

Parent Surveys for School Teacher Evaluation

KENNETH D. PETERSON

School of Education, Portland State University, Portland, OR 97207

The improvement of public school teacher evaluation is a topic of considerable interest in the research literature and in professional practice. One suggested direction for better evaluation is to develop sources of data about teacher performance in addition to the traditional administrator reports. Principal visits, particularly as a sole data source about teachers, have been criticized as inadequate and even misleading. Parent viewpoints of teacher performance have been discussed by a number of authors as a promising but undeveloped source of information for evaluation. The purpose of this study was to investigate the use of parent surveys in two Utah school districts where such surveys are a part of the career ladder teacher evaluation systems.

Background

The practice and understanding of teacher evaluation are in need of development. Recent educational reform documents have pointed to increased and improved evaluation as a top priority (Educational Commission of the States, 1983). Current teacher evaluation, which relies upon principal visits and rating forms (Lewis, 1982), has been soundly criticized by researchers (Darling-Hammond, Wise, & Pease, 1983; Scriven, 1981; Stodolsky, 1984). One problem identified in present practice is that principals have sparse systematic information about teachers because of inadequate time and opportunity to collect data. Another problem comes from the sociological tension of teachers and their supervisors (Lortie, 1975); their dependance upon one another often results in biased reporting by principals. It is clear that current evaluation techniques will not meet the demands of school reforms such as teacher career ladders and leadership schemes.

One suggested improvement for teacher evaluation is to develop multiple and variable lines of evidence of teacher quality (Peterson, 1984). Multiple data sources enable a range of information to be added to observation of classroom process

This study was supported by Grant # G008410033 from the U.S. Department of Education. However, the contents do not necessarily reflect the views of the USDOE, and no endorsement is implied.

(Popham, 1985). As examples, student surveys disclose pupil judgments about a teacher's provision of an opportunity to learn (Aleamoni, 1981) and peer review of instructional materials provides a check on curriculum quality (French-Lazovik, 1981). Variable lines of evidence means that data sources should differ for individual teachers; no one line of evidence is appropriate for all practitioners. Since components of teacher excellence vary, documentation should be sought for performance dimensions in which teachers have an impact. The innovation of variable lines of evidence permits greater use of data sources that are controversial. For example, standardized tests for teachers are doubtful indicators of quality for *all* experienced teachers, but may serve to highlight the high verbal knowledge of *some* practitioners. Variable lines of evidence enable parent surveys to be used for some teachers in an evaluation system without penalizing unfair cases where data about parents are unavailable, not pertinent, or misleading. In a number of ways, using variable lines of evidence is a kind of "goal free" evaluation described by Scriven (1973).

The arguments for including parent views in systematic teacher evaluation are considerable. Epstein (1985) pointed to both the membership of parents in the school organization and to the shared responsibility of teachers and parents for student learning. Peterson and associates, (1985) added other reasons to include parents in teacher evaluation systems. First, parents *are* now involved in teacher evaluation. At present, parent input consists of haphazard, unreliable single case reports, and hearsay. Teachers are at risk when information is unsystematically gathered and reported. Second, parents are clients and taxpayers. The rights of consumers have been established in current evaluation practice (Scriven, 1981). Finally, and perhaps most important, parents are in a position to know about the teacher's performance in relation to their own children and in their (the parents') interactions with the teacher.

Some arguments for parent participation are formalized in the evaluation concept of *stakeholder evaluation*. Mark and Shotland (1985) described two dimensions of stakeholders in evaluation: power and legitimacy. Stakeholders are involved in evaluation in the sense of power because of their capacity to affect the enterprise. Groups may be included in the second sense of legitimacy because they have substantive contributions to add to the process or outcome. In either event, the decision to include stakeholder groups in evaluation is a value judgement to be made by the evaluators. Problems may arise from a lack of careful involvement if the only purpose is preemptory involvement (e.g., to avoid criticism) or pseudo-empowerment that achieves a diversionary facade of participation. Mark and Shotland provided guidelines for considerate and effective stakeholder participation. They emphasized that sufficient time be set aside, training provided, contributions acknowledged, and limits to the involvement clearly set.

There are substantial arguments against parent participation in teacher evaluation (Peterson & Kauchak, 1982). Most teachers are familiar with instances of parent influence based upon hearsay and unrepresentative episodes. Mere popularity with parents is a positive consideration but not compelling in determining teacher excellence. While in many cases parents are expert in raising their own

children, they are not expert at the tasks of classroom teaching. There are substantial problems with comparing parent views of teachers of different age students because of the large variation in parent interests, e.g., between kindergarten and high school age students.

Lortie (1975) described a number of sociological concerns about parent-teacher relations that complicate the evaluation procedure. Essentially, parents and teachers have different points of view that cloud the relationship. For example, parents see the child as a prized individual, while the teacher must see her or him as a member of a class group. Other examples of difficulties in parent-teacher relationships are that teachers resent the power to disrupt possessed by parents, while parents envy the authority and influence possessed by teachers.

Despite these considerable reservations about parent involvement in teacher evaluation, a great need remains to explore all possibilities for additional data about teacher performance. Epstein (1985) reported one of the few empirical studies of parent surveys for public school teacher evaluation. She found a correlation of .27 between principal and parent ratings. Epstein concluded that parents and principals base their judgments on different dimensions of teacher performance. Epstein reported that parent ratings are influenced by student classroom experiences and by resources and ideas offered to parents by the teachers. Among other findings was that the variation among ratings increased with the grade level. Epstein concluded that "parents can make valuable contributions in the evaluation of teachers" (p. 3) because they have an important perspective on performance and provide another view of excellence.

Little additional empirical research and development have been reported on parent involvement in teacher evaluation. For example, the *Handbook of Teacher Evaluation* (Millman, 1981) has no section on parent participation. In 1985 Epstein stated that "we have no practical examples of how multiple judges, including parents, are involved systematically in the evaluations of teachers" (p. 9).

The Study

The purpose of this study was to investigate a practical use of parent surveys in a teacher evaluation system that employs them as an optional data source. Specific questions addressed frequency of use by teachers, return rates by grade level, descriptive statistics of ratings, variance in ratings by age of student and sex of teacher, differences in variation by grade level, stability of ratings over two years, effects of written comments, and costs of parent surveys.

Sample

Volunteer teachers in two Utah school districts participated in this study. One district was mid-sized (635 teachers, 29 schools) and located in a suburban-rural setting. The second was small (66 teachers, three schools) and located in a ski resort

town. Approximately two-thirds of the teachers in each district participated in an evaluation system as part of a career ladder program. The evaluation system was designed to enable teachers to be promoted and receive additional pay for demonstration of quality teaching performance. Parent survey data on 337 K-12 teachers were included in this study.

Evaluation and Promotion System

Both districts used multiple and variable lines of evidence in their evaluation systems (Peterson & Mitchell, 1985; Peterson, 1987). Teachers who had three or more years of successful experience were eligible to apply for promotion to a level of "Associate Teacher." Teachers selected four or more lines of evidence of quality performance from student reports, peer review of materials, teacher tests (National Teachers' Examinations), documentation of professionalism, administrator reports, student achievement data, systematic observation, and parent surveys. Parent surveys were requested by participating teachers. Forms were distributed and returned by mail to parents and scored by aides or research assistants hired to be independent of teachers, parents, and administrators. Three class surveys were required for junior high and high school teachers. Teachers were permitted to inspect result summaries, and to decide whether to include them in promotion deliberations. Teacher data were assembled in dossiers limited to 15 pages and judged by Promotion Panels. The Panels consisted of four teachers, two administrators, and two parents. The rewards for demonstration of quality as evidenced in the dossier were promotions for three years and \$1,100 per year in one district, and five years and \$1,500 per year in the other.

Parent Survey Instrument

The parent survey instrument was developed by the Utah Teacher Evaluation Project at the University of Utah based on current research literature. Items were added by the League of Women Voters of Salt Lake and the forms refined by the Utah State Parent-Teachers' Association (Peterson et al., 1985). The summative promotion decisions were made on the class mean rating on the global item: "Were you satisfied with your daughter or son's overall classroom experience as provided for by the teacher?" (Peterson, Gunne, Mitchell, & Rivera, 1984).

The parent rating form was one page in length and contained two sections. The first section estimated parental interest by having respondents check whether they had asked for information about (1) the class goals, (2) student progress, or (3) suggestions for home support of learning. Next to these items were spaces to check (1) if the parent had requested the information and (2) if the teacher had provided this information. This first section permitted an estimation of the extent to which parents were involved in information exchanges. Parents of younger students were anticipated to be more inquisitive of class and student information. The general

expectation for teachers was that they provided more information than they were asked for. For example, if six parents asked for progress information, the expectation would be for more than six to have indicated information received.

The second section of the parent survey instrument solicited three views about teacher performance from parent perspective. First, did the student seem to know what was expected of him or her. Confusion or confidence may be seen by parents. Second, were the challenges of the class appropriate. Classes reported by students as too easy to too hard can be noted by parents. Third, parents were asked for a global rating of their satisfaction of the classroom experience as provided for by the teacher. The survey instrument concluded with a space for comments.

The parent survey form used in this study differed from that of Epstein (1985) in that it addressed only information for which the parent could be expected to have first-hand information. It did not use Epstein's topics of classroom lessons, knowledge of subjects, discipline, and creativity for which the parent would likely have knowledge only second hand from their children, if at all.

Data and Analysis

Data collected and analyzed in this study pertained to a range of exploratory questions. The first analysis was a tally of numbers of teachers electing to use parent survey data. Additional analysis addressed descriptive statistics and graphic presentations of the distribution of class mean parent ratings on the global item. Ratings were further described according to grade level of student and sex of teacher.

Because parent interest in classrooms and teacher evaluation varies according to the grade level of the student, return rates for surveys were analyzed. This consisted of an analysis of variance in the independent variable of return rates according to four grade levels: primary (K-3), upper elementary (4-6), junior high (7-8), and high school (9-12). The unit of analysis and dependent variable was a class return rate percentage. The purpose of this analysis was to provide perspective for summative decision makers.

A correlation was computed between two consecutive years of parent rating data. The purpose of this analysis was to estimate stability of the measure across one year and different classes. A correlation was computed for 28 teachers who gathered parent survey data for two successive years.

In addition to the question of stability of ratings, there was an interest in possible changes as a result of teachers receiving parent feedback data. For example, student surveys have been found to foster improvement (Braunstein, Klein, & Pachio, 1973). A dependent t-test was performed to examine changes. In addition, differences between two years of data were examined in a larger group that included the 23 additional teachers who collected data in only the first or second year. This second comparison ($n = 51$) was made with an independent t-test (unequal variances). The unit of analysis and dependent variable for both analyses was a class mean on the global item.

The variance in parent ratings was analyzed by two-way ANOVA for the inde-

pendent variables of grade level of student and sex of teacher. The unit of analysis and dependent variable was a class mean on the global item. The purpose of this analysis was to determine the statistical significance of observed differences among grade levels and between teacher sexes. Significant differences might require different standards for judgments based on parent survey data.

Finally, two supplementary analyses were performed to provide information essential to establishing and maintaining sophisticated evaluation systems. Cost-benefits data included money and time costs of parent surveys for teachers, administrators, parents, and evaluators. A stratified random sample of 30 teachers was interviewed with the open-ended question "what worked and what didn't work" for reactions to the process and results.

Findings

Parent surveys were used by 337 of the 701 (48.1 percent) teachers in two districts in which the instrument was employed as an elective data source for promotion reward decisions. However, not all of the teachers were eligible for the promotion system or chose to participate in it. Teacher scores in this analysis represent the class mean for elementary school teachers and the mean of three class means for junior high and high school teachers. Thus this analysis represents approximately 12,000 individual surveys. Teacher means on the global item ranged on a five-point scale (5 most positive) from 5.00 to 1.50 for the combined sample of 337 teachers in both districts. The sample had a mean rating of 4.49 (S.D. = .39). Figure 1 depicts the distribution of ratings.

Parent ratings differed slightly according to the sex of the teacher. Male teachers ($n = 130$) received a mean rating of 4.44 (S.D. = .30), while females ($n = 122$) received a mean rating of 4.61 (S.D. = .33). The number of subjects in this analysis was lower than the total using parent surveys because not all descriptive data were available on all result summaries.

Parent ratings differed according to the grade level taught by the teacher. Mean teacher ratings by four grade levels (primary, upper elementary, junior high, high school) are presented in table 1. The variation in parent ratings increased from primary grades to junior high but then declined in high school.

Differences in parent ratings according to the sex and grade level of the teacher were examined in the larger district sample with a two-way analysis of variance. The difference in scores according to sex was found not to be statistically significant ($F = 2.72$; $df = 1, 251$; $p = .10$). Differences in scores according to grade level were found to be statistically significant ($F = 14.76$; $df = 3, 251$; $p < .01$). Pairwise contrasts (Student-Newman-Keuls) showed significant differences between each pair of grade levels except for primary and upper grades. The interaction between sex and grade level was found not to be significant ($F = .07$; $df = 2, 251$; $p = .93$). However, the empty cell of males and primary grades confounded this analysis. The combined independent variables (R^2) explained 21.5 percent of the observed variance. The adjusted betas for teacher sex and grade level were .11 and .41, respectively.

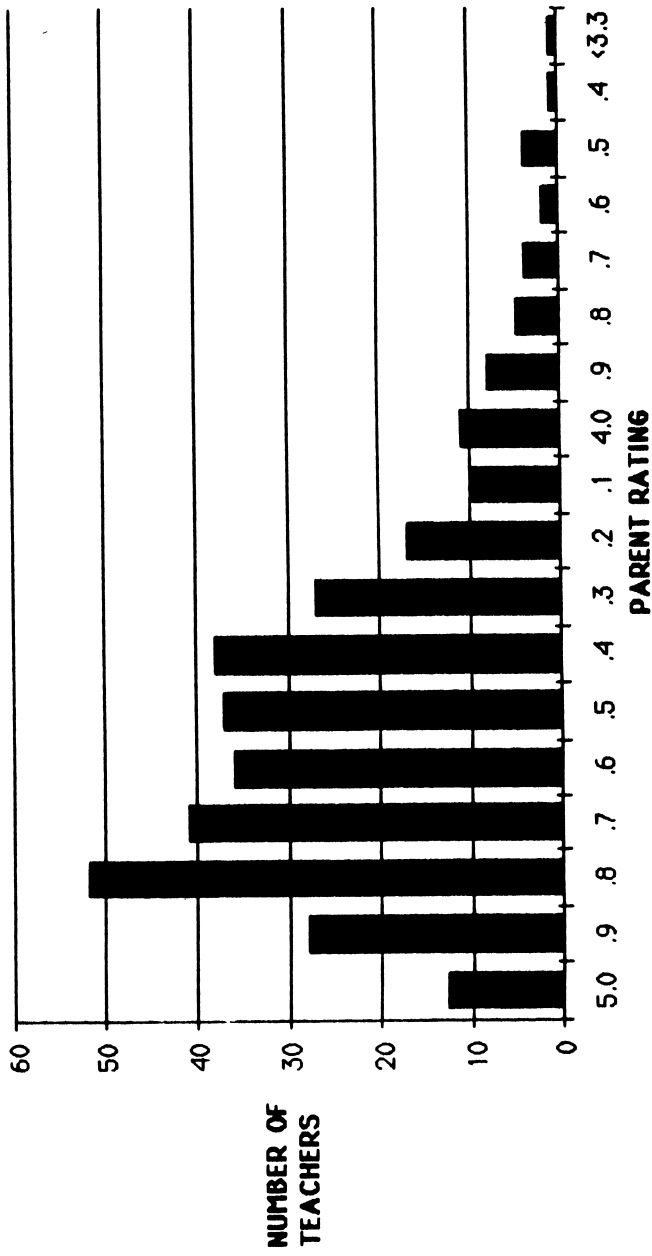


Figure 1. Parent Ratings of Teachers on Global Item.

Table 1. Parent ratings by grade level.

	N	Mean	S.D.
Primary	49	4.72	.18
Upper	59	4.63	.23
MS, JHS	80	4.35	.37
High Sch	64	4.48	.26

The mean parent survey return rate for the larger district sample of 178 teachers was 66.8 percent (S.D. = 22.6 percent). However, as shown in table 2, the return rates varied considerably by grade level. An analysis of variance of return rate by four grade levels (primary, upper elementary, junior high, high school) showed statistically significant differences among groups ($F = 69.57$; $df = 3, 177$; $p < .01$). Pairwise contrasts (SNK) identified significant differences between each pair of groups except for primary and upper grades.

A correlation was computed for the 28 teachers who gathered data over a two-year period. The product-moment correlation (r) was 0.414 ($r^2 = 0.17$). For comparative perspective, the same teacher population student report stability for 51 teachers was $r = 0.673$ ($r^2 = 0.45$).

Table 2. Percent parent form returns

	N	Mean	S.D.
Primary	35	86.2%	.10
Upper	37	85.3%	.12
MS, JHS	57	60.8%	.15
High Sch	49	46.0%	.20

Statistical tests for differences between ratings in two successive years showed nonsignificant changes in the global item rating between first and second years of parent survey usage. The comparison of 28 teachers rated in two years showed a statistically nonsignificant ($t = .03$; $df = 27$; $p = .99$) increase, while the comparison of 51 teachers (first year $n = 48$; second year $n = 37$) showed a statistically nonsignificant ($t = -1.08$; $df = 83$; $p = .14$) decrease.

Parent surveys cost approximately 44 cents per student for mailing and \$2.50 for aide time, for an approximate cost of \$15 per class. Time required was approximately five minutes for teachers and 20 minutes for the aide per class.

The open-ended nature of the interview questions produced suggestions for changes and information about problem applications (e.g., results returned later than promised). Parent surveys were spontaneously reported as valuable feedback for teachers by 76.6 percent of the sample. The items not analyzed in this study nor used in the summative judgments were seen as helpful formative information. Interviews with teachers concerning the parent surveys suggested that teachers found the comments to be as helpful or more helpful than the global item ratings for understanding parent reactions to their classes.

Discussion and Conclusions

Parent surveys were selected by a significant number of teachers in this study. In addition, the survey results showed sufficient range variation to permit discriminations among teachers. Both the levels of usage and discrimination suggest that parent surveys present an important additional data source for teacher evaluation.

Grade level differences were found to be statistically significant, both in return rates and in how positive parents were. These findings suggest that parent survey results should be examined separately by grade level. Differences for female and male teachers were not found to be significantly different.

The variance in parent ratings increased from primary to upper grade to junior high school but declined in high school. This finding corroborates that of Epstein (1985) for the age levels she studied, but raises questions about parental views of teachers in the high school grades. Perhaps less contact and communication results in more global or halo ratings for the teachers of older students.

The correlation of ratings between years was moderate. One implication of this finding is that data of several years are needed to establish stable patterns of teacher performance. A preliminary recommendation might be for three years. With these data, some teachers may have stable ratings while others do not.

No statistically significant changes were found between two years of parent surveys. The sample teachers did not have inservice or formal discussions of raising ratings. Perhaps inservice or a longer period of time would result in higher ratings by parents.

This analysis of results of teachers who elected to gather and report data cannot be generalized to all teachers. However, the optional use of parent data provided a unique opportunity to collect and study such information. The safe setting created by two levels of teacher choice (electing the line of evidence, inspecting the results) perhaps was responsible for the general acceptance of the procedure. Those teachers with objections or demonstrable handicaps in parent rating procedures were not penalized in the promotion judgment system. It is likely that the monetary reward was another reason that teachers elected to collect evaluation data, including parent surveys.

Teachers reacted positively to parent comments in this study. Lortie (1975) provided a perspective on such reactions. Parent surveys in this study systematized what before had been hearsay. The strong, supportive comments were well received by teachers who described them as authoritative reassurance. In follow-up interviews teachers described parent comments as pertinent, specific, and ultimately credible. Further survey or interview studies are warranted to better understand how teachers use parent survey data.

Dollar costs for parent surveys can be considerable if mail is used for all forms, as was done in this study. These costs may be necessary for initial establishment of parent perceptions. However, costs can be lessened by teachers who have established a stable pattern by the following: gathering data only three times in five years, sampling classes, and scoring only the global item. Class sampling is appropriate where the return rates are high. One strategy is to score the global-item only if this

item is above a "cut-in" one standard deviation below the district mean. For example, in this sample only 38 of the 373 (about 10 percent) would have the whole form scored if this strategy were used.

Parent surveys demonstrated their potential as a resource in a multiple and variable lines of evidence evaluation system. Unanswered questions in this study included: analysis of formative items, additional items that are useful in questionnaires, lower limits for sampling reliability, and stability of ratings over time. This study did not include a detailed look at teachers' attitudes and thinking about parent response data. Certainly further studies of parent surveys for school teacher evaluation are warranted.

References

- Aleamoni, L.M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Braunstein, D., Klein, B., & Pachio, M. (1973). Feedback expectancy and shifts in student rating: Reliability, validity and usefulness. *Journal of Applied Psychology*, 58, 254-258.
- Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Educational Commission of the States, (1983). *A summary of major reports on education*. Denver: Educational Commission of the States.
- Epstein, J.L. (1985). A question of merit: Principals' and parents' evaluations of teachers. *Educational Researcher*, 14, 3-10.
- French-Lazovik, G. (1981). Peer review: Documentary evidence in the evaluation of teaching. In J. Millman (Ed.), *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Lewis, A.C. (1982). *Evaluating educational personnel*. Arlington, VA: American Association of School Administrators.
- Lortie, D.C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Mark, M.M., & Shotland, R.L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review*, 9(5), 605-626.
- Millman, J. (Ed.) (1981). *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Peterson, K. (1984). Methodological problems in teacher evaluation. *Journal of Research and Development in Education*, 17(4), 62-70.
- Peterson, K.D. (1987). Multiple and variable lines of evidence in teacher evaluation. *American Educational Research Journal*, in press.
- Peterson, K., & Kauchak, D. (1982). *Teacher evaluation: Perspectives, practices and promises*. Salt Lake City: Center for Professional Practice, University of Utah. (ERIC Document Reproduction Service No. ED 233 996).
- Peterson, K., & Mitchell, A. (1985). Teacher controlled evaluation in a career ladder program. *Educational Leadership*, 43(3), 44-47.
- Peterson, K., Gunne, M., Mitchell, P., & Rivera, O. (1984). Multiple audience rating form strategies for student evaluation of college teaching. *Research in Higher Education*, 20(3), 309-321.
- Peterson, K., Kauchak, D., Mitchell, A., McCarthy, S., & Stevens, D. (1985). *Utah Teacher Evaluation Project: The Park City career ladder design*. Salt Lake City: University of Utah. (ERIC Document Reproduction Service No. ED 265 143).
- Popham, W.J. (1985). The evaluation of teachers: A mission ahead of its measures. Invited address at the American Educational Research Association, Chicago, Illinois, April.

- Scriven, M. (1973). Prose and cons about goal-free evaluation. *Journal of Educational Evaluation*, 3(4), 1-4.
- . (1981). Summative teacher evaluation. In J. Millman (Ed.) *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Stodolsky, S.S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13, 11-18.