

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia - Social and Behavioral Sciences 69 (2012) 1548 – 1554

Procedia
 Social and Behavioral Sciences

International Conference on Education and Educational Psychology (ICEEPSY 2012)

Teachers' Attitude to Test-driven Accountability; Developing a Comprehensive Approach to Assessment

"a Fatemeh Khalvandi*, b Mahin Chenari" *

*"a Allameh Tabatabaei, Faculty of Psychology and Educational Science, Tehran, Iran"**"b University of Qom, Faculty of Literature and Human Science, Qom, Iran"***Abstract**

If accountability means obligating to give a justification of what one has done then the critical question is how educational systems try to realize accountability. In Iran, like most other educational systems, a test-driven accountability system (TDAS) has been used. Teachers believe that TDAS is not a comprehensive assessment tool for this purpose. It is a one-dimensional tool that merely emphasizes on knowledge memorization and it is used for grading students or schools rather than educational progress. There should be other classroom assessment techniques that enable both instructors and students to mutually improve learning.

This study aims to determine how teachers in one district view the use of TDAS when being compared to a similar group in other districts and to explore teachers' opinion on other assessment techniques for developing a comprehensive approach to assessment. The research methodology is a mixed method. So a questionnaire with open-ended questions was deployed among 302 teachers who randomly selected from all secondary schools of the district and a questionnaire with closed-ended questions was used for 22 teachers who purposefully selected. Quantitative data showed that teachers had slightly positive attitude toward TDAS. However, there were no significant differences between male and female teachers' attitude. According to qualitative data and related literature, a comprehensive approach to assessment is proposed which involves inclusion of all three cognitive, affective, and psychomotor learning domains. This approach relies on patterns of data collected over time.

Keywords: *test-driven accountability system (TDAS), assessment techniques, teachers' attitude, learning domains*

© 2012 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.

1. Introduction

Education is a social enterprise for promoting shared values and common goals. The most shared priority is the development of the next generation as effectively and efficiently as possible given competing demands for limited resources. Education does not exist in a vacuum. Structural changes over the past decades have placed tremendous pressure on such public services as education to cope with the changing social, economic, political, and cultural forces. The current educational reform movement began in reaction to perceived shortcomings in education and international competition in all sectors. This worldwide movement has led to expectations for greater accountability and an increase in monitoring and evaluation of schools and systems (McEwen, 1995).

The scope of accountability is potentially quite large. Basically, when speaking about accountability, we mean who is held accountable by whom, for what, and with what practical consequences (Hoffer, 2000). Darling-Hammond and Ascher (1991) defined an accountability system as:

a set of commitments, policies, and practices that are designed to:

1. Heighten the probability that students will be exposed to good instructional practices in a supportive learning environment;
2. Reduce the likelihood that harmful practices will be employed; and
3. Provide internal self-correctives in the system to identify, diagnose, and change courses of action that are harmful or ineffective. (p. 2)

* E-mail address: Khalvandi@yahoo.com

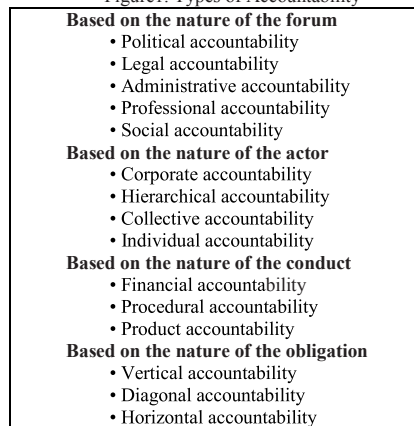
They identified at least five types of accountability mechanisms applicable to education: political, legal, bureaucratic, professional, and market. Legislators and school board members must regularly stand for election (political accountability). Citizens can ask the courts to hear complaints about public schools’ violation of laws regarding, for example, equal educational opportunity (legal). District and state/provincial education offices promulgate rules and regulations intended to ensure that schooling takes place according to set standards (bureaucratic). Teachers and other school staff must acquire specialized knowledge, pass certification exams, and uphold professional’s standards of practice (professional). Parents and students may choose the courses or schools they believe most appropriate. They may also be involved in school decision making (market) (p. 3).

Ideally each level of the education system would assume its share of responsibility:

- *States/provinces* would be responsible for providing equal and adequate resources to schools and for ensuring the enforcement of equity standards and standards of professional certification.
- *School districts* would be accountable for the policies they adopt, for equity in the distribution of school resources, and for creating processes that make them responsive to the needs and concerns of parents, students, and school-level staff.
- *Schools* would be accountable for equity in the internal distribution of resources, for adopting policies that reflect professional knowledge, for establishing means for continual staff learning, for creating problem-identification and problem-solving processes that drive continual improvements, and for responding to parent, student, and staff ideas.
- *Teachers* would be accountable for identifying and meeting the needs of individual students based on professional knowledge and standards of practice, for continually evaluating their own and their colleagues’ practices, for seeking new knowledge, and continually revising their strategies to better meet the needs of students. (Darling-Hammond & Ascher, 1991, p. 11)

Bovens (2007) has explained that the accountability is used in a rather narrow sense: a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct. He has pointed that types of accountability are a series of dimensions that can be used to describe the various accountability relations and arrangements.

Figure1. Types of Accountability



As the figure 1 shows accountability relation can be classified on each of the four dimensions separately.

According to Kirst (1990), accountability can be achieved through six broad approaches: performance reporting, monitoring and compliance with standards or regulations, incentive systems, reliance on the market, changing the locus of authority or control of schools, and changing professional roles. Since these strategies are not mutually exclusive, state/provincial or local governments can use several of these approaches simultaneously. Policy makers need to consider local contexts when determining the emphasis and balance among alternatives.

McDonnell and Elmore (1987) proposed a conceptual framework to help policy makers understand alternative policy instruments. They included four generic classes of instruments: mandates, inducements, capacity building, and system changing. *Mandates* are rules governing the action of individuals and agencies, intended to produce compliance. Mandates provide specific benefits to individuals and diffuse benefits to society. Examples include environmental regulation, nondiscrimination requirements, and speed limits. *Inducements* provide money to individuals or agencies in return for certain actions, for example, grants-in-aid to governments and in-kind grants to individuals. *Capacity building* involves the transfer of money for investment in material, intellectual, or human resources. Basic research is an example of capacity building. *System changing* involves the transfer of official authority among individuals and agencies to alter the system by which public goods and services are delivered.

Established deliverers lose their authority and new deliverers gain authority. Examples of system changing instruments include vouchers and deinstitutionalization (p. 8).

Student assessment has been the central tool for holding education accountable for results. Teachers assess students' work and assign marks for work completed. In most instances this type of assessment (whether for formative or summative purposes) lets parents and administrators know how well students are learning. In addition to assessment by classroom teachers, most provinces have provincial assessment and examination programs for accountability purposes. Both use the same instruments for all students writing the tests; they are distinguished by whether or not the marks count for individual students. Assessment programs are intended to monitor student achievement on the programs of study and provide stakeholders with information about how well students are learning. Typically, only selected subjects are assessed at two or more grade levels; the marks on these tests do not count for individual students. Examination programs, on the other hand, certify individual student achievement. This allows school jurisdictions and schools to compare their results (usually provided by the province) with aggregated results for all students in the province. Both types of program also provide information on the curriculum's effectiveness. Accountability is concerned with evaluating the effectiveness and efficiency of public education. In addition to student evaluation, many jurisdictions evaluate programs, schools, and school systems.

Noblit and Eaker (1988) argued that both the process and outcomes of evaluations are political and that the decision to subject a program and its participants to evaluation is a policy decision. Evaluations are political because they establish the bases for judgment (p. 127). These authors identified and discussed six evaluation approaches or designs: positivism, interpretivism, critical theory, aesthetics, collaborative research, and action research. For each approach, they classified the credibility of knowledge, the social relations among evaluation parties, and the political result.

Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1994) recognizes that evaluation is inevitable in that all people make choices, and that they do so by assessing the worth or merit of options. Evaluations should be conducted effectively, fairly, and efficiently. Four attributes of an evaluation organize the 30 standards: (1) *utility*, intended to ensure that an evaluation will serve the information needs of intended users; (2) *feasibility*, to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal; (3) *propriety*, to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results; and (4) *accuracy*, to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. The Joint Committee also reorganized the standards according to function to help users see that the standards can be used in planning, conducting, and reporting an evaluation: deciding whether to evaluate, defining the evaluation problem, designing the evaluation, collecting and analyzing information, reporting the evaluation, budgeting the evaluation, contracting for, managing, and staffing the evaluation (pp. vii–x).

In sum, accountability is not a new phenomenon in education. The distinctive feature of current calls for greater accountability refers to formal accountability, typically involving quantification of student outcomes. The different sets of variables involved in producing student outcomes each point to different actors engaged in specific activities that are accountable for different things. We are not going to consider all types of accountability since it is not possible in this article. Our focus is on bureaucratic accountability at class level factors that are controlled by teachers. Student assessment is the case that is usually used to measure this type of accountability. In Iran, like many other education systems all over the world, standard tests are used for this purpose. Educational systems expect the result of these tests show student educational progress. But the deficiencies of these scores for this purpose are readily apparent!

Actually, this study was started when I was present as a district headmaster of school principals in teachers' council sessions for analyzing the causes of growth and reduction of students' educational progress in a centralized test-driven accountability system (TDAS) which had been held by province evaluation commission at the first mid-year of 2011-2012 school year among all schools of Tehran province. The result of this evaluation indicated that the district under present study was gained 23th grade between 35 districts. The most teachers in response to official analysts said that if accountability means giving a justification of what one has done TDAS is not a comprehensive evaluation tool for this purpose. It is a one-dimensional tool that merely emphasizes on knowledge memorization. As these sessions proceeded, these similar reactions of teachers forced me to seek for an appropriate evaluation tool for accountability system.

The main purpose of this study is to provide a comprehensive framework for students' assessment while investigating teachers' attitude to previous method that is a test driven one. So, the *research questions are*:

Q1: What is the attitude of teachers to TDAS?

Q2: Are there any significant differences between male and female teachers' attitude to TDAS?

Q3: What is a comprehensive framework to student assessment?

2. Methodology

2.1 Sample

In 2011-2012 school year, there are 1400 teachers in Bahastan region, according to Krejcie and Morgan's table (1970), Sample size was 302. A total of 302 surveys were distributed by the researchers. Participation in this study was voluntary, participants' identities were anonymous and confidentiality of responses was assured. For assurance of correct responding by participants and explaining possible problems or ambiguities, researchers were present when participants were completing the surveys. A few surveys were not delivered. At last, a total of 290 questionnaires were found acceptable to use in the analysis. For the qualitative section of study, 22 people were selected by purposeful method for 22 open-ended questions.

2.2 Measures

Demographics. The three demographic items examined age, gender and total work experience. Two of these variables, age and work experience, were measured using a ratio scale and gender being measured by use of a nominal scale. *A questionnaire* that combines two existing surveys was used. The first section of the questionnaire used a survey developed and validated by Stronge (1984) which measured attitudes of teachers toward competency tests. The twenty questions from the original survey have validity, and an internal consistency correlation of .91 was established using Cronbach's Coefficient Alpha (Moore, 1988). For all measurement items, respondents were asked to indicate their agreement on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The second portion of the survey gained teachers' expectations for a comprehensive framework to student assessment and contained 22 open-ended questions.

2.3 Data analysis

The survey determined the teachers' attitudes toward the TDAS. Responses were statistically analyzed according to frequency, percent, mean, and standard deviations. To determine whether any significant differences in teacher attitude based on the demographic information, The Pearson product-moment correlation coefficient and T Test were applied to demographic categories in relation to teacher attitude. Content analysis was used for the data which acquired from interviews. Analytic activities were done through the spiral model of Boeijie (2010); Sampling, data collection, data, open coding (segmenting), list of codes, axial coding, list of categories, selective coding (reassembling) and conceptual model.

3. Results

3.1 Quantitative data

Demographic information was obtained from the sample group in terms of age range, gender, and number of years of experience as a teacher. The frequencies and range of responses are included in Table 1. The majority of teachers in the sample was between the ages of 30 to 38, was female, taught various grades in high school, and has been teaching between 8 and 15 years.

Table 1: Demographics of Sample

Demographic items	Number	Percent
• Age Range		
• 22- 29 years	98	33.79%
• 30- 38 years	153	52.75%
• 38- 45 years	34	11.72%
• 46- 53 years	5	1.74%
• Over 53 years	0	0%
• Total	290	100%
• Gender		
• Female	211	72.75%
• Male	79	27.25%
• Total	290	100%
• Teaching Experience		
• 1-7 years	65	22.41%
• 8-15 years	138	47.58%
• 15- 22 years	75	25.86%
• More than 23 years	12	4.15%
• Total	290	100%

The attitudes of teachers toward the TDAS test were measured by responses to statements on the survey. The frequency and percentage were calculated for each response, in addition to the mean and standard deviation. The 20 statements, beginning with statement 4, concerning attitude toward the TAAS test were designed in order to measure the degree of agreement with each statement, with a five point variation ranging from strongly agree to strongly disagree. These scores ranged from 88 to 29 on a 100 point scale. A median score of 67 on the attitude survey established the divide between positive and negative attitudes toward the TDAS. Respondents scoring above the median score of 67 for attitude toward the TDAS were rated as positive attitude, and those with scores below 67 rated as negative attitude toward the TDAS. Table 2 lists the ratings of attitude toward TDAS.

Table 2: The Ratings of Attitude toward TDAS

Positive Attitude		Negative Attitude	
No.	(%)	No.	(%)
56	19.31%	234	80.69%

In order to determine whether a significant difference exists between the respondents rated as having positive attitudes toward the TAAS and those rated as having negative attitudes toward the TDAS, the means between the top five positive attitudes rated respondents and the bottom five negative attitude-rated respondents were compared. A t-test was used for the analysis, showing a significant difference ($t = 7.713, p < .05$). Table 19 illustrates the relationship between the teachers with the most positive attitudes toward the TDAS and those with the most negative attitude toward the TDAS.

Table 3: T-Test between Teachers with Positive Attitudes and Teachers with Negative Attitudes

Positive Attitude Score	Negative Attitude scores	df	t
88	29	8	7.713
81	32		
79	37		
78	42		
76	52		

The Pearson product-moment correlation coefficient, performed on the means of respondent gender scores and the means of teacher attitudes toward the TAAS did not show any significant relationships ($p > .05$) between gender and attitude. The values of $r = .000$ indicated that there was no relationship between gender and teacher attitudes toward the TDAS. Table 3 illustrates the information regarding the Pearson product moment correlation between gender and teacher attitude towards TDAS

Table 4: Pearson Product-Moment Correlation Coefficient between Gender and Teacher Attitudes toward the TDAS

Gender vs.	N	r	p
Teacher Attitudes Toward TDAS	290	.000	.999

3.2 Qualitative Data

The main problem that we were going to find out is “what is a comprehensive assessment tool for evaluating educational progress? This main question was declined in 10 open-ended sub-questions. First, all of the suggested assessment tools and methods by teachers were extracted. Then, similar themes have been categorized in the same topic. After all, based on similarities between different topics, main cores categories were recognized. The following tables show the related information.

Table 5: all of the suggested assessment tools and methods by teachers

Row	Suggested assessment tools and methods by the teachers	Frequency	Percent
1	Observation of the classrooms twice in per midyear by principal or province representatives	15	68%
2	longitudinal assessment data	20	90%
3	use of performance-based versus traditional assessments	18	81%
4	assessment that calls for the actual performance of a skill and the creation of a response	18	81%
5	cumulative portfolios of students' work,	20	90%
6	assessments measure the full range of standards	22	100%
7	Focus groups including teachers and experts involved in the articulation of standards	12	60%
8	avoiding faulty test	20	90%
9	When standards are determined externally, teachers lose a measure of autonomy	22	100%
10	Using good measurement entails samples of a comprehensive range of skills	22	100%
11	Using various teaching methods: Minute Paper, Muddiest Point, One Sentence Summary, Pro Con Grid, Student Generated Test Questions, Classroom Opinion Poll, Goal Ranking and Matching, Process Analysis, Chain Notes, Group Work Evaluations, Using Pretests	22	100%
12	Provide short-term feedback about the day-to-day learning and teaching process at a time when it is still possible to make mid-course corrections.	18	81%
13	Encourage the view that teaching is a formative process that evolves over time with feedback.	22	100%
14	Provide useful information about student learning with a much lower investment of time compared to tests, papers, and other traditional means of learning assessment.	8	36%
15	Help to foster good rapport with students and increase the efficacy of teaching and learning.	5	23%
16	Assessing Skill in Analysis and Critical Thinking	19	86%
17	Using background Knowledge Probe – Short	12	54%
18	Assessing Skill in Syntheses and Critical Thinking	7	32%
19	Using annotated Portfolios	5	23%
20	Assessing Skill in Problem Solving	4	18%
21	Using audio- and Videotaped Protocols	8	36%
22	Assessing Learner Reactions to Teachers and Teaching	21	95%

23	Using notes Teacher-Designed Feedback Forms	7	32%
24	Using Punctuated Lectures	20	90%
25	Assessing Skill in Application and Performance	22	100%
26	Using Directed Paraphrasing	4	18%
27	Using Student-Generated Test Questions	13	59%
28	Assessing Students' Awareness of Their Attitudes and Values	12	54%
29	Student self-assessing	22	100%

Table 6: Categorizing similar themes

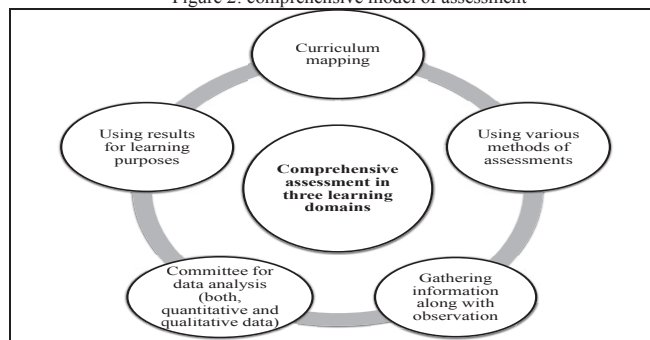
Row	Categorizing similar themes
1	<ul style="list-style-type: none"> – Observation of the classrooms twice in per midyear by principal or province representatives – longitudinal assessment data
2	<ul style="list-style-type: none"> – use of performance-based versus traditional assessments – assessment that calls for the actual performance of a skill and the creation of a response – cumulative portfolios of students' work, – Assessing Skill in Application and Performance – Assessing Skill in Problem Solving – Assessing Learner Reactions to Teachers and Teaching – Assessing Skill in Analysis and Critical Thinking
3	<ul style="list-style-type: none"> – assessments measure the full range of standards – Focus groups including teachers and experts involved in the articulation of standards – avoiding faulty test – Using good measurement entails samples of a comprehensive range of skills – When standards are determined externally, teachers lose a measure of autonomy
4	<ul style="list-style-type: none"> – Using Minute Paper, Muddiest Point, One Sentence Summary, Pro Con Grid, Student Generated Test Questions, Classroom Opinion Poll, Goal Ranking and Matching, Process Analysis, Chain Notes, Group Work Evaluations, Using Pretests, Student-Generated Test Questions, Directed Paraphrasing, audio- and Videotaped Protocols, notes Teacher-Designed Feedback Forms, Punctuated Lectures, annotated Portfolios, background Knowledge Probe – Short – Student self-assessing
5	<ul style="list-style-type: none"> – Encourage the view that teaching is a formative process that evolves over time with feedback. – Provide short-term feedback about the day-to-day learning and teaching process at a time when it is still possible to make mid-course corrections. – Help to foster good rapport with students and increase the efficacy of teaching and learning.

Table 7: Common heading for thematic categories

Common heading for thematic categories	Row(S)
Cognitive domain	1 & 4
Attitude domain	3 & 4
Psychomotor domain	5, 2 & 4

According to these findings and related literature, a comprehensive model of assessment is proposed. Figure 2 shows this comprehensive model.

Figure 2: comprehensive model of assessment



As we see, assessment in all three cognitive, attitude and psychomotor learning domains is the core and basic of this model. A comprehensive assessment should include all of what teachers expect to observe. These considerable themes are noticed in table

7. In comparison with TDAS, this model has positive points such as feedback for learning, formative (ongoing), quick and easy, and not graded. Observations of the classrooms, twice in per midyear, which usually are structured around checklists are derived from popular conceptions of best practice. Longitudinal assessment data also allow decision makers to explore and analyze test score gains and the other student products for school-level analysis. Constructing educational committee in school and province level provides a complete feedback and Helps to foster good rapport with students and increases the efficacy of teaching and learning. When all the people who are responsible for teaching and learning are involved in assessment process they encourage becoming more active and will have positive attitude. So they will motivate to do everything to gain better results.

It seems that this comprehensive model is a reliable road map to get rid of one-dimensional tools such as test driven assessment.

4. Conclusions

Hoffer (2000) has pointed that the evidence is murkier concerning whether increased efforts to build accountability have had the ultimate desired effect of raising students' academic achievement. Since TDAS is designed to measure only the minimal standards have been employed by some states and localities in the past decades, these assessments may lead to a leveling-down of teaching and learning if only minimal become the focus of these educational assessment systems. Comprehensive assessment systems avoid this danger. Another concern is validity. The validity of an assessment can, and should be, evaluated in several ways. The most basic level of validity is face validity, whether the instrument in question appears to relevant parties as a reasonable measure of important standards. Researchers have developed several methods of gaining feedback from key groups on school assessments. According to findings, focus groups including teachers and experts involved in the articulation of standards can be helpful in screening the proposed assessment topics down to a manageable size that includes the most important topics. Other ways of gauging validity rely on data collected from the proposed items. Items and scales that are not correlated with relatively well-established outcomes or other factors are often found on closer scrutiny to be confusing or difficult to administer.

In this proposed model, teachers and the other persons like principals who are responsible for learning are involved in assessment process. So as Coleman (1997) said we can expect that autonomy can be increased as an incentive for schools and for teachers to reach higher standards. Furthermore, this kind of assessment entails samples of a comprehensive range of skills, so that the feedback to students and teachers are not abstract. TDAS loses legitimacy because means of improvement are difficult to identify. On the other hand, if the feedback is specific and concrete, there is a risk that corrective actions will improve skills not sampled in the next assessment. There is also a risk that teaching and learning will become focused on the requirements of the tests as opposed to the subject matter more practically conceived. Besides, School-average achievement test scores give little or no guidance to classroom teachers trying to assess and improve their effectiveness.

The proposed model gives you an idea of the range of possibilities when seeking for a comprehensive assessment tool that promote learning. Undoubtedly, many other educational systems have tried tactics similar to these as well as others I haven't thought about. The common theme here is finding ways to make a comprehensive way for assessment and make teachers and principals explicitly aware of the broader learning goals of the course and require them to assess the full range of students' performance and their learning practices.

5. References

- Boeije, H. 2010. *Analysis in Qualitative Research*. Sage Publications Ltd.
- Bovens, M. (2007), Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal*, 13: 447–468. doi: 10.1111/j.1468-0386.2007.00378.x
- Coleman, J. S. (1997). Output-driven schools: Principles of design. In J. S. Coleman, B. Schneider, S. Plank, K. S. Schiller, R. Shouse, & H. Wang (Eds.), *Redesigning American education*. Boulder, CO: Westview Press.
- Darling-Hammond, L. & Ascher, C. (1991, March). *Creating accountability in big city systems*. New York, NY: ERIC Clearinghouse on Urban Education. Distributed by the Midwest Center for Drug-Free Schools and Communities, Oak Brook, IL.
- Darling-Hammond, L., & Ascher, C. (1996). *Creating accountability in big city schools*. New York: ERIC Clearinghouse of Urban Education at Teachers College, Columbia University.
- Hoffer, T.B. (2000). Accountability in education. In M.T. Hallinan (Ed.), *Handbook of the sociology of education* (pp. 529 – 544). New York: Kluwer.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Kirst, M. W. (1990). *Accountability: Implications for state and local policymakers*. Washington, DC: U.S. Department of Education.
- McDonnell, L. M., & Elmore, R. F. (1987). *Alternative policy instruments*. Santa Monica, CA: RAND Corporation.
- Noblitt, G. W., & Eaker, D. J. (1988). Evaluation designs as political strategies. In J. Hannaway & R. Crowson (Eds.), *The politics of reforming school administration: The 1988 yearbook of the Politics of Education Association* (pp. 127–138). Philadelphia, PA: Falmer.
- Stronge, J.H. (1984). The impact of the attitudes of high school principals on minimum competency testing. *Doctoral dissertation*, University of Alabama, 44/09, 2648A.