

Spring 3-1-1998

Creating Legally Valid School Administrator Evaluation Policy in Utah

Steve Baldrige

Follow this and additional works at: <https://digitalcommons.law.byu.edu/elj>



Part of the [Educational Administration and Supervision Commons](#), and the [Education Law Commons](#)

Recommended Citation

Steve Baldrige, *Creating Legally Valid School Administrator Evaluation Policy in Utah*, 1998 BYU Educ. & L.J. 19 (1998).
Available at: <https://digitalcommons.law.byu.edu/elj/vol1998/iss1/3>

.

This Article is brought to you for free and open access by BYU Law Digital Commons. It has been accepted for inclusion in Brigham Young University Education and Law Journal by an authorized editor of BYU Law Digital Commons. For more information, please contact hunterlawlibrary@byu.edu.

CREATING LEGALLY VALID SCHOOL ADMINISTRATOR EVALUATION POLICY IN UTAH

*By Steve Baldrige**

I. INTRODUCTION

The use of personnel evaluations as a tool to spur improved performance has been a common element of public school reform efforts. Teacher evaluations were the first form of personnel evaluations to become widely used in education. In the last decade, however, teacher evaluation policies across the country were revised because of concerns about their validity. In conjunction with their efforts to improve the quality of teacher evaluations, many states and districts began to require administrator evaluations as well.

Today, public school administrator evaluation is required by law in forty-seven states.¹ The 1990 amendments to the Utah Educator Evaluation Act require “valid and reliable” evaluations of all certificated school employees—including school administrators.² To this author’s knowledge, Jordan School District remains the only district in Utah to have taken steps necessary to comply with the statutory standards of validity. An adverse decision in a 1991 administrative hearing motivated the district to create new teacher and administrator evaluation policies.³ The hearing examiner found that an educator could not be denied merit pay because of sub-standard evaluations under an evaluation policy for which the district could not produce legally adequate validation evidence.⁴ Many other districts would likely fare no better.

* Assistant Professor of Educational Leadership, Barry University. B.A., 1989, Washington & Lee University; J.D., 1992, Brigham Young University; Ph.D. in Educational Leadership, 1996, Brigham Young University. The author appreciates suggestions from Drs. Steven J. Hite, E. Vance Randall, Cindy Skaruppa and Kevin Worthen on previous drafts of this article.

1. See C. Furtwengler, *State Requirements for the Evaluation of School Administrators*, Paper presented at the American Educational Research Association annual meeting, New Orleans, LA, 1994.

2. See generally UTAH CODE ANN. § 53A-10-101 *et seq.* (1990).

3. Dr. David Sperry, Report of the Hearing Examiner (January 14, 1991) (on file with Pat Thompson at the Jordan School District Office of Program Services and Evaluation).

4. *Id.* at 18.

This article explains the statutory language and sound professional practices related to school administrator evaluation in the State of Utah by responding to a series of questions. As an example of sound practice, this article describes elements of the Jordan Administrator Evaluation Policy—implemented by the Jordan School District in the 1996-97 school year. The core elements of its validation study are noted to give guidance to other districts throughout Utah and across the country. A summary of the validation study results are found in Appendix B.

II. IS THE DISTRICT'S ADMINISTRATOR EVALUATION POLICY LEGALLY VALID?

Although some districts may still be confused by a discontinuity between the preamble and body of the Utah Educator Evaluation Act, administrator evaluations are now required by the statute which originally applied only to teachers. The statutory intent that evaluations “allow the educator and the school district to promote the professional growth of the *teacher*, to identify and encourage *teacher* behaviors which contribute to student progress, to identify *teachers* according to their abilities, and to improve the education system”⁵ demonstrates the original teacher focus when the statute was passed in 1988. Yet this language is at odds with the 1990 amendments which define “educator” as “any individual, except the superintendent, employed by a school district who is required to hold a professional certificate issued by the State Board of Education.”⁶

Preambulatory language—language which expresses the intent of a law—is given significantly less weight by courts than directive language. The Utah State Office of Education clarified this point in a letter to Jordan School district by stating that the Utah Educator Evaluation Act applies to school administrators as well as to teachers.⁷ All districts in Utah should evaluate their administrators using evaluation policies that meet the specific provisions of this statute.

To answer the question of how well a district's administrator evaluation policy will hold up against a legal challenge requires attention to both sound policy development and appropriate handling of particular evaluations. First, does the administrator evaluation policy as approved by the school board comply with all laws applicable to such policies? Second,

5. UTAH CODE ANN. § 53A-10-101 (1988)(emphasis added).

6. UTAH CODE ANN. § 53A-10-102(2) (1990).

7. See Letter from Dr. Bob Ellison, The Institute for Behavioral Research and Creativity, to Dr. Barry Newbold, Jordan School District Office of Program Services and Evaluation (February 2, 1993), and Letter from Dr. Dave Nelson, Utah State Office of Education, to Dr. Barry Newbold, Jordan School District Office of Program Services and Evaluation (February 2, 1993) (both on file with Pat Thompson at the Jordan School District Office of Program Services and Evaluation).

does the evaluation policy as it actually functions in a particular case comply with relevant law? The legality of actions taken as a result of administrator evaluations rests on compliance in both the evaluation policy's form and implementation; however, a district's evaluation policy validation study must focus on the first question while inservice must focus on the second. This article focuses on the policy question with particular attention paid to the creation of legally valid school administrator evaluation policies in the State of Utah.

The question of facial legality will almost always be the first direction of a legal attack on administrator evaluations. If the form of an argument is invalid, then the truthfulness of each statement within it is irrelevant. As applied to the legal context of administrator evaluation, this rule indicates that if the design of an evaluation is flawed, then a court will never need to reach the legitimacy of specific steps taken under it.

A. *Was the Administrator Evaluation Policy Created as Statutorily Mandated?*

The evaluation policy must have been created in a manner consistent with the law. If not done, then no amount of validation evidence will legitimize the evaluation policy. This point is crucial, and one on which the Utah statute is clear. Sections 53A-10-103 and 105 describe mandatory elements of the evaluation policy's development.

1. *Are there equal numbers of administrators and teachers on the committee appointed by the district board of education to develop the administrator evaluation policy?*

The first element of evaluation development required by the statute is that "[e]ach local school board [must] develop an evaluation program in consultation with its educators through appointment of a joint committee . . . comprised of an equal number of classroom teachers and administrators appointed by the board."⁸ The statutory language allows nominations.⁹ The statute contemplates one school board approved committee charged with formulating an evaluation policy for all educators within the district.

A district may choose to create separate task forces for teacher and administrator evaluation policies, so long as "equal number[s] of classroom teachers and administrators appointed by the board"¹⁰ sit on the committee that ratifies the evaluation policies. Indeed, the Utah State Office of

8. UTAH CODE ANN. § 53A-10-103(1) and (2) (1990).

9. UTAH CODE ANN. § 53A-10-103(3) (1990).

10. UTAH CODE ANN. § 53A-10-103(2) (1990).

Education raised no concerns when informed in 1993 that Jordan School District intended to create a separate Jordan Administrator Evaluation Task Force (JAETF), and that the task force would "include[] administrators, teachers, classified personnel, and parents."¹¹ This precedent should protect districts that choose to follow Jordan's lead from legal attack on this point so long as the composition of the ratifying committee preserves equal numbers for teachers and administrators.

2. Will the principal of each school orient all educators assigned to the school toward the purpose and method of the administrator evaluation policy before evaluations under it begin?

The second element begins to be fulfilled with information included in questionnaires sent to district employees through their principals as part of the validation study. The purpose of the evaluation policy should be expressed in the questionnaire's cover letter. The usefulness of potential indicators proposed for use in assessing administrator performance should be a focus of validation questionnaires designed by the task force. The opportunity to suggest content and procedures should be another. In this way, respondents will learn about potential elements of the policy in the creation process. Later, when the evaluation policy content and procedures are approved by the school board, principals must receive orientation information to share with their faculties in order to comply with the statute.¹² Principals' orientation of their faculties toward the approved policy must include "the purpose of the evaluations and the method used to evaluate."¹³

B. Does the Administrator Evaluation Policy, as Approved by the District School Board, Meet the Statutory Standard for Validity and Reliability?

Compliance with the statutory standard for validity and reliability is achieved by looking to the statutory standard, then organizing district efforts to meet it. The standard is found in § 106(6) which reads: "Any educator evaluation program adopted by a local school board in consultation with a committee shall provide the following: . . . a reliable and valid evaluation consistent with generally accepted professional standards for personnel evaluation systems."¹⁴

11. Letter from Dr. Barry Newbold, Jordan School District Office of Program Services and Evaluation, to Dr. Bob Nelson, Utah State Office of Education (January 28, 1993) (on file with Pat Thompson at the Jordan School District Office of Program Services and Evaluation).

12. UTAH CODE ANN. § 53A-10-105 (1990).

13. *Id.*

14. UTAH CODE ANN. § 53A-10-106(6) (1990).

One common use of evaluation policies is to make administrative decisions regarding personnel. In 1991, Jordan School District and one of its teachers addressed, in an administrative hearing, the question of whether the district could withhold merit pay from an educator who had sub-standard evaluation scores. Because the statutory language of § 106(6) reads the same today as it did then, comments in the Report of the Hearing Examiner¹⁵ regarding the validity and reliability of Scale 11 of the JITC¹⁶ remain persuasive on the question of who is authorized to define “reliable and valid” as used in this section. The Examiner quoted from the Utah State Attorney General’s Opinion in response to an informal request to clarify the law.¹⁷

It is clear that the board has ultimate authority to develop and adopt the evaluation program for the district. In order to do this the board must be able to define the terms and conditions of the program so long as the program complies with the law. Because the terms referred to are not defined in the statute itself, the board in its reasonable discretion may define those terms. Otherwise, the board would be prevented from carrying out the intent of the law for the local board to formulate an appropriate evaluation program.

However, while the board’s discretion is entitled to considerable weight it is not unlimited and conceivably a board could arrive at an extreme and *arbitrary* definition which a court might find beyond the board’s authority. *Presumably a definition based on the joint committee proposal would be acceptable to most educators in the district and therefore a reasonable standard for evaluation*, subject to the board’s ultimate approval. . . . Since the term consistent means compatible, harmonious, or in agreement with, it should not be interpreted that the board’s definition need duplicate, be identical to, or be exactly the same as those found in the appropriate professional standards. There is room for adaptation as long as it meets the test of being “consistent with.”¹⁸

15. Dr. David Sperry, Report of the Hearing Examiner (January 14, 1991) (on file with Pat Thompson at the Jordan School District Office of Program Services and Evaluation).

16. Jordan School District’s Teacher Evaluation Instrument prior to the current J-PAS.

17. Informal Opinion Request No. 87-43 - Educator’s Evaluation Act - Question 2, Utah State Attorney General (September 25, 1987).

18. Dr. David Sperry, *supra* note 3, at 12-14 (quoting from Informal Opinion Request No. 87-43 - Educator’s Evaluation Act - Question 2, Utah State Attorney General (September 25, 1987)) (emphasis added).

To further clarify the meaning of “generally accepted professional standards” in § 106(6), the Examiner stated,

Common definitions of the term “general” include “not precise or definite,” “not limited to a precise application,” and “not specific.” Had the legislature intended that the specific standards contained in [a] manual and only those standards be applied, one would logically assume that the actual language of the statute would have stipulated that.¹⁹

Thus, to begin a validation study of the content for an administrator evaluation policy, a district task force should agree on definitions for the terms “valid” and “reliable”, which can be noted by the board of education when it approves final versions of the administrator evaluation policy. These definitions should be guided by—but not mechanically confined to—professional standards in the field of educational personnel evaluation. Three sources of such standards are worthy of particular attention.

1. *How does The Personnel Evaluation Standards: How to Assess Policies for Evaluating Educators define valid and reliable?*

*The Personnel Evaluation Standards (PES)*²⁰ contains two especially relevant “Accuracy Standards”—Valid Measurement and Reliable Measurement. *PES* defines the standard for valid measurement as: “procedures . . . chosen or developed and implemented on the basis of the described role and intended use [of the evaluation], so that the inferences concerning the evaluatee are valid and accurate.”²¹ Explanations, guidelines, common errors and illustrative cases are included to aid in defining the proposed standards.

These supplementary materials point out various issues related to validity. “Validity means that what was intended to be measured was measured.”²² “The various strategies for presenting evidence of validity include correctly inferring a trait or characteristic . . . from empirical evidence; correctly inferring a relationship between an assessed characteristic used to predict a level of performance and the observed future performance; . . . presenting the self-evident relationship between the content of a measurement device or procedure and performance, where the measure-

19. Dr. David Sperry, *supra* note 3, at 11.

20. JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (CHAIRIED BY DANIEL STUFFLEBEAM), *THE PERSONNEL EVALUATION STANDARDS: HOW TO ASSESS POLICIES FOR EVALUATING EDUCATORS* (1988) (referred to by name in Dr. David Sperry’s Hearing Report, at 15).

21. *Id.* at 98.

22. *Id.*

ment content includes a representative sample of the job tasks involved;”²³ and comparing the results of the chosen style of measurement with the results of other styles of measurement to validate the choice.²⁴

The standard for reliability reads: “Measurement procedures should be chosen or developed and implemented to assure reliability, so that the information obtained will provide consistent indications of the performance of the evaluatee.”²⁵ The accompanying explanation states that “[a] reliable measure is one that provides consistent information about the performance being assessed . . . Consistency should be sought across different indicators of the same criterion (internal consistency), across different observers of the same behavior (observer agreement), and across different occasions on which a behavior is observed (stability).”²⁶

PES suggests the employment of “multiple measures, multiple observers, and multiple occasions for data collection as appropriate to minimize inconsistency and discern consistent patterns and trends.”²⁷ Finally, it points out that “[r]eliability is a necessary, but not a sufficient, condition for validity.”²⁸

2. *How do validation study design experts define valid and reliable?*

Dr. Steven J. Hite, one of the consultants employed by Jordan School District while creating its Jordan Administrator Evaluation System, gave a presentation to the JAES Task Force regarding the definitions of validity and reliability. Meeting minutes summarize his presentation by stating the following:

Reliability is determined when measurement can take place repeatedly resulting in the same outcome. Validity deals with the instrument’s ability to accurately assess those skills or behaviors for which it was constructed. . . . [T]he challenge of the [task force] is to choose indicators which are not constructs to avoid . . . an [evaluation] instrument that is not measurable.”²⁹

Reviewers must strive for consistency in collection and interpretation of information gathered for evaluation, and admit limitations to inferences

23. JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION at 98.

24. *Id.* at 99.

25. *Id.* at 104.

26. *Id.* (citations omitted).

27. *Id.* at 105.

28. *Id.*

29. Jordan Administrator Evaluation Task Force, Meeting Minutes (September 15, 1994) (on file with Pat Thompson in the Jordan School District Office of Program Services and Evaluation).

drawn from collected data. The content and procedures established by the task force must guide future reviewers in these tasks.

3. How are "valid and reliable" defined in the relevant literature?

A review of the literature on administrator evaluation also sheds some light on the standards proposed for educational evaluation and the generality with which they are accepted for use. Most competency domains promoted by professional organizations and state legislatures are purported to be valid nationally or statewide for named categories of administrators.³⁰ Competency domains promoted by other types of authors usually are not accompanied by validation claims.³¹ Because the evaluation policy must comport with the state statute, some state-wide validation procedure is appropriate, but procedure should not replace the district-wide validation procedure's centrality. Such a dual study was done by Jordan School District.³²

The formulation of validation procedures involves numerous potentially confusing validity-related terms. For example, Dr. Nelson of the Utah State Office of Education suggests that an administrator evaluation policy should be "subjected to careful *content validation*" and "a small *construct validation* study."³³ To strengthen claims of the instrument's validity, he suggests comparing the results of evaluations based on the new instrument to "supervisory or peer nominations" of administrators who excel in the evaluation domains.³⁴ This type of comparison would demonstrate that reviewers can measure excellence in the domains with the instrument in a way that produces results similar to the alternative, as suggested by *PES*.

In contrast, one of the consultants employed by Jordan School District during the creation of its teacher evaluation policy has opined that "some minimal level of *criterion validated* statistical support would meet the [legal] requirements"³⁵ for validation of an administrator evaluation policy. Further, when administrator numbers in categories are less than 30, a "systematic *content validation* [approach]," "that people generally believe is

30. Jordan Administrator Evaluation Task Force Planning Committee, Meeting Minutes (October 25, 1994) (on file with Pat Thompson in the Jordan School District Office of Program Services and Evaluation).

31. *Id.*

32. Summary results of that study and the final set of categories and performance indicators that are used to evaluate administrator performance in Jordan School District are found in Appendix B following this article. For further information, please write the author at Barry University, 11300 NE 2 Ave., Powers #258, Miami Shores, FL 33161, or call 1(800) 756-6000, extension 3715.

33. Letter from Dr. Bob Nelson, *supra* note 7, at 1.

34. *Id.*

35. *Id.*

meaningful and appropriate," is suitable in lieu of statistically based validation procedures.³⁶ He has also suggested a "peer nomination procedure" as a check on the validity of evaluations under the new policy, but stressed *content validation* as the necessary first step.³⁷

Examiner Sperry determined that the JTC validation study was too "modest" even though it included a "consensus *validity*" step, and a test of objectivity.³⁸ "The demands of the law," he wrote, "require greater technical assurances."³⁹ The Examiner emphasized that, "The record does not adequately support the proposition that the District . . . has sufficient evidence in the manner prescribed by law to demonstrate that the grievant has received a reliable and valid evaluation."⁴⁰ The examiner referred to the *PES* standards and counseled the district to establish "a continuous validity measurement program" as part of validation efforts.⁴¹ Naturally, one wonders how all these terms relate to formulation of an administrator evaluation policy.

Measurement and Evaluation in the Schools (MES) explains that today validity is generally considered an integrated concept, while evidence marshaled to demonstrate the degree of validity an evaluation policy enjoys is commonly divided into four categories: face validity-related, content-related, criterion-related, and construct-related.⁴² Face validity-related evidence is the least technical type.⁴³

a. *What is meant by "face validity" and how can the district create relevant evidence for its administrator evaluation policy?* "Face 'validity' . . . refers to the degree to which a measurement instrument appears to measure that which it is intended to measure."⁴⁴ For example, the wording and format of an administrator evaluation policy's job categories and performance indicators should be easy for evaluatees and other affected groups to read and understand. The categories and performance indicators should also "appear . . . to be relevant and to measure what is intended."⁴⁵ Because many of the audiences for administrator evaluation are not adept at analyzing the technical issues of validation, the evaluation pol-

36. Letter from Dr. Bob Nelson, *supra* note 7, at 1.

37. *Id.* at 2.

38. Dr. David Sperry, *supra* note 17, at 17.

39. *Id.*

40. Dr. David Sperry, *supra* note 17, at 18 (emphasis in original).

41. *Id.*

42. B. R. WORTHEN, ET AL, MEASUREMENT AND EVALUATION IN THE SCHOOLS 180 (1993).

43. *Id.* at 188.

44. *Id.*

45. *Id.*

icy is more likely to receive their respect and cooperation if it seems fair and appropriate to them.

Fundamentally, the question of face validity concerns rapport and public relations . . . Certainly if test content appears irrelevant, inappropriate, silly, or childish, the result will be poor cooperation, regardless of the actual validity of the test. Especially in adult testing, it is not sufficient for a test to be objectively valid. It also needs face validity to function effectively in practical situations.⁴⁶

This type of evidence relates closely to the feasibility standard in *PES* called "political viability." Worthen, Borg and White counsel evaluation policy creators that "[t]he personnel evaluation policy should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the policy work."⁴⁷

Face validity can be demonstrated through evidence that comments by evaluators, evaluatees, and other affected groups are considered by the district as revisions are made. Stylistic changes to the document based on reviews by formatting specialists, editing performed by a task force, incorporating or responding to editorial comments, and periodic reviews are other indicia. To be prepared against the possibility of future litigation based on an evaluation under the policy, the district should maintain records of these efforts to ensure face validity evidence is available.

b. What is meant by "content validation" evidence and how can the district create relevant evidence for its administrator evaluation policy?

Evidence of this type is logical/rational rather than empirical/statistical. The basic issue in content validation is *representativeness*. In other words, how adequately does the content of the [evaluation] represent the entire body of content to which [its] user intends to generalize? Since the responses to a[n evaluation] are only a sample of a[n evaluatee's] behavior, the validity of any inferences about that [person] depends upon the representativeness of that sample . . . Ideally, a[n evaluation] should sample *all* important aspects of the content domain. No important parts of the domain should be underrepresented or excluded.⁴⁸

In addition, the district task force which creates the evaluation instrument must seek a final product that is "free from the influence of irrelevant variables that would threaten the validity of inferences based on the observed scores."⁴⁹

46. ANNE ANASTASI, *PSYCHOLOGICAL TESTING* 144 (6th ed. 1988).

47. JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, *supra* note 20, at 75.

48. B. R. WORTHEN, ET AL, *supra* note 42, at 181.

49. *Id.* at 182.

MES explains that content validation evidence is often gathered through recourse to colleagues. This is usually done by asking them to rate instrument items' congruence with instrument objectives.⁵⁰ Seven steps from MES are summarized here:

1) Specify, as clearly as possible, the domain of behaviors to be measured.

2) Analyze the behavioral domain, then subcategorize it into more specific topics, subject-matter areas, or clusters of objectives.

3) Draw up a set of evaluation specifications that shows not only the content areas or topics to be covered, but also the relative emphasis to be placed on each.

4) Decide how many questions to include in the evaluation instrument. Remember that although adding questions tends to increase reliability under research conditions, evaluator and evaluatee fatigue can reduce this in day-to-day practice.

5) Determine how many items will need to be developed in each category to make sure there is representative coverage of all content areas and categories of objectives.

6) Construct or select evaluation instrument items for each category.

7) Enlist colleagues or other content experts to review your items. This will help to identify unwitting biases you might bring to the item-writing task, as well as to strengthen the final set of items that are selected.⁵¹

Records of the steps taken by the district to develop its administrator evaluation policy serve as content validity evidence. These steps demonstrate a feedback loop process of comparisons between the instrument creators and various sets of relevant colleagues. First, the creation task force may conduct a review of proposed administrator evaluation domains, indicators and instruments found in the professional literature. The results of such a review and a demonstration of the degree of convergence between other sources and the domains found in the National Policy Board for Educational Administration's *Principals for our Changing Schools: Knowledge and Skill Base (Principals)* was created for Jordan School District and presented in 1994.⁵²

Based on the results of that literature review, the Jordan School District task force focused on the twenty-one domains described in *Principals*. In deference to the usability of the evaluation policy, the task force limited the number of domains in the instrument to seven. Several of the domains

50. B. R. WORTHEN, ET AL, *supra* note 42, at 182.

51. *Id.*

52. Jordan Administrator Evaluation Task Force, Meeting Minutes (June 27, 1994) (on file with Pat Thompson in the Jordan School District Office of Program Services and Evaluation).

were combined and called subcategories. Several other important concepts were grouped together and designated as a professional standards domain. A number of proficiency indicators were selected from *Principals*, then adapted, and approved by consensus of the task force members.

The task force may send out questionnaires to professional educators both in the district and throughout the state. These may incorporate Likert-scaled questions and allow respondents to name other job categories deemed equally important but overlooked by the task force. Jordan School District sent questionnaires to superintendents, district administrator supervisors, district evaluation directors, and an elementary, a junior high and a senior high principal from each district in the state. In addition, the Jordan District sent questionnaires to all its full-time employees, 100 parents of elementary students, 100 parents of junior high students, and 100 parents of senior high students. Inviting participation of such diverse constituencies evidences the task force's effort to solicit and consider comments from the relevant stakeholders of quality school administration.

Finally, the evaluation policy and a description of its validation procedures is reviewed by various evaluation experts. Their participation provides yet another step in the iterative process of creating, revising, implementing and adjusting a district's administrator evaluation policy. Each time the policy cycles through refinement by the taskforce, stakeholders, and external reviewers, the district strengthens its case that its administrator evaluation policy is legally sound.

c. What is meant by "criterion-related validation" evidence and how can a district create relevant evidence for its administrator evaluation policy?

Criterion-related evidence is gathered by examining the correspondence between two measures of some behavior or skill. In the case of administrator evaluation, this can be done by evaluating a representative group of administrators under the proposed new evaluation policy as a pilot study and then comparing its results with a peer nomination of administrators who are considered to be excellent. If the results highly correlate, then it is likely that the new instrument is accurately measuring attributes that peers feel are important. This avenue of evidence collection is supported by *PES*,⁵³ as well as the Utah State Office of Education.⁵⁴ If included in a validation study, it should be completed after the task force has responded to questionnaire comments and has agreed upon standards defining acceptable administrator performance.

53. JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, *supra* note 20, at 99.

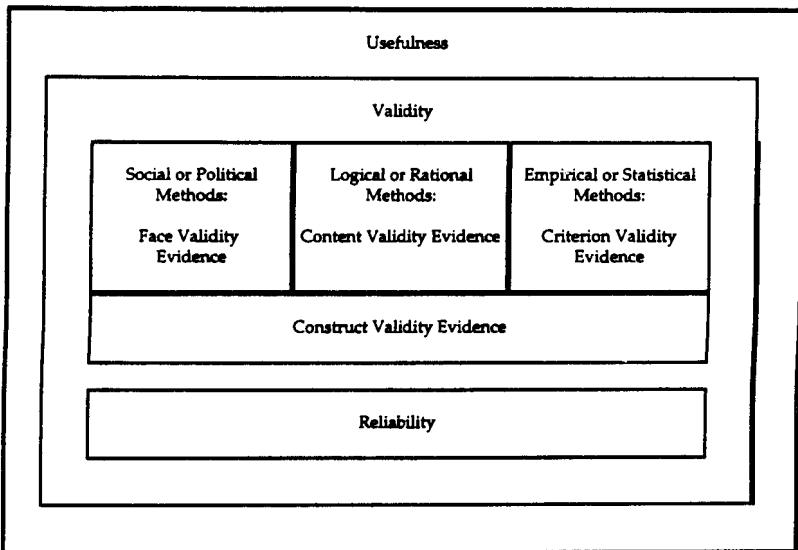
54. Letter from Dr. Dave Nelson, *supra* note 7, at 1.

d. What is meant by "construct validation" evidence and how can the district create relevant evidence for its administrator evaluation policy?

Constructs are abstractions that do not exist outside of the human mind, such as "self-esteem," "creativity" or "scholastic aptitude." "[T]hey are not directly measurable."⁵⁵ Evidence gathered to prove face, content, and construct validity also helps establish that a newly created instrument does indeed measure what it intends to measure.

The following diagram clarifies the kinds of validity evidence and their relationships. To be truly useful, evaluation policies must be valid and reliable. Face validity evidence is gathered through documenting the social and political acceptability of the evaluation policy. Content validity evidence is demonstrated by logical and rational explanation of the administrator evaluation policy's components. Criterion validity evidence is demonstrated through empirical or statistical analyses which compare evaluations under the new policy and other generally accepted measures of performance. Evidence of face, content and criterion validity is also used to demonstrate that the construct being assessed through the evaluation policy is well understood. This is denoted by the dotted line between these. Reliability is a part of validity which is also required of an evaluation policy that is useful.

RELATING USEFULNESS, VALIDITY EVIDENCE, AND RELIABILITY



55. B. R. WORTHEN, ET AL, *supra* note 42, at 184.

C. Does the District's Administrator Evaluation Policy Meet Due Process and Statutory Procedural Standards?

The provisions of the statute are designed to meet the demands of due process under the Utah State Constitution⁵⁶ and the United States Constitution.⁵⁷ Despite the vaguaries surrounding definition of "due process," most scholars agree that due process entails at least three elements in administrative settings: notice, hearing and appeal.⁵⁸ The Educator Evaluation Act of Utah also specifies numerous additional procedural steps which go well beyond basic due process expectations.

1. How does the district policy meet the notice element of due process while conforming to the statute?

The first element of due process is notice. Under the statute, the evaluatee must receive personal notice and "a copy of the evaluation instrument, if an instrument is to be used" at least fifteen days before first being evaluated.⁵⁹

A less transparent element of notice is the requirement that stated procedures be followed. The idea is that people should know what to expect when their "life, liberty or property" hang in the balance.⁶⁰ Thus, even though due process clauses do not require that evaluations be done on Tuesdays only—for instance, it would be a violation of due process to evaluate on other than a Tuesday if the district policy specifies that Tuesdays are the only day on which they may be done.

2. How does the district policy meet the hearing element of due process while conforming to the statute?

The second element of due process is a fair hearing. Specifically, within [fifteen] days after the completed evaluation process the evaluation in writing shall be discussed with the educator. Following any revisions made after the discussion, a copy of the evaluation shall be filed in the educator's personnel file together with any related reports or documents. A copy of the evaluation and attachments shall be given to the educator.⁶¹

56. UTAH CONST., art. I, § 7.

57. U.S. CONST., amend. XIV, § 1.

58. See MAX RADIN, RADIN LAW DICTIONARY 8 and 103 (Lawrence G. Greene ed., 2d ed. 1970) ("Administrative Law" and "Due Process of Law" entries).

59. UTAH CODE ANN. § 53A-10-106(2) (1990).

60. UTAH CONST., art. I, § 7, (See "In General" Comment).

61. UTAH CODE ANN. § 53A-10-106(7) (1990).

Thus, the evaluatee and evaluator hold a post-evaluation conference after which the evaluatee may amend responses to the written report. The evaluatee “may make a written response to all or any part of the evaluation [that is thereafter] . . . attached to the evaluation.”⁶²

3. How does the district policy meet the appeal element of due process while conforming to the statute?

The third element of due process is the opportunity of appeal. The right to appeal and the manner thereof are specified in the statute.⁶³ The evaluatee has thirty days in which to request a review of an evaluation with which he or she is dissatisfied.⁶⁴ The review may not be done by an employee of the district unless the evaluatee and superintendent agree to such a method of review.⁶⁵

4. What additional procedural requirements must the school district policy include to conform to the statute?

The statute is particular about a number of other procedural elements of educator evaluation. For example, the evaluation must be completed by the principal, his or her designee, the administrator’s immediate supervisor or as specified in the evaluation policy.⁶⁶ It must employ “several types of evaluation *and* evidence, such as self-evaluation, student evaluation, peer evaluation, or systematic observations.”⁶⁷ The evaluation policy also must provide clear, written descriptions of “deficiencies, the available resources for improvement, and a recommended course of action that will improve the educator’s performance.”⁶⁸ For educators who are not yet entitled to rely on continued employment, the district must provide mentors who, when possible, “perform[] substantially the same duties as the provisional educator and [have] at least three years of educational experience.”⁶⁹ Final evaluations of administrators who are judged to be inadequate must be completed at least sixty days before the end of the contract year.⁷⁰

Inevitably, some errors may be made while evaluations are being performed. If the errors in an evaluation are no more than technical, there is a legal argument to preserve the legitimacy of actions taken on the basis of

62. UTAH CODE ANN. § 53A-10-106(5) (1990).

63. UTAH CONST., art. I, § 7, (*See* “In General” and “Construction and Interpretation” Comments).

64. UTAH CODE ANN. § 53A-10-110(1) (1990).

65. UTAH CODE ANN. § 53A-10-110(2) (1990).

66. UTAH CODE ANN. § 53A-10-106(1) (1990).

67. UTAH CODE ANN. § 53A-10-106(4) (1990) (*emphasis added*).

68. UTAH CODE ANN. § 53A-10-107(1) (1990).

69. UTAH CODE ANN. § 53A-10-108(1) and (2) (1990).

70. UTAH CODE ANN. § 53A-10-109(1) (1988).

the evaluation. The concept is "harmless error."⁷¹ A breach of protocol which does not actually cause injury is forgiven by the courts. For example, a misspelled name of an administrator evaluated as deficient will not be an invalidating error unless the resulting confusion actually causes a delay in remediation or frustrates notice. To facilitate this argument, it is advisable to include a provision in the evaluation policy stating that errors which do not result in actual injury and were made in a good faith attempt to follow procedure will not nullify the result of an evaluation.

III. CONCLUSION

This article helps to clarify what should be done to ensure that actions taken as the result of administrator evaluations will be upheld in the face of legal challenges. These comments focus on the various aspects of gathering validity evidence as a district task force develops the instruments, procedures, and inservice for an administrator evaluation policy.

Three topics related to such an effort are addressed in the body of the article: 1) the proper steps to follow in creating the administrator evaluation policy, 2) the kinds of validation evidence a district should incorporate into such an effort and how to gather such evidence, and 3) the proper procedural elements to include in administrator evaluation policies to conform to the applicable legal standards. The questions are constructed and answered in an attempt to help school districts, administrators, consultants and stakeholders in quality education focus their efforts to create sound school administrator evaluation policy.

Although this article does not focus on adjusting valid administrator policy, two additional points are important to consider for districts that may be looking at the big picture. First, maintaining the validity of evaluations under an appropriately created policy will largely depend on providing future evaluators with adequate inservice on the use of the evaluation system. Once this has been done, some type of certification and recertification of evaluators is advisable so that the district can show that it has made good faith efforts to ensure that evaluations are done accurately and consistently. Second, it is to be expected that an entirely appropriate administrator evaluation system will become outdated over time by changing practices, roles and technologies. Adoption of a periodic review provision at the time the policy is approved by the district school board will help the district become aware of problems or errors and remedy these.

Throughout the article, reference has been made to how Jordan School District has applied the Utah Educator Evaluation Act and other applicable

71. WILLIAM P. STASKY, *WEST'S LEGAL THESAURUS/Dictionary: A RESOURCE FOR THE WRITER AND THE COMPUTER RESEARCHER* 364 (1985).

law in creating its new administrator evaluation system. This information is shared as a potential example to other districts in and out of Utah as they seek to access the literature and studies that validate school administrator evaluation policies. The appendix gives a more detailed account of the validation study results and clarifies Jordan's final content selections for more interested readers. Together, the article and appendix may serve as a helpful resource for similar efforts by other districts to bring their administrator evaluation practice into closer harmony with applicable law, possibly reducing legal exposure due to use of evaluation systems that may not be legitimated if grievances arise under them.

APPENDIX A: THE UTAH EDUCATOR EVALUATION ACT

53A-10-101. LEGISLATIVE FINDINGS.

The Legislature recognizes that the quality of public education can be improved and enhanced by providing for systematic, fair, and competent evaluation of public educators and remediation of those whose performance is inadequate. The desired purposes of evaluation are to allow the educator and the school district to promote the professional growth of the teacher, to identify and encourage teacher behaviors which contribute to student progress, to identify teachers according to their abilities, and to improve the education system.

53A-10-102. DEFINITIONS.

As used in this chapter:

- (1) "Career educator" means a certified employee entitled to rely upon continued employment under the policies of a local school board.
- (2) "Educator" means any individual, except the superintendent, employed by a school district who is required to hold a professional certificate issued by the State Board of Education. Educator does not include individuals who work less than three hours per day or who are hired for less than half of a school year.
- (3) "Probationary educator" means any educator employed by a school district who, under local school board policy, has been advised by the district that his performance is inadequate.
- (4) "Provisional educator" means any educator employed by a school district who has not achieved status as a career educator within the school district.

53A-10-103. ESTABLISHMENT OF EDUCATOR EVALUATION PROGRAM — JOINT COMMITTEE.

- (1) Each local school board shall develop an evaluation program in consultation with its educators through appointment of a joint committee.
- (2) The joint committee shall be comprised of an equal number of classroom teachers and administrators appointed by the board.
- (3) A board may appoint members of the joint committee from a list of nominees voted on by classroom teachers in a nomination election and from a list of nominees voted on by the administrators in a nomination election.
- (4) The evaluation program developed by the joint committee must comply with the requirements of Section 53A-10-106.

53A-10-104. FREQUENCY OF EVALUATIONS.

A local school board shall provide for the evaluation of its provisional and probationary educators at least twice each school year.

53A-10-105. EVALUATION ORIENTATION.

- (1) The principal of each school shall orient all educators assigned to the school concerning the school board's educator evaluation program, including the purpose of the evaluations and the method used to evaluate.
- (2) Evaluations may not occur prior to the orientation by the principal.

53A-10-106. COMPONENTS OF EDUCATOR EVALUATION PROGRAM — EVALUATOR — NOTICE — CRITERIA — RESPONSE.

Any educator evaluation program adopted by a local school board in consultation with a committee shall provide the following:

- (1) unless otherwise provided in the adopted program, the principal, the principal's designee, or the educator's immediate supervisor shall perform the educator evaluation;
- (2) personal notice to the educator of the evaluation process at least 15 days prior to the first evaluation and receipt of a copy of the evaluation instrument, if an instrument is to be used;
- (3) a reasonable number of observation periods for any evaluation to insure adequate opportunity for evaluation;
- (4) the use of several types of evaluation and evidence, such as self-evaluation, student evaluation, peer evaluation, or systematic observations;
- (5) that the educator may make a written response to all or any part of the evaluation and that the response will be attached to the evaluation;
- (6) a reliable and valid evaluation consistent with generally accepted professional standards for personnel evaluation systems; and
- (7) within 15 days after the completed evaluation process the evaluation in writing shall be discussed with the educator. Following any revisions made after the discussion, a copy of the evaluation shall be filed in the educator's personnel file together with any related reports or documents. A copy of the evaluation and attachments shall be given to the educator.

53A-10-107. DEFICIENCIES — REMEDIATION.

- (1) An educator whose performance is inadequate or in need of improvement shall be provided with a written document clearly identifying deficiencies, the available resources for improvement, and a recommended course of action that will improve the educator's performance.
- (2) The district shall provide the educator with reasonable assistance to improve performance.
- (3) An educator is responsible for improving performance by using the resources identified by the school district and demonstrating acceptable levels of improvement in the designated areas of deficiencies.

53A-10-108. CONSULTING EDUCATOR FOR PROVISIONAL EDUCATOR.

- (1) The principal or immediate supervisor of a provisional educator shall assign a consulting educator to the provisional educator.
- (2) If possible, the consulting educator shall be a career educator who performs substantially the same duties as the provisional educator and has at least three years of educational experience.
- (3) The consulting educator shall assist the provisional educator to become informed about the teaching profession and school system, but may not serve as an evaluator of the provisional teacher.

53A-10-109. FINAL EVALUATION.

- (1) At least 60 days prior to the end of the contract school year, the principal, immediate supervisor, or appointed evaluator of an educator whose performance has been determined to be inadequate or in need of improvement, shall complete all written evaluations and recommendations regarding the educator evaluated during the contract school year.
- (2) The final evaluation shall contain only data previously considered and discussed with the individual educator as required in Section 53A-10-106.
- (3) Nothing in this section prevents a school district from performing supplementary evaluation for good cause after the issuance of the final evaluation.

53A-10-110. REVIEW OF EVALUATION — TIME LIMIT ON REQUEST.

- (1) An educator who is not satisfied with an evaluation has 30 days after receiving the written evaluation to request a review of the evaluation.
- (2) If a review is requested, the district superintendent or the superintendent's designee shall appoint a person, not an employee of the district, who has expertise in teacher or personnel evaluation to review and make recommendations to the superintendent regarding the teacher's evaluation.
- (3) Nothing in this section prevents the teacher and district superintendent or the superintendent's designee from agreeing to another method of review.

53A-10-111. ADDITIONAL COMPENSATION FOR SERVICES.

The district may compensate a person employed as a consulting educator or participant in the evaluation for those services, in addition to the person's regular salary, if additional time is required in the evaluation process.

Amended by Chapter 78, 1990 General Session

APPENDIX B: SUMMARY RESULTS OF A STATE-WIDE AND DISTRICT VALIDATION STUDY*

Outline of the Study

I. Overview of a school administrator evaluation policy validation study

II. Mean responses for task force-proposed competency domains

A. Aggregate mean responses

1. Jordan stakeholder aggregate mean responses
2. Utah administrator aggregate mean responses

B. Mean responses and differences partitioned by respondent category

1. Jordan stakeholder mean responses partitioned by respondent category
2. Utah administrator mean responses partitioned by respondent category
3. Respondent stakeholder category differences
4. Administrator job title differences

C. Mean responses and differences partitioned by administrator contact

1. Jordan stakeholder mean responses partitioned by administrator contact
2. Utah administrator mean responses partitioned by administrator contact
3. Administrator contact differences

D. Mean responses and differences partitioned by school level contact

1. Jordan stakeholder mean responses partitioned by school level contact
2. Utah administrator mean responses partitioned by school level contact
3. School level differences

E. Mean responses and differences partitioned by years of experience

1. Jordan stakeholder mean responses partitioned by years of experience
2. Utah administrator mean responses by years of experience

* See Generally STEVE BALDRIDGE, AN INQUIRY INTO THE CONTENT VALIDITY OF PUBLIC SCHOOL ADMINISTRATOR EVALUATION DOMAINS AND PROFICIENCY INDICATORS 97-113, 191-197, and 365 (1996).

3. Years of school employment differences

F. Mean responses and differences partitioned by respondent gender

1. Jordan stakeholder mean responses partitioned by respondent gender
2. Utah administrator mean responses partitioned by respondent gender
3. Gender differences

G. Mean responses and differences partitioned by district urbanization

1. Utah administrator mean responses partitioned by district urbanization
2. District urbanization level differences

III. Cummulative rankings of competency domains among partitioned groups

IV. Conclusions and recommendations

A. Priorities for retention of indicators

1. Mean
2. Goodness-of-measure
3. Goodness-of-fit

I. Overview of a school administrator evaluation validation study

The Jordan Administrator Evaluation System was created by a task force composed of parents, classified employees, teachers and school-level administrators representing the K-12 range, and district-level administrators, with the assistance of professional consultants. The task force first addressed the system's content, then its procedures. Content validation questionnaires were sent to a sample of administrators throughout the state, all full-time employees of the district, and a representative sample of parents in the district. The content validation study results are summarized below. The procedures were then piloted with about 10% of the administrators in the district, then revised. Finally the system was approved by the district educator evaluation committee and the school board.

II. Mean responses for task force-proposed competency domains

The first major question to be answered in the validation study was: What competency domains on which Jordan School District administrators could be evaluated are most essential to administrator success? The stakeholder taskforce narrowed and revised the National Policy Board for Educational Administration list from twenty-one to seven domains. To gather validating or invalidating information on their selection, the long survey sent to

the various stakeholder groups in Jordan School District, and the short survey sent to administrators throughout the State of Utah, both asked respondents to rate the seven taskforce-created domains on the following five-point Likert scale: Without proficiency in this domain a school administrator would fail (Strongly disagree/Disagree/No opinion/Agree/Strongly agree). Aggregated and partitioned mean responses for the domain questions are described below. Responses from outside the district served as a comparison check to ensure that alternative perspectives were considered.

To begin answering the question of the extent to which respondents agreed on the importance of the domains to success of school administrators, one should consider the mean ratings (0 to 4) of competency domains. First those of Jordan stakeholders were considered all together, then those of school administrators across the state. Because all of the domains received favorable aggregate mean ratings in Jordan (2.94 to 3.37) and throughout the state (3.13 to 3.55), none could be summarily eliminated—and the taskforce’s hard work in preparing the content of the new evaluation system was validated. Even when mean ratings of the various respondent groups were parsed out, the scores continued to be high across the board (2.75 to 4.00).

By partitioning respondents in various ways one can verify whether some groups’ opinions are being unheeded because they are “drowned out” by the other groups’ responses. Subgroups that were examined include: the respondents’ stakeholder group, the type of administrator with which respondents have most contact, the school level with which respondents have most contact, and respondent gender. Stakeholder groups were selected because those with differing vantage points on school administration may see the priorities of the job differently. The type of administrator with whom respondents have most contact was likewise selected because competence in a domain might be especially crucial for administrators in one position but not for those in another. Finally, school level was selected because the competencies required of administrators in one level of schools may be different than those required in administration of other levels. Even when partitioned in these various ways, all of the different Jordan respondent groups rated the domains between 2.76 and 3.80. The ratings by groups in the Utah administrator questionnaire were more variable, ranging from 2.50 to 4.00.

A. *Aggregate mean responses*

Aggregate mean responses give a first, “broad brush” picture of how crucial the seven taskforce-created domains are. If a domain scored an aggregate mean of 2.0 (No opinion) or lower, elimination or revision would have been strongly indicated because the respondents would have been

saying that competence in the domain is unconnected to administrator failure. The weakness of looking only at aggregate means is the potential of missing important disagreements regarding the value of domains by adding together the scores given by differently situated respondents. By providing information about both aggregate and partitioned responses, this report allows closer scrutiny of the level of acceptance the domains and indicators received from all parties affected by administrator evaluations in Jordan School District.

1. Jordan stakeholder aggregate mean responses. The aggregate mean responses from the long survey—reflecting the opinion of 3188 school administrator evaluation stakeholders in Jordan School District—rated six of the domains between “Strongly agree” and “Agree” on the Likert scale; the seventh rated just under “Agree” at 2.94. As in the pilot, the Interpersonal Skills domain rated highest, this time at 3.37. Professional Standards and Judgment, which had also tied for first in the pilot, rated next highest at 3.35 and 3.27 respectively. Next came Leadership at 3.23 and Staff Development at 3.18, again in the same order as in the pilot. Contrary to the pilot, Resource Management, which scored a 3.17, nudged ahead of the lowest rated domain—Instructional Program. Table 1, containing the aggregate mean responses to the domain questions from the pilot, the long and the short questionnaires, is found below.

2. Utah administrator aggregate mean responses. The aggregate mean responses from the short survey, reflecting the opinion of 182 school administrators throughout the State of Utah, rated all seven domains between “Strongly agree” and “Agree” on the Likert scale. As in the pilot and the long questionnaire, the Interpersonal Skills domain rated highest, this time at 3.55. In contrast, Judgment, Leadership and then Professional Standards rated next highest at 3.47, 3.46 and 3.33 respectively. Next came Resource Management at 3.24. Similar to the pilot and the long questionnaire, Resource Management, Instructional Program, and Staff Development came in the last three slots; however, in the short questionnaire, Staff Development came in last at 3.13 and Instructional Program came in penultimate at 3.17. In general, administrators in the statewide questionnaire rated the domains higher than did the Jordan stakeholders in aggregate but about on par with Jordan administrators as shown in the next section. The aggregate mean responses to the domain questions from the pilot, the long and the short questionnaires are displayed in Table 1 below.

Table 1: Aggregate Domain Mean Ratings and Ranks

Q#	PMn	Rk	JMn	Rk	UMn	Rk	Domain
3.	3.78	4	3.23	4	3.46	3	LEADERSHIP: The school leader develops a shared strategic vision and facilitates the realization of the vision with staff, students, parents, and the community.
4.	3.85	1	3.27	3	3.47	2	JUDGMENT: The school leader makes wise decisions in a timely fashion based on the best available information.
5.	3.53	6	2.94	7	3.17	6	INSTRUCTIONAL PROGRAM: The school leader knows the school curriculum and provides leadership in curriculum planning, instruction, and evaluation.
6.	3.63	5	3.18	5	3.13	7	STAFF DEVELOPMENT: The school leader provides encouragement, support, and opportunities for staff to develop and strengthen professional knowledge and skills.
7.	3.47	7	3.17	6	3.24	5	RESOURCE MANAGEMENT: The school leader effectively utilizes available resources to address basic needs and achieve essential educational goals in an efficient and ethical manner.
8.	3.85	1	3.37	1	3.55	1	INTERPERSONAL SKILLS: The school leader treats others in a professional manner with respect and sensitivity, and facilitates a caring and motivating environment.
9.	3.85	1	3.35	2	3.33	4	PROFESSIONAL STANDARDS: The school leader knows and implements legal standards and policies, and is a role model of a professional educator.
Key: Q# = questionnaire question number; PMn = pilot mean;							
JMn = Jordan mean; UMn = Utah administrators mean; Rk = rank							

B. Mean responses and differences partitioned by respondent category

1. Jordan stakeholder mean responses partitioned by respondent category.

School board members consistently rated all domains—and indicators—higher than any other group in the Jordan questionnaire. Administrators consistently rated all domains second highest. Certificated employees usually came next, then parents and classified employees, but these three really switch order quite a bit. Despite these overall differences among the groups, their rankings are surprisingly similar (2.92 to 3.80).

The rankings of the group rates are also quite similar. As examples, administrators, certificated employees and parents gave Interpersonal Skills their highest rating while classified employees and school board members chose

Professional Standards. Administrators, parents and school board members gave their second highest rating to Judgment, which was also rated third highest by certificated and classified employees. All but school board members gave Leadership their fourth highest rating. Three groups gave Staff Development their fifth highest rating and the other two groups put it at fourth and sixth. Three groups rated Resource Management the sixth most crucial domain. All groups but parents gave their lowest rating to Instructional Program—which parents gave to Resource Management. In all cases but Interpersonal Skills, the ranks of the ratings by each group were within one above or below the other groups (e.g., 1, 2, 3 or 5, 6, 7).

By adding up the rankings of the domains by each group, one gets a sense of how strongly favored the domains are among the groups. The highest possible total rating would be 5 (5 x 1), the lowest would be 35 (5 x 7). In these terms Interpersonal Skills and Professional Standards are strongly supported by the groups, with totals of 10, and Instructional Program receives the weakest support, with a total of 34. All of the domain mean ratings and ranks partitioned by respondent category are found below in Table 2.

Table 2: Jordan Domain Mean Ratings and Ranks

Partitioned by Respondent Category

Q#	Adm	R	Cert	Rk	Clas	R	Prnt	Rk	SchB	R	Tot	Domain
		k				k				k		
3	3.34	4	3.33	4	3.05	4	3.12	4	3.65	3	19	LEADERSHIP
4	3.48	2	3.36	3	3.10	3	3.15	2	3.70	2	12	JUDGMENT
5	3.00	7	2.92	7	2.98	7	2.99	6	3.15	7	34	INSTRUCTIONAL PROGRAM STAFF
6	3.17	6	3.27	5	3.03	5	3.01	5	3.60	4	25	DEVELOPMENT RESOURCE
7	3.31	5	3.26	6	3.01	6	2.97	7	3.45	6	30	MANAGEMENT INTERPERSONAL
8	3.53	1	3.50	1	3.13	2	3.23	1	3.50	5	10	SKILLS PROFESSIONAL
9	3.45	3	3.46	2	3.15	1	3.14	3	3.80	1	10	STANDARDS

Key: Q# = questionnaire question number; Adm = administrator means;
 Cert = certificated employee means; Clas = classified employee means;
 Prnt = parent means; SchB = school board member means; Rk = rank; Tot = total

2. *Utah administrator mean responses partitioned by respondent category.* All Utah administrator groups rated the domains between 3.06 and 3.84, slightly higher than Jordan questionnaire respondent groups. Among Utah school administrators, those who evaluate principals consistently rated all domains higher than any other group. Superintendents were the consistently second highest rating group. School board members were usually the group that rated each domain lowest. Again, as in the Jordan Stakeholder Questionnaire, these differences are probably no more important than the consistent trends among the groups regarding the domains to which they gave their highest and lowest ratings.

Principals, research directors and school board members all gave their highest rating to the Interpersonal Skills domain. Leadership received the highest rating of superintendents and those who evaluate principals—who both gave their second highest rating to Interpersonal Skills. Although Instructional Program and Resource Management received similar total rankings, it was for different reasons. There was disagreement among the groups on how highly to rate Instructional Program with rankings from second (research directors) to seventh (principals). Resource Management, however, received only 4s, 5s and a 6. Staff Development came out clearly on the low end with the lowest rating from all groups except principals, who placed it second lowest.

Table 3: Utah School Administrators Domain Mean Ratings
and Ranks Partitioned by Respondent Category

Q#	Prin	R k	Rch	Rk	Eval	R k	Supt	Rk	SchB	R k	Tot	Domain
3	3.31	3	3.29	4	3.84	1	3.84	1	3.20	3	12	LEADERSHIP
4	3.46	2	3.29	4	3.74	3	3.65	3	3.28	1	13	JUDGMENT
5	3.06	7	3.36	2	3.47	5	3.38	6	3.12	6	26	INSTRUCTIONAL
6	3.10	6	3.07	7	3.37	7	3.27	7	3.08	7	34	PROGRAM STAFF
7	3.15	5	3.21	6	3.47	5	3.41	4	3.16	4	24	DEVELOPMENT RESOURCE
8	3.54	1	3.50	1	3.79	2	3.70	2	3.28	1	7	MANAGEMENT INTERPERSONAL
9	3.29	4	3.36	2	3.68	4	3.41	4	3.16	4	18	SKILLS PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number; Prin = principal means;
Rch = research director means; Eval = evaluation/testing director means;
Supt = superintendent means; SchB = school board member means; Rk = rank; Tot = total

3. *Respondent stakeholder category differences.* One difference between respondents to the Utah school administrators questionnaire and the Jordan questionnaire is that there was less agreement among Utah administrator groups on how highly to rate domains. Ranks of the ratings varied widely from a narrower range of all 1s and 2s (Interpersonal Skills), 4s and a 2 (Professional Standards), and 7s and a 6 (Staff Development)—to wider ranges of a 2, a 5, 6s and a 7 (Instructional Program), and a 1, a 2, 3s and a 4 (Judgment). By adding up the rankings of the domains by each group, one can get an sense of how strongly favored the domains are among the groups. The highest possible total rating would be 5 (5 x 1), the lowest would be 35 (5 x 7). In these terms Interpersonal Skills is strongly supported by the groups, with a total of 7, and Staff Development receives the weakest support, with a total of 34. All of the Utah school administrators domain mean ratings and ranks partitioned by respondent category are found above in Table 3.

In the Jordan Questionnaire, administrators consistently rated all domains second highest, after school board members. Certificated employees usually came next, then parents and classified employees. Despite the overall differences among the groups, their rankings are surprisingly similar (2.92 to 3.80). Differences in opinion regarding the usefulness of indicators falling into the four factor groups were significant along the facet of stakeholder category. Classified employees scored factor Group A indicators higher than administrators or certificated employees, while parents scored factor Group A indicators higher than certificated employees. Administrators scored factor Group D indicators higher than certificated employees, classified employees or parents.

Within the demographic partitions, 90% of classified employees engaged in mild acquiescence (the high), while only 70% of parents did so (the low). On the low end, only 15% of certificated employees engaged in excessive acquiescence. By stakeholder category, Parents were the least likely to acquiesce at 70%, while Classified employees were the most likely at 90%. Among stakeholder groups, excessive acquiescence ranged from 15% of certificated employees up to 25% of classified employees. Among stakeholder categories, it is noteworthy that the skew is equally apparent among administrators and teachers (93%) who engaged in mild acquiescent behavior. Positive skew is highest among school board members (100%), and least apparent among those who failed to identify their stakeholder category (40%).

4. *Administrator job title differences.* All Utah administrator groups rated the domains between 3.06 and 3.84, slightly higher than Jordan questionnaire respondent groups. Among Utah school administrators, those who have closest contact with principals almost always rated the domains high-

est (3.26 to 3.72) while those whose closest contact is with assistant principals rated all domains lower (2.85 to 3.56). Superintendents were the consistently second highest rating group. School board members were usually the group that rated each domain lowest. Again, as in the Jordan Stakeholder Questionnaire, these differences are probably no more important than the consistent trends among the groups regarding the domains to which they gave their highest and lowest ratings. The results from the Utah administrator questionnaire are less clear than those from Jordan. As in Jordan, Instructional Program was noticeably lower than any other domain for assistant principals. But unlike Jordan, the ranks of rates are not similar between the two groups. Still, it is clear that all the domains fit principals better than they fit assistant principals.

C. Mean responses and differences partitioned by administrator contact

1. Jordan stakeholder mean responses partitioned by administrator contact. The hypothesis that the type of administrator with whom respondents have most contact might lead people to rate the domains differently was borne out by the questionnaire responses. Those who have closest contact with principals rated the domains highest (3.04 to 3.45) while those whose closest contact is with assistant principals rated all domains lower (2.74 to 3.37) while retaining the same rank order. Instructional Program was noticeably lower than any other domain for assistant principals. In other words, although all the domains fit both positions well, the fit is clearly better for principals.

Those whose closest contact is with district office administrators rated Judgment, Instructional Program, Staff Development, and Resource Management between principals and assistant principals. They rated Leadership, Interpersonal Skills and Professional Standards below assistant principals. Those whose contact is with some other type of school administrator rated all domains approximately a quarter point lower than any other group (2.76 to 2.89), the exception being Instructional Program for assistant principals mentioned earlier. Unfortunately, they did not specify with what type of administrator they have contact. Even so, their order of preference follows those whose closest contact is with the principal or assistant principal.

Similar to the results with other groupings, Interpersonal Skills, Professional Standards and Judgment receive universally higher ranked rates than Instructional Program, Staff Development and Resource Management. By adding up the rankings of the domains by each group, one can get an sense of how strongly favored the domains are among the groups. The highest possible total rating would be 4 (4 x 1), the lowest would be

28 (4 x 7). In these terms Interpersonal Skills and Professional Standards are most strongly supported by the groups, with totals of 7, and Instructional Program receives the weakest support, with a total of 26. All of the Jordan domain mean ratings and ranks partitioned by administrator contact are found in Table 4 below.

Table 4: Jordan Domain Mean Ratings and Ranks
Partitioned by Administrator Contact

Q#	APrn	R	Prin	Rk	DOF	R	Othr	Rk	Tot	Domain
3	3.16	4	3.32	4	3.13	5	2.81	4	17	LEADERSHIP
4	3.26	3	3.33	3	3.28	1	2.85	3	10	JUDGMENT
5	2.74	7	3.04	7	3.02	7	2.79	5	26	INSTRUCTIONAL PROGRAM
6	3.04	5	3.29	5	3.07	6	2.79	5	21	STAFF DEVELOPMENT
7	3.04	5	3.27	6	3.17	3	2.76	7	21	RESOURCE MANAGEMENT
8	3.37	1	3.45	1	3.15	4	2.89	1	7	INTERPERSONAL SKILLS
9	3.33	2	3.43	2	3.20	2	2.89	1	7	PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number; APrn = assistant principal means;
Prin = principal means; DOF = district office administrator means;
Othr = other administrator means; Rk = rank; Tot = total

2. *Utah administrator mean responses partitioned by administrator contact.* The results from the Utah administrator questionnaire are less clear than those from Jordan. Nonetheless, the hypothesis that the type of administrator with whom respondents have most contact might lead people to rate the domains differently was borne out by the questionnaire responses. Those who have closest contact with principals almost always rated the domains highest (3.26 to 3.72) while those whose closest contact is with assistant principals rated all domains lower (2.85 to 3.56). As in Jordan, Instructional Program was noticeably lower than any other domain for assistant principals. But unlike Jordan, the rankings of mean responses are not similar between the two groups. Still, it is clear that all the domains fit principals better than they fit assistant principals.

Those whose closest contact is with district office administrators rated Leadership, Judgment, Instructional Program, and Staff Development between principals and assistant principals. They rated Interpersonal Skills and Professional Standards below assistant principals, and Resource Management slightly above principals. Those whose contact is with some other type of school administrator rated domains most erratically, registering both the highest and lowest rates by any groups among the partitioned means (2.50 to 4.00). Unfortunately, there is no way to know what jobs these administrators fulfill; fortunately, very few people—only two—had most contact with “other” administrators.

Table 5: Utah Administrator Domain Mean Ratings
and Ranks Partitioned by Administrator Contact

Q#	APrn	R	Prin	Rk	DOF	R	Othr	Rk	Tot	Domain
		k				k				
3	3.27	4	3.72	1	3.37	3	3.00	3	11	LEADERSHIP
4	3.30	3	3.61	3	3.46	2	3.00	3	11	JUDGMENT
5	2.85	7	3.37	5	3.21	6	2.50	6	24	INSTRUCTIONAL PROGRAM
6	3.00	6	3.26	7	3.14	7	2.50	6	26	STAFF DEVELOPMENT
7	3.16	5	3.26	7	3.27	4	3.00	3	19	RESOURCE MANAGEMENT
8	3.56	1	3.65	2	3.48	1	4.00	1	5	INTERPERSONAL SKILLS
9	3.34	2	3.46	4	3.26	5	3.50	2	13	PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number; APrn = assistant principal;
Prin = principal means; DOF = district office administrator means;
Othr = other administrator means; Rk = rank; Tot = total

By adding up the rankings of the domains by each group, one can get a sense of how strongly favored the domains are among the groups. The highest possible total rating would be 4 (4 x 1), the lowest would be 28 (4 x 7). In these terms Interpersonal Skills is most strongly supported by the groups, with a total of 5, while Instructional Program and Staff Development receive the weakest support, with totals of 24 and 26. All the Utah administrator domain mean ratings and ranks partitioned by administrator contact are found above in Table 5.

3. Administrator contact differences. The hypothesis that the type of administrator with whom respondents have most contact might lead people to rate the domains differently was borne out by the Jordan Questionnaire responses. Those who have closest contact with principals rated the domains highest (3.04 to 3.45) while those whose closest contact is with assistant principals rated all domains lower (2.74 to 3.37) while retaining the same rank order. Instructional Program was noticeably lower than any other domain for assistant principals. In other words, although all the domains fit both positions well, the fit is clearly better for principals. In the statewide questionnaire, those whose closest contact is with district office administrators rated Leadership, Judgment, Instructional Program, and Staff Development between principals and assistant principals. They rated Interpersonal Skills and Professional Standards below assistant principals, and Resource Management slightly above principals.

Those whose closest contact is with district office administrators rated Judgment, Instructional Program, Staff Development, and Resource Management between principals and assistant principals. They rated Leadership, Interpersonal Skills and Professional Standards below assistant principals. Those whose contact is with some other type of school administra-

tor rated all domains approximately a quarter point lower than any other group (2.76 to 2.89), the exception being Instructional Program for assistant principals mentioned earlier; however, their order of preference follows those whose closest contact is with the principal or assistant principal. The groups whose perspectives are most unique are people whose closest contact is with district office administrators and other administrators. Both groups gave high or low ratings to two indicators that appear nowhere else in all the partitioned analyses. These were 12 (high) and 43 (low) for those associated with district office administrators, and 24 (low) and 39 (low) for those whose closest contact is with other administrators.

Differences in opinion regarding the usefulness of indicators in factor groups appear along the facet of the type of administrator with which respondents have most contact. Those who have contact predominantly with assistant principals score factor Group A indicators (Leadership, Judgment, Interpersonal Skills) lower than those who have contact predominantly with district office administrators. But they also score factor Group D indicators (Resource Management) higher than those who work predominantly with principals or district office administrators.

D. Mean responses and differences partitioned by school level contact

1. Jordan stakeholder mean responses partitioned by school level of contact. School level was selected because the competencies required of administrators at one school level may differ from those required of administrators at other school levels. In Jordan, those whose closest contact is with elementary schools or district offices found the seven domains to be the most essential to avoiding administrator failure. Those whose contact is with high schools found the domains least essential; however, even there the lowest rating was a 2.68 for Instructional Program.

Regardless of school level, respondents felt that ability in the area of Instructional Program was least essential to administrator competence. The difference between respondents with closest contact to special schools and all other groups on the Interpersonal Skills Domain is another striking find. These respondents rated it second least essential to competence while all other groups rated it highest or second highest. Of note is also the three-way tie at special schools among Judgment, Staff Development and Resource Management at third highest. This, too, represents a different set of priorities from the other contexts, for which Staff Development and Resource Management ranked fifth or sixth (except that those with district office contact also ranked resource management high, at second highest). One way to make sense of the differences is to recall that the greater needs

of students in special schools may make inservice and funding issues more critical for their school administrators.

Table 6: Jordan Domain Mean Ratings and Ranks
Partitioned by School Level of Contact

Q#	Elem	Rk	JrHi	Rk	Hig	Rk	Spcl	Rk	DOF	Rk	Tot	Domain
3	3.32	4	3.20	4	3.04	3	3.28	2	3.24	5	18	LEADERSHIP
4	3.34	3	3.27	3	3.11	4	3.23	3	3.36	1	14	JUDGMENT
5	3.09	7	2.79	7	2.68	7	3.03	7	3.12	7	35	INSTRUCTIONAL PROGRAM STAFF
6	3.29	5	3.11	5	2.95	6	3.23	3	3.23	6	25	DEVELOPMENT RESOURCE
7	3.26	6	3.10	6	2.96	5	3.23	3	3.30	2	22	MANAGEMENT INTERPERSONAL
8	3.43	1	3.40	1	3.24	1	3.21	6	3.30	2	11	SKILLS PROFESSIONAL
9	3.41	2	3.40	1	3.14	2	3.36	1	3.30	2	8	STANDARDS

Key: Q# = questionnaire question number; Elem = elementary; JrHi = junior high; Spcl = special;
DOF = district office; Rk = rank; Tot = Total

As has been done previously, a sense of how strongly favored the domains are among the groups can be gained by adding up the rankings of the domains by each group. The highest possible total rating would be 5 (5 x 1), the lowest would be 35 (5 x 7). In these terms Professional Standards and Interpersonal Skills are most strongly supported by the groups, with totals of 8 and 11, while Instructional Program receives the clearly weakest support, with a resounding 35. All Jordan domain mean ratings and ranks partitioned by school level of contact are found above in Table 6.

2. *Utah administrator mean responses partitioned by school level of contact.* Administrators throughout the state whose contact is with special schools found all seven domains more essential to competence than did any other group. Although the trend is not nearly so clear as in Jordan School District, elementary school administrators also gave relatively high marks for all domains. Those with primary contact at the junior high school level found the domains least compelling but still rated all of them between 2.93 (Instructional Program) and 3.19 (Leadership and Interpersonal Skills). Resource Management received most support in special schools and least in elementary and high schools. Those in special schools

rated four of the domains the same and, surprisingly, put Interpersonal Skills at the top rather than sixth as in Jordan.

Table 7: Utah Domain Mean Ratings and Ranks

Partitioned by School Level of Contact

Q#	Elem	R	JrHi	Rk	Hig	R	Spcl	Rk	DOf	R	Tot	Domain
		k			h	k				k		
3	3.58	2	3.19	1	3.52	3	3.67	2	3.40	2	10	LEADERSHIP
4	3.56	3	3.11	5	3.59	2	3.67	2	3.45	1	13	JUDGMENT
5	3.33	5	2.93	7	3.24	5	3.33	6	3.11	7	30	INSTRUCTIONAL
6	3.31	6	3.00	6	2.98	7	3.33	6	3.26	6	31	PROGRAM STAFF
7	3.31	6	3.12	3	3.16	6	3.67	2	3.34	4	21	DEVELOPMENT RESOURCE
8	3.67	1	3.19	1	3.72	1	4.00	1	3.40	2	6	MANAGEMENT INTERPERSONAL
9	3.47	4	3.12	3	3.38	4	3.67	2	3.28	5	18	SKILLS PROFESSIONAL STANDARDS
Key: Q# = questionnaire question number; Elem = elementary; JrHi = junior high/middle; Spcl = special; DOf = district office; Rk = rank; Tot = Total												

The rating-ranks reveal strongest support for inclusion of Interpersonal Skills among evaluation domains and weakest support for Staff Development and Instructional Program. Totals could range from 5 (5 x 1), to 35 (5 x 7). Interpersonal Skills received a 6; Staff Development received a 31 and Instructional Program a 30. Leadership and Judgment followed Interpersonal Skills closely at 10 and 13, while Professional Standards and Resource Management filled in the middle rungs. All the Utah domain mean ratings and ranks partitioned by school level of contact are found above in Table 7.

3. School level differences. School level was selected because the competencies required of administrators at one school level may differ from those required of administrators at other school levels. In Jordan, those whose closest contact is with elementary schools or district offices found the seven domains to be the most essential to administrator competence. Those whose contact is with high schools found the domains least essential; however, even there the lowest rating was a 2.68 for Instructional Program. Regardless of school level, respondents felt that ability in the area of Instructional Program was least essential to administrator competence.

The difference between respondents with closest contact to special schools and all other groups on the Interpersonal Skills Domain is another striking find. These respondents rated it second least essential to competence while all other groups rated it highest or second highest. Of note is also the three-way tie at special schools among Judgment, Staff Development and Resource Management Domains at third highest. This, too, represents a different set of priorities from the other contexts, for which Staff Development and Resource Management ranked fifth or sixth (except that those with district office contact also ranked resource management high, at second highest). One way to make sense of the differences is to recall that the greater needs of students in special schools may make inservice and funding issues more critical for their school administrators.

Administrators throughout the state whose contact is with special schools found all seven domains more essential to competence than did any other group. Although the trend is not nearly so clear as in Jordan School District, elementary school administrators also gave relatively high marks for all domains. Those with primary contact at the junior high school level found the domains least compelling but still rated all of them between 2.93 (Instructional Program) and 3.19 (Leadership and Interpersonal Skills). Resource Management received most support in special schools and least in elementary and high schools. Those in special schools rated four of the domains the same and, surprisingly, put Interpersonal Skills at the top rather than sixth as in Jordan.

Differences of opinion regarding the usefulness of indicators in all of the factor groups appear along the facet of school level with which respondents have most contact. Those who have contact predominantly with the elementary or high school levels score indicators in the factor Group A domains higher than those who have contact predominantly with the district office. Those who have contact predominantly with the junior high level score indicators in factor Group B higher than those who have contact predominantly with the district office. Numerous differences of opinion regarding the usefulness of indicators in all factor groups when responses to the question about administrator contact and school level contact are cross-tabulated.

Differences in levels of acquiescence appeared among those most closely associating with the different school levels. By level of school contact, those associating with the district office engaged in mild acquiescence least, but the level was still 76%. Those whose closest contact is with special schools engaged in the most repetition at a rate of 87% of such respondents. Those whose closest contact is with junior high schools engaged in extreme acquiescence—at 16%, at rates notably less than those associated with the district office—at 25%. But both rates remain near 20%. Among school-level groupings, the positive skew is most apparent among those

associating with the district office (92%) and least apparent among those not responding to the school level question (40%) and those associating with high schools (71%).

E. Mean responses and differences partitioned by years of experience

1. Jordan stakeholder mean responses partitioned by years of experience.

Partitioning respondents into groups by years of experience allows one to see whether the perceived essentialness of domains changed over the years. The only two domains for which this appears significant were Staff Development and Resource Management which were both rated higher as experience grew. More striking were the differences between groups. Those who did not work in the schools rated all domains but Judgment lower than any of the employed groups. Those just starting their educational careers and those closest to retirement rated Interpersonal Skills and Professional Standards higher than their peers in between. New employees also favored Staff Development more than the other employed groups. Those with two to eight years of experience found Leadership and Judgment to be more essential than the other groups did, while those closest to retirement rated Judgment lower than all other groups.

Table 8: Jordan Domain Mean Ratings
and Ranks Partitioned by Years of Experience

Q#	Not	R	<2	Rk	2-8	R	9-15	Rk	16+	R	Tot	Domain
		k				k				k		
3	3.05	4	3.26	4	3.27	4	3.24	4	3.25	3	19	LEADERSHIP
4	3.20	1	3.27	3	3.32	3	3.25	3	3.15	6	16	JUDGMENT
5	2.90	7	2.94	7	2.93	7	2.97	7	2.98	7	35	INSTRUCTIONAL
6	2.96	6	3.26	4	3.18	5	3.18	5	3.21	5	25	PROGRAM STAFF
7	2.97	5	3.16	6	3.16	6	3.17	6	3.25	3	26	DEVELOPMENT RESOURCE
8	3.19	2	3.41	1	3.36	1	3.38	1	3.44	1	6	MANAGEMENT INTERPERSONAL
9	3.15	3	3.40	2	3.36	1	3.35	2	3.40	2	10	SKILLS PROFESSIONAL STANDARDS
<p>Key: Q# = questionnaire question; Not = no experience as a school employee; <2 = less than two years experience as a school employee; 16+ = sixteen or more years of experience as a school employee; Rk = rank; Tot = total</p>												

Rating-ranks most strongly support Interpersonal Skills at 6, just one point away from being rated most essential by all groups. Rating-ranks most weakly support Instructional Program which was rated lowest by all groups, thus receiving a 35. All Jordan domain mean ratings and ranks partitioned by years of experience are found above in Table 8.

2. Utah administrator mean responses partitioned by years of experience.

The only Jordan trend in responses over the years of employment that was repeated among administrators statewide was that Interpersonal Skills remained highly rated as experience increased. The essentialness of Judgment waned over time in the statewide sample. An odd contradiction was that Staff Development was rated much lower by administrators with the least amount of experience than any of their more experienced peers while they rated Leadership and Judgment much higher than the other groups. Maybe they admire ability in these areas but believe that inservice will not help them develop it. In general, Interpersonal Skills, Leadership and Judgment were found more essential to administrator competence, and Staff Development, Instructional Program and Resource Management were found less essential. Finally, it appears that Professional Standards are valued comparatively more highly in Jordan School District than in other districts throughout the state. All Utah administrator domain mean ratings and ranks partitioned by years of experience are found below in Table 9.

Table 9: Utah Administrator Domain Mean Ratings and Ranks Partitioned by Years of Experience

Q#	<2	R	2-8	R	9-15	Rk	16+	R	Tot	Domain
		k		k				k		
3	3.89	1	3.47	3	3.52	2	3.40	3	9	LEADERSHIP
4	3.78	2	3.50	1	3.49	3	3.46	2	8	JUDGMENT
5	3.22	5	3.24	6	3.26	6	3.16	6	23	INSTRUCTIONAL PROGRAM
6	2.78	7	3.24	6	3.29	4	3.12	7	24	STAFF DEVELOPMENT
7	3.22	5	3.35	4	3.26	6	3.21	5	20	RESOURCE MANAGEMENT
8	3.67	3	3.50	1	3.68	1	3.53	1	6	INTERPERSONAL SKILLS
9	3.33	4	3.32	5	3.29	4	3.36	4	17	PROFESSIONAL STANDARDS
Key: Q# = questionnaire question; Not = no experience as a school employee; <2 = less than two years experience as a school employee; 16+ = sixteen or more years of experience as a school employee; Rk = rank; Tot = total										

3. Years of school employment differences. Partitioning respondents into groups by years of experience revealed whether the perceived essentialness of domains changed over the years. The only two domains for which this appears significant were Staff Development and Resource Management

which were both rated higher as experience grew. More striking were the differences between groups. Those who did not work in the schools rated all domains but Judgment lower than any of the employed groups. Those just starting their educational careers and those closest to retirement rated Interpersonal Skills and Professional Standards higher than their peers in between. New employees also favored Staff Development more than the other employed groups. Those with two to eight years of experience found Leadership and Judgment to be more essential than the other groups did, while those closest to retirement rated Judgment lower than all other groups.

The only Jordan trend in responses over the years of employment that was repeated among administrators statewide was that Interpersonal Skills remained highly rated as experience increased. The essentialness of Judgment waned over time in the statewide sample. An odd contradiction was that Staff Development was rated much lower by administrators with the least amount of experience than any of their more experienced peers while they rated Leadership and Judgment much higher than the other groups. Maybe they admire ability in these areas but believe that inservice will not help them develop it. In general, Interpersonal Skills, Leadership and Judgment were found more essential to administrator competence, and Staff Development, Instructional Program and Resource Management were found less essential. Finally, it appears that Professional Standards are valued comparatively more highly in Jordan School District than in other districts throughout the state.

In the excessive repetition category, 33% of those who did not mark the number of years of experience they had in school employment consecutively answered more than half of the indicator questions the same. By number of years of work experience in schools, the most experienced group (16+ years' experience) engaged in acquiescence the least at 77%, and those who did not say how many years' experience they had repeated the most at 88%. By years of experience in school employment, 16% of those with less than 2 years of experience got into excessive ruts, while 33% of those who did not mark their number of years' experience engaged in acquiescence. By years of experience, possible positive bias is most apparent among those with 9 to 15 years experience (94%) and least apparent among those who did not answer the experience question (45%).

F. Mean responses and differences partitioned by respondent gender

1. Jordan stakeholder mean responses partitioned by respondent gender.

Comparisons across gender show great similarity in the order of priorities—both groups rate Instructional Program clearly lowest and Interper-

sonal Skills highest. Only one or two hundredths of a point transpose Staff Development and Resource Management for fifth and sixth places (although men did rate Staff Development sixth and women gave Resource Management that distinction). The most noteworthy difference between the sexes is seen in the level of essentialness they feel for the domains. Across the board, men were less enthusiastic about the domains than were women. The difference is about a quarter of a point (spanning from .21 on Judgment to .28 on Leadership). One explanation would be that men simply are not as likely as women to strongly express their opinions. Yet the average responses for the domains by men range from 2.75 to 3.21 (or .46 points) while that for women is narrower at .42 (from 3.02 to 3.44). Thus one could say that men displayed a greater range of opinion than did women while women's responses showed consistently greater intensity. All Jordan domain mean ratings and ranks partitioned by respondent gender are found below in Table 10.

Table 10: Jordan Domain Mean Ratings and Ranks

Partitioned by Respondent Gender

Q#	Male	Rank	Female	Rank	Domain
3	3.03	4	3.31	4	LEADERSHIP
4	3.13	3	3.34	3	JUDGMENT
5	2.75	7	3.02	7	INSTRUCTIONAL PROGRAM
6	3.00	6	3.25	5	STAFF DEVELOPMENT
7	3.01	5	3.23	6	RESOURCE MANAGEMENT
8	3.21	1	3.44	1	INTERPERSONAL SKILLS
9	3.17	2	3.42	2	PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number

2. Utah administrator mean responses partitioned by respondent gender.

The results of the statewide questionnaire were quite unlike those from Jordan School District. As to the order of priorities, the only agreement was on Resource Management being the fifth most essential domain to administrator competence. Women tied in their estimation of the essentialness of Instructional Program, Staff Development and Interpersonal Skills—all at second. They ranked Professional Standards least essential, at 3.33, and Judgment most essential, at 3.57. In contrast, men ranked Interpersonal Skills as most essential, at 3.58, and Staff Development least essential, at 3.10. One constant remained—men expressed a wider range of opinion than women. In the statewide questionnaire, women rated Judgment, Instructional Program, Staff Development, and Resource Management much higher than men (differences ranging from .10 on Judgment to .37 on Resource Management). But men rated Leadership, Interpersonal Skills, and Professional Standards higher than women with narrower margins (differences ranging from .01 to .11). It is striking

that neither the men nor the women in the Jordan and statewide questionnaires demonstrated similar priorities, possibly indicating that some other demographic characteristic—or combination of characteristics, accounts for priorities better than gender. All Utah administrator domain mean ratings and ranks partitioned by respondent gender are found below in Table 11.

Table 11: Utah Administrator Domain Mean Ratings
and Ranks Partitioned by Respondent Gender

Q#	Male	Rank	Female	Rank	Domain
3	3.45	3	3.43	5	LEADERSHIP
4	3.47	2	3.57	1	JUDGMENT
5	3.12	6	3.47	2	INSTRUCTIONAL PROGRAM
6	3.10	7	3.47	2	STAFF DEVELOPMENT
7	3.19	5	3.43	5	RESOURCE MANAGEMENT
8	3.58	1	3.47	2	INTERPERSONAL SKILLS
9	3.34	4	3.33	7	PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number

3. Gender differences. Comparisons across gender show great similarity in the order of priorities—both groups rate Instructional Program clearly lowest and Interpersonal Skills highest. Only one or two hundredths of a point transpose Staff Development and Resource Management for fifth and sixth places (although men did rate Staff Development sixth and women gave Resource Management that distinction). The most noteworthy difference between the sexes is seen in the level of essentialness they feel for the domains. Across the board, men were less enthusiastic about the domains than were women. The difference is about a quarter of a point (spanning from .21 on Judgment to .28 on Leadership). One explanation would be that men simply are not as likely as women to strongly express their opinions. Yet the average responses for the domains by men range from 2.75 to 3.21 (or .46 points) while that for women is narrower at .42 (from 3.02 to 3.44). Thus one could say that men displayed a greater range of opinion than did women while women's responses showed consistently greater intensity. Women scored factor Group A indicators (Leadership, Judgment and Interpersonal Skills) higher than men, they also scored factor Group C indicators (Professional Standards) higher than men.

The results of the statewide questionnaire were quite unlike those from Jordan School District. As to the order of priorities, the only agreement was on Resource Management being the fifth most essential domain to administrator competence. Women tied in their estimation of the essentialness of Instructional Program, Staff Development and Interpersonal Skills—all at second. They ranked Professional Standards least essential, at 3.33, and Judgment most essential, at 3.57. In contrast, men

ranked Interpersonal Skills as most essential, at 3.58, and Staff Development least essential, at 3.10. One constant remained—men expressed a wider range of opinion than women. In the statewide questionnaire, women rated Judgment, Instructional Program, Staff Development, and Resource Management much higher than men. But men rated Leadership, Interpersonal Skills, and Professional Standards higher than women with narrower margins. It is striking that neither the men nor the women in the Jordan and statewide questionnaires demonstrated similar priorities, possibly indicating that some other demographic characteristic—or combination of characteristics, accounts for priorities better than gender.

In comparisons between men and women, men engaged in acquiescence less—at 74%, than women—at 84%, but those who did not divulge their gender engaged in it the least—at 73%. Women engaged in excessive acquiescence more than men—at 20% to 17%. But neither did as much as those who were unresponsive to the gender question (21%). As between men and women, women exhibited a stronger tendency to mark consecutive items favorably (92%), but nonresponders to the gender question exhibited even less of a tendency to mark positive response-sets (67%). The interpretation of the positive skew as possible bias is supported by the fact that the longer strands of identical answers also tended to be the more extreme.

G. Mean responses and differences partitioned by district urbanization level

1. Utah administrator mean responses partitioned by district urbanization Level. The difference between responses from administrators in urban and rural settings is clear throughout the domains. Urban administrators found the domains more essential than their rural peers. The difference is most apparent in the Resource Management domain, which was rated 3.15 by rural administrators and 3.50 by their urban peers. It is least apparent in Leadership, which was rated 3.45 by those in rural school districts and 3.48 by those in urban districts. Both groups found Interpersonal Skills to be most essential, seconded by Judgment, and both placed Staff Development last. Urban district administrators rated Resource Management the same as Judgment. Rural administrators rated it second least important and tied instead on Instructional Program and Staff Development at least essential to competence. Urban administrators showed a greater range of opinion, going from 3.20 to 3.68 while their rural counterparts' average responses for domains spanned from 3.11 to 3.50. It is again surprising that there appears to be no clear parallel between Jordan School District responses and those of either urban or rural districts combined from across the state, except that all rate Interpersonal Skills highest. All Utah adminis-

trator domain mean ratings and ranks partitioned by district urbanization are found in Table 12 below. Jordan aggregate mean ratings and ranks as well as Jordan administrator mean ratings and ranks are listed to help compare Jordan School District to rural and urban districts throughout the state.

Table 12: Utah Administrator Domain Mean Ratings
and Ranks Partitioned by District Urbanization Level

Q#	Rural	Rank	Urban	Rank	JAggr	Rank	JAdmin	Rank	Domain
3	3.45	3	3.48	4	3.23	4	3.34	4	LEADERSHIP
4	3.46	2	3.50	2	3.27	3	3.48	2	JUDGMENT
5	3.11	6	3.34	6	2.94	7	3.00	7	INSTRUCTIONAL
6	3.11	6	3.20	7	3.18	5	3.17	6	PROGRAM STAFF
7	3.15	5	3.50	2	3.17	6	3.31	5	DEVELOPMENT RESOURCE
8	3.50	1	3.68	1	3.37	1	3.53	1	MANAGEMENT INTERPERSONAL
9	3.30	4	3.43	5	3.35	2	3.45	3	SKILLS PROFESSIONAL
Key: Q# = questionnaire question number									

2. *District urbanization level differences.* The difference between responses from administrators in urban and rural settings is clear throughout the domains. Urban administrators found the domains more essential than their rural peers. The difference is most apparent in the Resource Management domain, which was rated 3.15 by rural administrators and 3.50 by their urban peers. It is least apparent in Leadership, which was rated 3.45 by those in rural school districts and 3.48 by those in urban districts. Both groups found Interpersonal Skills to be most essential, seconded by Judgment, and both placed Staff Development last. Urban district administrators rated Resource Management the same as Judgment. Rural administrators rated it second least important and tied instead on Instructional Program and Staff Development at least essential to competence. Urban administrators showed a greater range of opinion, going from 3.20 to 3.68 while their rural counterparts' average responses for domains spanned from 3.11 to 3.50. It is again surprising that there appears to be no clear parallel between Jordan School District responses and those of either urban or rural districts combined from across the state, except that all rate Interpersonal Skills highest.

III. Cumulative rankings of competency domains among partitioned groups

If one adds up all the rankings given to each of the domains by each of the partitioned groups in the Jordan Questionnaire and compares these totals to the straight totals if domains had been ranked the same by all groups, then one gets a sense of the level of consistency in rankings of the domains across the groups in Jordan School District. For example, Interpersonal Skills is the domain with the best cumulative rank at 36 which is just 8 points above a pure second best ranking. Professional Standards came in next with a 39 which is 3 points below a pure second best ranking. Judgment comes in third at 58, 5 points above a pure third. Leadership comes in fourth at 81, 3 points above a pure fourth. Staff Development and Resource Management both approximate pure fifth places with scores of 107 and 110 respectively. Instructional Program scores the worst cumulative rank at 144 which is 3 points above a pure seventh place rank.

Similarly, if one adds up all the rankings given to each of the domains by each of the partitioned groups in the Statewide Questionnaire and compares these totals to the straight totals if domains had been ranked the same by all groups, one gets a sense of the level of consistency in rankings of the domains across the groups in the state of Utah. Among these groups—as in Jordan— Interpersonal Skills received the best cumulative ranking at 29 which is 7 points below a pure best ranking. Judgment received the second best ranking at 52 which is 8 points below a pure second best cumulative ranking. Leadership came in a close third at 57, 9 points above a pure third best rating. Professional Standards ranked fourth in the cumulative ranks at 86 which is 2 point below a pure fourth place. Resource Management ranked fifth at 101 or 9 points above a pure fifth highest rank. Instructional Program and Staff Development both approximated pure sixth place ranks with Instructional Program 9 points higher than a pure sixth at 123 and Staff Development 5 points lower than a pure sixth at 137.

This set of comparisons once again reinforces a clear distinction between the most essential domains for administrator success—Interpersonal Skills, Professional Standards, Judgment and Leadership—and the task force selected domains that are not quite so essential to success—Resource Management, Staff Development and Instructional Program. Despite this similarity, it appears that greater disagreement existed among the groups in the Statewide Questionnaire because cumulative ranks are farther away from pure ranks than in Jordan School District. The numeric analyses above are largely borne out in the respondent comments which are synthesized below. The cumulative ranking comparisons are found in Table 13 below.

Table 13: Cumulative Rankings of Competency Domains Among Partitioned
Groups in the Jordan and Utah Administrator Questionnaires

Q#	Jordan	Pure	Off	Utah	Pure	Off	Domain
3	81	4-	3	57	3-	9	LEADERSHIP
4	58	3-	5	52	2+	8	JUDGMENT
5	144	7-	3	123	6-	9	INSTRUCTIONAL PROGRAM
6	107	5+	2	137	6+	5	STAFF DEVELOPMENT
7	110	5+	5	101	5-	9	RESOURCE MANAGEMENT
8	36	2-	8	29	1+	7	INTERPERSONAL SKILLS
9	39	2-	3	86	4-	2	PROFESSIONAL STANDARDS

Key: Q# = questionnaire question number; Pure = closest ranking if all groups had rated the domain the same; Off = amount above or below a pure ranking

IV. Conclusions and recommendations

The results of this inquiry suggest that valid content for administrator evaluation can be achieved with far fewer responsibility categories than have been proposed by such entities as the NAESP, NASSP and NPBEA. Results did not indicate a need to vary evaluation instruments along the dimensions specified in the rationale section, such as type of administrator or school level, other than for assistant principals who may not engage in activities related to all domains (e.g., instructional program or resource management). It highlighted the importance of competence in the more social indicators found in the Leadership, Judgment and Interpersonal Skills domains. It also highlighted a discontinuity between the opinions or practitioners and researchers regarding the importance of administrator behaviors within the realm of Instructional Program. The levels of acquiescence introduce a heightened sense of caution regarding Likert-scaled questionnaire research worthy of further investigation, but not sufficient to invalidate the research strategy or results of this inquiry. Finally this research demonstrates how engaging in broad-based consultation and district tailoring of evaluation systems can increase their validity and credibility.

Such findings should prove useful as a basis for local school districts' efforts to create evaluation instruments and meaningful staff development for their administrators. With this information, educational leaders will be able to make better informed, and judicially supportable, decisions regarding administrative staff development, evaluation and remediation. The information should be of interest to school administrators, policy makers and researchers in this field.

A. *Priorities for retention of task force-proposed performance indicators*

In analyzing the value of various domains, subcategories and indicators for the Jordan Administrator Evaluation System, Steve Baldrige has considered the results of statistical procedures which answer the following questions:

1. *Mean.* To what extent did the respondents, as a whole, agree that the performance indicator is helpful in evaluating administrator performance? The question was answered by looking at the mean rating (0 to 4) of indicators. Because all of the indicators received favorable mean ratings (2.69 to 3.43), none can be thoughtlessly eliminated—and the Taskforce should feel satisfied with favorable feedback on their hard work done to this point.

2. *Goodness-of-measure.* To what extent did the respondents agree that each performance indicator is a good measure of its domain? The question was answered by looking at factor analyses measuring the consistency of indicator ratings to domain ratings. As further confirmation of the conceptual rigor with which the Taskforce has addressed its work, factor correlations from indicators to domains ranged from .59 to .83.

3. *Goodness-of-fit.* To what extent did the respondents agree that performance indicators belong together in the same domain? The question was answered by looking at factor analyses measuring respondents' patterns in rating indicators. Indicator correlations to group patterns ranged from .45 to .79.

The fifteen indicators ranking highest and lowest in terms of their mean (M#1 to M#15 and M#63 to M#77), in terms of their domain correlation (D#1 to D#15 and D#63 to D#77), and in terms of their group correlation (G#1 to G#15 and G#63 to G#77), were identified. The following sets are proposed for discussion by the Taskforce based on the results of the statistical procedures. CLEAR KEEPERS (CK) are indicators that scored among the top twenty percent under at least one of the three statistical tests and were not found among the bottom twenty percent in any of the tests. KEEPERS (K) are indicators that scored in the middle sixty percent on all three statistical tests, or in the top twenty percent on at least one test but also in the bottom twenty percent on one test. POSSIBLE DELETIONS (PD) are indicators that ranked in the bottom twenty percent under at most two of the statisti-

cal procedures. CLEAR DELETIONS (CD) are those indicators ranked in the bottom twenty percent under all three procedures.

In broad terms, differences in respondent groups' opinions regarding the usefulness of indicators are significant along many of the facets. Because all of the groups held the indicators in such high regard, however, the differences do not make any of the indicators invalid as content for evaluation of administrators in any of the categories. They are all apt measures of administrator performance, but some are clearly considered more apt by some groups than others.

APPENDIX C: CONSTRUCT VALID REGROUPINGS FOR THE
JORDAN ADMINISTRATOR EVALUATION SYSTEM*

Recomm'd	Set	Domain, Subcategory or Indicator
Same		JUDGMENT: The school leader makes wise decisions in a timely fashion based on the best available information.
Modify		PROBLEM ANALYSIS
A-Same	CK	23. Recognizes problem situations
A-Add	K	21/22. Gathers, organizes and analyzes information to help resolve problems
A-Same	CK	24. Involves all stakeholders in problem solving
Same		JUDGMENT
A-Same	CK	25. Uses appropriate strategies for resolving problems
A-Same	K	26. Makes decisions in a timely fashion
A-Same	CK	27. Considers the rights and concerns of others when making judgments
A-Same	CK	28. Controls bias when making judgments
A-Same	K	29. Makes conclusions with self-control and without unnecessary emotion or stress
Same		INTERPERSONAL SKILLS: The school leader treats others in a professional manner with respect and sensitivity, and facilitates a caring and motivating environment.
Same		SENSITIVITY
A-Add	K	59. Identifies, understands, and respects the diversity of values and cultures in school and in society
A-Same	CK	61. Exhibits behaviors that promote a positive and caring climate
A-Same	CK	62. Works to reduce conflict and increases mutual respect
Same		MOTIVATION
A-Same	CK	63. Encourages teamwork and collegiality
A-Same	K	64. Recognizes achievements and professional contributions
Same		COMMUNICATION SKILLS
A-Same	K	65. Articulates ideas and beliefs clearly in both oral and written form
A-Same	CK	66. Communicates effectively with teachers, students, parents, peers, district and state personnel, and community members
Same		STAFF RELATIONSHIPS
A-Same	CK	67. Relates well with other staff members
A-Same	CK	68. Provides leadership in developing positive staff morale
Same		STUDENT RELATIONSHIPS

*. See Generally STEVE BALDRIDGE, AN INQUIRY INTO THE CONTENT VALIDITY OF PUBLIC SCHOOL ADMINISTRATOR EVALUATION DOMAINS AND PROFICIENCY INDICATORS 369-72 (1996).

- | | | |
|--------|----|---|
| A-Same | CK | 69. Encourages student growth and development |
| A-Same | K | 70. Assists teachers in helping students to develop realistic and positive self-concepts |
| A-Same | PD | 71. Works to equalize educational opportunities for students |
| A-Same | K | 72. Cooperatively develops and maintains procedures for a high level of positive student behavior |
| A-Same | CK | 73. Deals effectively with behavior problems |

Same COMMUNITY RELATIONS

- | | | |
|--------|----|---|
| A-Same | CK | 74. Develops public confidence and mutual respect |
| A-Same | K | 81. Holds the respect and confidence of the community, peers, administration, and board of education (See 74.) |
| A-Same | PD | 75. Encourages appropriate community participation in school activities, including development of school goals and program evaluation |
| A-Same | PD | 76. Encourages productive teacher-parent relationships |

Modify	I. LEADERSHIP: The school leader develops a shared strategic vision with staff, students, parents, and the community, facilitates the realization of the vision, and provides encouragement, support, and opportunities for staff development.
--------	---

Modify LEADERSHIP

- | | | |
|--------|----|--|
| A-Same | CK | 10. Models high personal work standards |
| A-Same | PD | 11. Develops shared strategic vision with the involvement of staff, students, parents, and community |
| A-Add | K | 42. Articulates performance expectations (See 10., 14., 72.) |
| A-Add | PD | 17. Empowers others and assists them in completing tasks |
| A-Add | K | 14. Gives assignments on a fair basis with clearly explained expectations |
| A-Add | K | 13. Manages the issues and transitions that occur with change |

Modify ORGANIZATIONAL OVERSIGHT

- | | | |
|-------|----|---|
| A-Add | K | 58. Uses effective organization and management procedures |
| A-Add | CK | 40. Demonstrates competency in supervision |
| A-Add | K | 44. Demonstrates competency in performing staff evaluation responsibilities |
| A-Add | K | 48. Facilitates staff involvement in planning staff development |
| A-Add | PD | 50. Informs staff of opportunities for professional growth |

Same	INSTRUCTIONAL PROGRAM: The school leader knows the school curriculum and provides leadership in curriculum planning, instruction, and evaluation.
------	--

Modify CURRICULUM

- | | | |
|--------|----|---|
| B-Move | PD | 30. Involves teachers in the design, development, and management of curriculum |
| B-Same | K | 32. Identifies and defines the relationships among the written curriculum, what is taught, and what is tested |
| B-Same | K | 31. Implements current research findings |

Same		INSTRUCTION
B-Same	PD	33. Sets instructional objectives based on district and school vision statements and desired student outcomes
B-Same	K	35. Demonstrates understanding of the main differences between effective and ineffective instructional practices
B-Same	PD	34. Identifies the key attributes of skilled instructional leaders including the use of classroom strategies that respond to various learning styles

Same		EVALUATION
B-Same	PD	37. Examines the relationships between school goals and student outcomes
B-Same	PD	39. Explains the implications of assessment data to teachers and patrons, and links them to school improvement programs
B-Add	PD	45. Uses staff evaluations and other individualized needs assessment data to form staff development and staff orientation programs

Same		PROFESSIONAL STANDARDS: The school leader knows and implements legal standards and policies, and is a role model of a professional educator.
------	--	---

Same		LEGAL AND POLICY APPLICATIONS
C-Same	CK	77. Demonstrates knowledge of and complies with local, state, and federal legal standards applicable to public schools
C-Same	CK	78. Knows and implements district policies

Same		ETHICS & ROLE MODELING
C-Same	CK	79. Demonstrates personal integrity and morality
C-Same	CK	80. Models district and school belief statements

Same		PROFESSIONALISM
C-Same	CK	82. Is well groomed and appropriately dressed
C-Same	CK	83. Meets the physical and mental demands of the job
C-Same	CK	84. Shows personal initiative and self-confidence
C-Add	PD	86. Participates in conferences and other professional activities dealing with educational issues

Same		RESOURCE MANAGEMENT: The school leader effectively utilizes available resources to address basic needs and achieve essential educational goals in an efficient and ethical manner.
------	--	---

Modify		MANAGEMENT OF FUNDS AND MATERIALS
D-Same	K	51. Develops and administers a school and activities budget
D-Same	K	52. Implements the district purchasing system
D-Same	PD	53. Administers a materials and equipment inventory
D-Same	K	54. Develops and administers an accountability system for budgeting and reapportioning resources

D-Same	PD	55. Develops and administers a system for staff participation in determining goals, apportioning resources, and evaluating use of resources
Same		<u>MANAGEMENT OF PHYSICAL PLANT</u>
D-Same	K	56. Assumes management responsibilities for school plant, facilities, and equipment
D-Same	K	57. Assumes responsibility for development and implementation of necessary schedules involving students, staff, community, facilities, and equipment

ELIMINATE

Delete	STAFF DEVELOPMENT: The school leader provides encouragement, support, and opportunities for staff to develop and strengthen professional knowledge and skills.
--------	--

Delete PHILOSOPHICAL/CULTURAL VALUES

Delete DELEGATION

Delete PERSONAL PROFESSIONAL DEVELOPMENT

Delete	PD	12. Establishes high expectations for staff
Delete	PD	15. Maintains active interest in the work of others
Delete	PD	16. Uses a team approach to school management
Delete	PD	18. Evaluates school programs and projects to determine effectiveness
Delete	PD	19. Facilitates the implementation of events, projects and activities
Delete	CD	20. Identifies appropriate strategies or tools for collecting information
Delete	PD	36. Demonstrates understanding of the main differences between effective and ineffective instructional practices
Delete	PD	38. Analyzes assessment data and draws inferences for revising school programs
Delete	K	41. Demonstrates competency in mentoring, coaching, and conferencing
Delete	PD	43. Initiates activities to improve instruction
Delete	K	46. Provides face-to-face and written performance feedback
Delete	PD	47. Encourages staff to develop, pursue, and evaluate educational goals and objectives
Delete	PD	49. Facilitates productive cooperation to improve instruction
Delete	PD	60. Demonstrates a personal philosophy that is compatible with the district philosophy
Delete	PD	85. Uses professional literature to stay informed about current educational practices