Multiple Sources of Evidence:

An Analysis of Stakeholders' Perceptions of Different Indicators of Student Learning

Thomas R. Guskey
College of Education
University of Kentucky
Lexington, KY 40506

Phone: 859/257-8666
E-mail: GUSKEY@UKY.EDU

Paper presented at the Annual Meeting of the
American Educational Research Association, Montreal, Canada.

April 2005

Multiple Sources of Evidence:
An Analysis of Stakeholders' Perceptions of Different Indicators of Student Learning

Abstract

This study compared different stakeholders' perceptions of the validity of various indicators of student learning used to judge the quality of schools and individual student's academic performance. Data are based on questionnaire responses of 314 educators from school districts in three states that have implemented comprehensive statewide assessment programs that include high-stakes consequences both for educators and for students. MANOVA results showed significant differences between school administrators and teachers, with administrators favoring the validity and trustworthiness of nationally-normed standardized assessments, state assessments, and district assessments, while teachers favored classroom observations, classroom assessments, homework completion and quality, class participation, and behavior. The implications of these differences for reform initiatives are discussed, particularly with regard to teachers' motivation to improve results.

Multiple Sources of Evidence:
An Analysis of Stakeholders' Perceptions of Different Indicators of Student Learning

Modern education reforms, especially those guided by the *No Child Left Behind* (NCLB) legislation (U.S. Congress, 2001), involve the use of large-scale assessments. Policy makers and legislators at the national and state levels are attracted to assessments as instruments for reform because they can be relatively inexpensive, relatively quick to implement, externally mandated, and the results are highly visible (Linn, 2000). These same policy makers and legislators also are convinced that good data on student performance drawn from large-scale assessments will help focus educators' attention and guarantee success, especially if consequences are attached to the results.

While government officials argue that the large-scale assessments used in most states today are designed primarily to measure students' "proficiency" on carefully articulated standards for student learning, these assessments also are used to evaluate schools and students for the purposes of accountability. As such, the results affect many different stakeholder groups, including school administrators, teachers, students, parents, school board members, future employers, and the community. Because a major intent of most states' assessment programs is to monitor and improve the educational system, however, the stakes are highest for school administrators and teachers (Lane & Stone, 2002).

While the psychometric quality and validity of large-scale assessments for accountability purposes are widely debated (see, Kane, 2002), one point on which both advocates and critics agree is that they represent only one of a variety of indicators of student learning that might be considered. Yet despite calls from professional organizations for protection against high-stakes decisions based on single tests or assessments (American Educational Research Association, 2000, AERA, APA, & NCME, 1999), the exclusive use of large-scale assessment results for making high-stakes decisions about schools and students remains widespread (Barton, 2002; Kifer, 2001).

This study was designed to compare different stakeholders' perceptions of the validity of various indicators of student learning, in addition to large-scale assessment results, which potentially might be used to judge the quality of a school's instructional program and the level of students' academic success. Specifically, its purpose was to determine if school administrators and teachers share similar perceptions regarding what evidence provides the best, most accurate, and most trustworthy information about students' academic performance. School administrators and teachers were selected because the consequences of accountability affect them, along with their students, most directly. The goal of the study was to show the extent of consensus or disagreement among these different stakeholder groups and to consider the implications of their shared or differing perceptions for education improvement efforts.

Data Sources and Method of Analysis

The data for this investigation were drawn from 320 educators from three different states who took part in summer professional development institutes. All three states have implemented comprehensive statewide assessment programs that include both rewards and sanctions for schools, along with specific consequences for students, based on assessment results. The schools and school districts from which these educators came varied widely in size and in the social, demographic, and economic characteristics of their student populations. The range included large schools in urban centers, large and small schools in suburban areas, and small schools in rural communities. The total sample is described in Table 1. It included 6 superintendents, 66 district level administrators and program directors, 67 principals and assistant principals, 17 counselors and special educators, 49 primary and elementary teachers, 74 middle school teachers, and 35 secondary school teachers. Six educators failed to provide complete information and could not be included in the analysis.

Generally the administrators were more experienced than the teachers (22.9 years versus 15.8 years). Male and females were evenly represented in all groups, with the exception of primary and elementary teachers, who were predominately female (40 versus 9).

[Insert Table 1]

All of the educators included in the study completed the same, one-page "Student Learning Evidence Questionnaire." This questionnaire asked respondents to record their name (optional), their years of experience in education, and their current position (Superintendent, District Level Administrator, Program Director/Coordinator, Principal or Assistant Principal, Counselor, Special Educator, or Teacher (Primary Grades K-2, Elementary Grades 3-5, Middle Grades 6-8, Secondary Grades 9-12).

Next, these educators were asked to rank order 15 different sources of evidence on student learning. The directions read: "Listed below are several sources of evidence on student learning. Please rank order these sources from 1 to 15 based on what you believe (or trust) best shows what students know and can do." The 15 sources of evidence included the following:

___ Nationally-normed standardized assessments

___ State assessments

___ District assessments

___ End-of-course examinations

___ Teacher-developed assessments
___ Teacher observations
___ Regular classroom quizzes
___ Homework completion and quality
___ Portfolios of students' work
___ Student exhibits (projects and reports)
___ Students' grades
___ Compositions and writing assignments
___ Students' class involvement and participation
___ Students' behavior and attitude in class
___ Data on promotion, retention, and dropouts

Two final questions asked "What three sources of evidence (those above or others) would you recommend using to judge the quality of a school?" and "What three sources of evidence (those above or others) would you recommend using to judge the quality of a teacher's teaching?

Responses to the rank-ordering were first combined across current position categories to form Administrator and Teachers groups. Superintendents, district level administrators and program directors, principals and assistant principals, formed the Administrator group. Because counselors and special educators typically work directly with students in schools, their responses were combined with those from the primary, elementary, middle, and secondary grade teachers to form the Teachers group.

Next, responses were analyzed by calculating means and standard deviations of the rankings by both Administrator and Teachers groups. Correlation coefficients among the various rankings were then computed. Finally MANOVA procedures were used in which the educators' position was considered the single design factor and their rankings of the different sources of evidence as 15 interrelated dependent variables. Univariate results were then explored to further clarify the results.

Results

Analyses of the means, standard deviations, and relative ranking of various indictors by Administrator and Teachers groups are shown in Table 2 and revealed several interesting trends. First, both groups agreed on the relative value of portfolios of students' work, teacher-developed assessments, and compositions and writing assessments. Both also agreed on the relative less importance of data from nationally-normed standardized assessments, and data on promotion, retention, and dropouts. Apparently Administrators and Teachers share the belief that portfolios, classroom assessments, and writing assessments provide valuable evidence on student learning. Such evidence is also more likely to be aligned with a school or teacher's curriculum and instructional goals. Nationally-normed standardized assessments, however, often lack such alignment. The low ranking of data on promotion, retention, and dropouts is more difficult to explain, especially given its importance in most state's accountability programs. Perhaps both Administrators and Teachers believe the many of the factors explaining these data are outside of their direct control. Hence, they offer a less valuable indicator of educators' influence on student learning.

Comparative rankings, however, were also revealing. Generally, Administrators tended to trust district assessments, state assessments, and nationally-normed assessments far more than Teachers. Apparently administrators believe these forms of large-scale assessment provide a clearer and perhaps more objective picture of what students have achieved than do other forms of evidence gathered by individual teachers. For their part, Teachers trusted teacher observations, homework completion and quality, and students' behavior and attitude in class more than did Administrators. Obviously, Teachers believe that the evidence they gather, based on what takes place in their classrooms, provides the most accurate depiction of what student have learned and are able to do. These differences are illustrated in Figure 1, which plots Administrators' and Teachers' average rankings of the various indicators of student learning along with 95% confidence intervals. Correlation results shown in Table 3 yielded similar patterns.

[Insert Tables 2 & 3 and Figure 1]

Finally, the results of MANOVA procedures are shown in Table 4. In this analysis, educators' position (Administrator versus Teacher) was considered the single design factor and rankings of the different sources of evidence as 15 interrelated dependent variables. The multivariate F proved statistically significant by all criteria tested. This provides confirmatory evidence that Administrators and Teachers do, indeed, differ in their perceptions of the value of these various indicators of student learning. Inspection of the separate univariate results indicate that Administrators rate nationally-normed standardized assessment, state assessments, and district assessments more valuable than to Teachers. On the other hand, Teachers rate teacher assessments, teacher observations, students' class involvement and participation, and students' behavior and attitude in class more valuable that do Administrators. Thus, on eight of these 15 indicators of student learning, Administrators and Teachers differ in their perceptions of their relative value as evidence of what students know and are able to do.

[Insert Table 4]

Educational Importance

The differences identified in this investigation in the perceptions of Administrators and Teachers regarding what evidence provides the best and most valid representation of students' level of achievement or performance have important implications for educational reform initiatives. Strong evidence shows that the motivation and effort put forth to improve instruction and student learning are affected by individuals' perceived meaningfulness and relevancy of the assessment results (Lane, Parke, & Stone, 1998). If school administrators and teachers differ in their perceptions of the meaningfulness, validity, and relevancy of specific sources of information on student learning, then it seems imperative that those sources of information be expanded. Minimally they should include indicators of student learning that are trusted and believed by individuals who are stakeholders in the improvement process and for whom the consequences of accountability are most significant.

While it seems unlikely that the use of nationally-normed standardized assessments and state assessments will be abandoned in the foreseeable future, broadening the sources of evidence used to judge the quality of schools or teachers' instructional programs to include a wider array of student learning indicators would likely enhance educators' motivation toward improvement efforts and their involvement in the process. It would also provide a basis for

studying the validity of various sources of information used in an accountability system and how different sources of information might be combined to improve the accuracy and reliability of accountability decisions.

References

American Educational Research Association (2000). Position statement of the American Educational Research Association concerning high-stakes testing in preK-12 education. *Educational Researcher, 29*(8), 24-25.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Barton, P. E. (2002). *Staying on course in education reform*. Princeton, NJ: Statistics & Research Division, Policy Information Center, Educational Testing Service.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*. *21*(1), 31-41.

Kifer, E. (2001). *Large-scale assessment: Dimensions, dilemmas, and policies*. Thousand Oaks, CA: Corwin Press.

Lane, S., Parke, C. S., & Stone, C. A. (1998). A framework for evaluating the consequences of assessment programs. *Educational Measurement: Issues and Practice*, *17*(2), 24-28.

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, *21*(1), 23-30.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(1), 4-16.

U.S. Congress (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.

Group	Number	$\bar{X}$ Years Experience (s.d.)
Superintendents	6	26.20 (7.69)
District Level Administrators	34	24.70 8.89
Program Directors/Coordinators	32	22.69 (8.16)
Principals or Asst. Principals	67	21.94 (7.05)
Counselors	9	17.67 (5.63)
Special Educators	8	14.00 (8.85)
Primary Grade Teachers	11	20.45 (8.87)
Elementary Grade Teachers	38	16.68 (10.28)
Middle Grade Teachers	74	15.30 (9.50)
Secondary Grade Teachers	35	14.26 (9.57)
Total	314	18.91 (9.48)

Table 1. Demographic Data on Study Sample

Indicator	Administrators (n = 139)		Teachers (n = 175)		Total (n = 314)	
	$\bar{X}$ (s.d.)	Rank	$\bar{X}$ (s.d.)	Rank	$\bar{X}$ (s.d.)	Rank
Nationally-normed Standardized Assessments	10.40 (4.35)	14	12.38 (2.97)	14	11.50 (3.77)	14
State Assessments	9.53 (4.19)	10	11.68 (2.69)	14	10.73 (3.59)	13
District Assessments	8.60 (3.77)	9	10.59 (2.82)	12	9.71 (3.42)	11
End of Course Examinations	8.50 (3.12)	8	9.17 (2.92)	10	8.87 (3.02)	9
Teacher-Developed Assessments	5.63 (3.20)	3	4.78 (3.09)	3	5.16 (3.16)	3
Teacher Observations	5.73 (3.71)	5	3.79 (2.70)	1	4.65 (3.32)	2
Regular Classroom Quizzes	7.82 (2.98)	7	7.31 (2.91)	7	7.54 (2.94)	7
Homework Completion & Quality	9.73 (3.50)	12	7.93 (3.41)	8	8.73 (3.56)	8
Portfolios of Students' Work	3.50 (3.23)	1	4.01 (3.37)	2	3.78 (3.53)	1
Student Exhibits (projects and reports)	5.13 (3.54)	2	5.76 (3.51)	6	5.48 (3.53)	4
Students' Grades	10.13 (3.68)	13	9.46 (2.92)	11	9.76 (3.29)	12
Compositions & Writing Assessments	5.72 (3.02)	4	5.45 (3.16)	4	5.57 (3.10)	5
Students' Class Involvement & Participation	6.80 (3.98)	6	5.74 (3.69)	5	6.21 (3.85)	6
Students' Behavior & Attitude in Class	9.68 (4.09)	11	8.53 (3.92)	9	9.04 (4.03)	10
Data on Promotion, Retention, & Dropouts	12.87 (3.01)	15	13.35 (2.54)	15	13.14 (2.76)	15

Table 2. Means, Standard Deviations, and Relative Ranks of Average Rankings Among Different Groups

	Post	Yrs	Natl	State	Dist	Crs	TchA	TchO	Quiz	Hwk	Port	Exhb	Grds	Wrtg	Part	Beh
Years	**-.37**															
Natl. Asmts.	**.26**	-.11														
State Asmts.	**.30**	-.11	.71													
Dist. Asmts.	**.29**	**-.20**	**.60**	**.73**												
Course Exams	.11	-.09	**.24**	**.26**	**.30**											
Tchr. Asmts.	-.13	.08	-.11	-.02	.05	**.20**										
Tchr. Obs.	**-.29**	.08	**-.29**	**-.38**	**-.26**	**-.23**	**.19**									
Class Quizzes	-.09	-.02	**-.24**	**-.27**	**-.18**	**.16**	**.24**	.05								
Homework	**-.25**	**.22**	**-.35**	**-.41**	**-.52**	**-.19**	-.01	.09	**.19**							
Portfolios	.08	-.08	**-.28**	**-.20**	**-.15**	**-.26**	**-.19**	.03	**-.17**	**-.23**						
Exhibits	.09	-.01	**-.30**	**-.33**	**-.31**	**-.20**	**-.20**	-.07	-.11	.01	**.38**					
Grades	-.10	-10	-.01	.02	-.08	.06	-.10	-.11	.07	**.16**	**-.26**	**-.26**				
Wrtg. Asmts.	-.04	.00	**-.30**	**-.25**	**-.30**	**-.22**	**-.26**	-.09	-.07	-.02	**.31**	**.27**	**-.17**			
Participation	**-.14**	.07	**-.50**	**-.51**	**-.52**	**-.39**	-.10	**.14**	-.07	**.20**	.06	.12	**-.14**	.12		
Behavior	**-.14**	.10	**-.37**	**-.46**	**-.46**	**-.41**	**-.23**	.08	**-.24**	**.22**	.03	.08	-.13	.11	**.56**	
Prom. & Retn.	.09	-.06	**.23**	**.20**	**.19**	-.09	**-.33**	**-.25**	-.12	**-.19**	-.04	**-.15**	.00	-.03	**-.21**	-.06

Table 3. Correlations Among Different Measures (n = 314)
[Bold: p < .01]

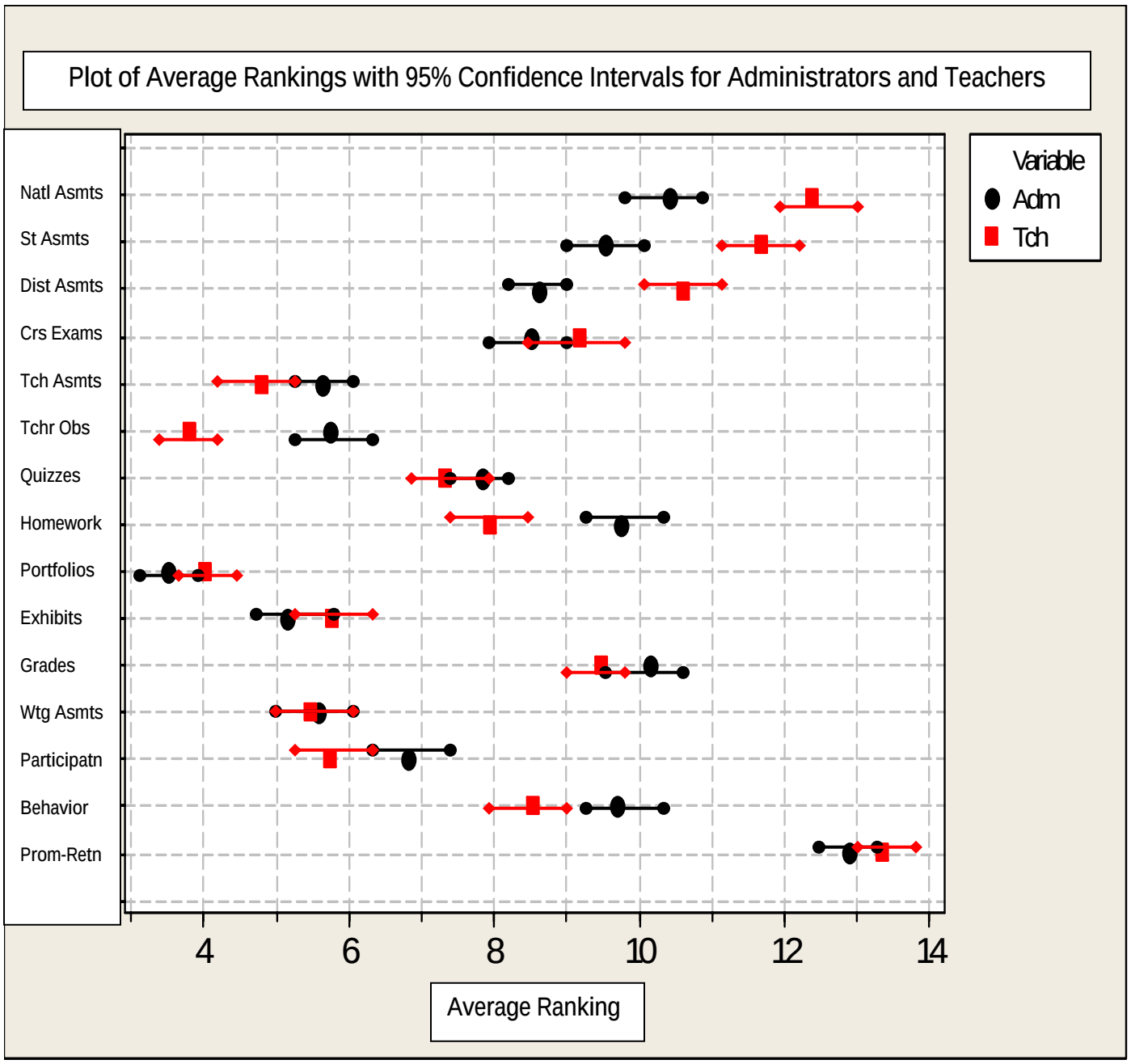

Figure 1. Plot of Average Ranking Per Indicator for Administrators and Teachers

MANOVA Results

Multivariate (Position)	**df**	**Criterion**	**Test Statistic**	**F**	**p**
	15	Wilks'	0.802	4.888	0.000
		Lawley-Hotelling	0.246	4.888	0.000
		Pillai's	0.197	4.888	0.000
Univariate					
Source		**df**	**SS**	**F**	**p**
Natl. Asmts.		1	305.91	23.00	0.000
State Asmts.		1	357.31	30.32	0.000
District Asmts.		1	305.01	28.44	0.000
Course Exams		1	33.96	3.74	0.054
Teacher Asmts.		1	56.00	5.68	0.018
Teacher Obs.		1	287.11	28.29	0.000
Quizzes		1	20.27	2.35	0.126
Homework		1	249.66	20.95	0.000
Portfolios		1	19.53	1.79	0.182
Exhibits		1	30.80	2.49	0.116
Grades		1	35.02	3.26	0.072
Wtg Asmts.		1	5.56	0.58	0.448
Participation		1	86.34	5.90	0.016
Behavior		1	102.55	6.42	0.012
Prom. & Retn.		1	17.70	2.33	0.128

Table 4. MANOVA Results