

What Do K-12 Teachers Think About Including Student Surveys in Their Performance Ratings?

Mid-Western Educational Research Association 2014 Division L Distinguished Paper

Beverly J. Dretzke
Timothy D. Sheldon
Alicia Lim
University of Minnesota

This study investigated K-12 teachers' opinions about the use of student surveys as a component of a teacher evaluation system. Surveys were administered to teachers at the beginning of the school year and again in the spring. Analyses of teachers' responses on the fall survey indicated tentative support for the inclusion of student feedback in teacher performance ratings, as well as some uncertainty. Teachers of lower elementary grades were more skeptical than teachers of other grade levels regarding the use of student survey data and were more doubtful about their students' ability to distinguish between effective and ineffective teaching behaviors. Comparisons of fall and spring responses showed a consistent decrease in teachers' support for the use of student surveys regardless of grade level taught. In addition, teachers were more skeptical than principals with respect to both the validity of student survey data and the usefulness of student feedback for improving teaching performance.

Throughout the United States, increased attention is being given to the systems utilized to evaluate K-12 teachers. Traditional evaluation systems are criticized because they typically use only two categories for teacher performance (satisfactory and unsatisfactory) and offer little useful feedback to teachers in either category on how they could improve (e.g., Hull, 2011, 2013; Weisberg, Sexton, Mulhern, & Keeling, 2009). Accordingly, educational researchers and policy analysts are encouraging states to use teacher evaluation systems comprised of multiple data sources that not only result in differentiations among teachers based on their effectiveness in enhancing student achievement, but also identify areas where teachers might benefit from professional development (e.g., Glazerman et al., 2011; Peterson, 2004; Peterson, Wahlquist, Bone, Thompson, & Chatterton, 2001; Weisberg et al., 2009). Similarly, Race to the Top, a Federal grant program, promoted the use of teacher evaluation systems that include multiple measures and multiple rating categories and allow for timely, constructive feedback (U.S. Department of Education, 2014).

A 2013 study by the Center for Public Education reported that 41 states now require or recommend the use of multiple measures to evaluate public school teachers (Hull, 2013). Student achievement gains and classroom observations are the measures most commonly utilized.

However, an increasing number of states are also incorporating the results of student perception surveys. The National Council on Teacher Quality reported that, as of 2013, 12 states require or allow for student surveys in their teacher evaluation models (Doherty & Jacobs, 2013).

Consequently, a growing number of educational researchers are carrying out investigations of student perception surveys utilized in elementary and secondary school settings. Goe, Bell, and Little (2008) reviewed research on approaches to evaluating K-12 teachers and concluded that student ratings could provide reliable and valid information regarding teacher performance. They also noted that a strong argument for the use of student perception data is based on students having more direct and extensive experience with a teacher than principals or other potential evaluators.

The Measures of Effective Teaching (MET) project funded by the Bill and Melinda Gates Foundation utilized student perception surveys, along with prior achievement and classroom observations, to predict students' achievement, and found that the combination of data sources not only increased prediction accuracy but also provided meaningful feedback that would help school administrators identify and prioritize teachers' professional development needs (MET Project, 2013). In his dissertation research, Balch (2012) investigated the validation of a newly developed student survey and gathered information on how teachers used feedback from student surveys to inform their instruction. The teachers in his study said they used their survey feedback reports to identify areas for improvement and frequently worked with their colleagues to develop appropriate strategies. Based on his findings, Balch suggested that teachers will likely become more invested in student surveys as a component of teacher evaluation when they see their summary feedback reports and realize their potential value for enhancing teaching effectiveness.

Schulz, Sud, and Crowe (2014) conducted interviews with school leaders and state officials to gather information on the experiences of early adopters of student perception surveys. The early adopters included the districts participating in the MET project, charter management organizations, and teacher preparation programs. Their investigation identified two main challenges related to the use of student surveys in teacher evaluation systems. One of these challenges was gaining teacher buy-in and support. Interviewees often reported that teachers expressed initial skepticism and resistance to student surveys related to their perceptions that student surveys were simply popularity contests. The second challenge was making effective use of student feedback to improve teacher practice. The researchers noted the importance of showing connections between implementation of a student survey and improving teacher practice.

Teacher buy-in is critical for the success of the new generation of evaluation systems (White, Cowhy, Stevens, & Spote, 2012). Teachers' perceptions of the validity of the various components of an evaluation system influence their acceptance of the components and ultimately their support for the entire system. Very little research, however, has been conducted on K-12

teachers' views of the systems that are used to evaluate their performance. One study that did investigate K-12 teachers' attitudes toward evaluation practices was carried out by Kauchak, Peterson, and Driscoll (1985). Kauchak and his colleagues interviewed and surveyed teachers about a variety of evaluation methods, one of which was the use of student evaluations. Their results indicated that, overall, teachers were equally divided among three different views regarding student evaluations. The most positive third felt that student evaluations could provide valuable information about a teacher's performance but that professional judgment should be exercised when interpreting the results. The middle third were quite skeptical about the use of student evaluations and indicated that, if used, student feedback should be interpreted with a great deal of caution. The least positive third openly opposed the use of student evaluations and doubted that students could provide valid or reliable information about teaching performance. When the responses were examined by teaching level, it was found that elementary school teachers were more likely than secondary teachers to express negative views regarding the use of student evaluations. Teachers with negative views stated that it was doubtful that students, especially younger ones, could understand the complexities of teaching and differentiate between competent and incompetent teacher behaviors.

The results of research reported by Schwab and Iwanicki (1988) are consistent with those of Kauchak and his colleagues (1985). Schwab and Iwanicki investigated teachers' attitudes regarding a performance-based salary program. Overall, 41% of the teachers who participated in their study indicated that they did not support the inclusion of student feedback in the evaluation process, 47% supported the inclusion of student feedback with some reservation, and only 11% strongly supported the inclusion of student feedback. Schwab and Iwanicki also found that high school teachers indicated more support for the use of student feedback than junior high or elementary school teachers.

The research reported in this paper was carried out as one component of an extensive study of the teacher evaluation model developed by the Minnesota Department of Education (MDE) and piloted in the 2013-14 school year (Dretzke et al., 2015). In 2011, MDE was charged with convening a work group to consult with the Commissioner of Education to develop a state model for teacher growth and development. Recognizing the need for a comprehensive teacher evaluation system based on multiple data sources, the Minnesota statutes stated that the model must include the following components: student engagement measures, student growth measures, an individual growth and development plan prepared by each teacher, peer review, and a summative evaluation by a qualified and trained evaluator (Minnesota Statutes Education Code, 2014). The statutes required all districts to begin evaluating teachers based on these criteria beginning in the 2014-15 school year. The state model was to serve as an example model for local districts.

MDE contracted with the Center for Applied Research and Educational Improvement at the University of Minnesota to evaluate the pilot of the state model. The authors of this paper were members of the University of Minnesota research team. The evaluation of the Minnesota state model included fall and spring interviews with teachers and summative evaluators and fall and

spring surveys administered to teachers and summative evaluators. The summative evaluators were the individuals who would formally evaluate teachers and assign a rating to their performance. Only one component of the evaluation is presented in this paper. The primary purpose of the research reported here was to investigate teachers' attitudes and beliefs regarding the use of student survey data in teacher performance ratings. Two main questions were addressed:

1. How did teachers feel about the use of student surveys before student perception surveys were administered in their districts?
2. Did their opinions change by the end of the process when the student survey results had been received and reviewed with their summative evaluators?

Analyses were carried out to answer additional questions regarding whether or not teachers' opinions differed by grade levels taught, subjects taught, or years of teaching experience. Teachers' and summative evaluators' opinions regarding the validity and usefulness of student surveys were also compared. Because most of the summative evaluators were principals or assistant principals, summative evaluators are referred to as principals in this paper.

Method

Participating Districts

During the 2013-14 school year, a total of 16 Minnesota school districts participated in the pilot of the student engagement component of MDE's teacher evaluation model. Student engagement was the component in which students' perceptions of classroom instruction would be measured by means of a student perception survey. Most of the participating districts were located in rural Minnesota. Student enrollments in the districts ranged from 202 to 7,510 and the number of teachers per district ranged from 18 to 391.

Instruments

Teacher questionnaires. Two questionnaires were developed by the research team that were designed to elicit teachers' opinions regarding the use of student surveys to evaluate teacher performance. The pre-implementation questionnaire was administered at the beginning of the school year before classes began and before teachers were introduced to the student survey instrument. The post-implementation questionnaire was administered in late spring near the end of the school year after teachers were expected to have received a summary report of their students' survey results. In this paper, these two teacher questionnaires are referred to as the fall survey and the spring survey, respectively. On both the fall and spring surveys, teachers were asked to indicate their level of agreement with each of 12 statements regarding student surveys (e.g., *I am in favor of using feedback from student surveys to evaluate my teaching; I am concerned that students will give low ratings in a subject that is difficult for them*). A seven-point response scale was utilized (*strongly disagree, disagree, somewhat disagree, neither disagree nor agree, somewhat agree, agree, strongly agree*). The content of the items was derived from literature on student perception surveys utilized in K-12 settings (e.g., Burden, 2010; Burniske & Meibaum, 2012; Hanover Research, 2013; Kauchak et al., 1985; MET Project, 2013; Peterson,

Wahlquist, & Bone, 2000; Peterson et al., 2001; Schwab & Iwanicki, 1988), feedback from educators, and feedback from MDE staff. Teachers were also asked to provide background information regarding grade level(s) taught, subject area(s) taught, and number of years of teaching experience. On the spring survey (but not the fall survey), teachers were asked how they would rate student surveys with respect to generating information for two intended outcomes of the evaluation model: a) assigning a valid rating to a teacher's performance, and b) identifying areas where a teacher could benefit from professional development. These items were included only on the spring survey because it was expected that teachers would have received the student survey feedback reports by the time the spring survey was administered and we wanted their responses to be informed by their student feedback reports.

Principal questionnaire. Only two items on the principal questionnaire, administered in the spring, were included in analyses carried out for this paper. With the exception of minor differences in wording, these were the same two items presented on the spring teacher survey that asked respondents to rate student surveys with respect to generating information for assigning a valid rating to a teacher's performance and identifying areas where a teacher could benefit from professional development.

Survey administration. The teachers' fall survey was administered as a paper/pencil instrument at the beginning of the school year, from late August to early September. The fall questionnaires were either hand delivered to the districts or were sent by mail, and were administered as the first activity in training sessions that provided teachers with an overview of the state model. Completed questionnaires were placed in envelopes, sealed, and returned directly to the University of Minnesota researchers. The spring teacher and principal surveys were administered online near the end of the school year in May and June. To maintain confidentiality, the links for the online surveys were sent by the researchers directly to the teachers' and principals' email addresses. The decision to change from a paper/pencil questionnaire to an online survey for teachers was made because teacher training sessions were not being conducted at the pilot sites near the end of the school year. We felt that the use of online surveys sent directly to teachers' email addresses would ensure that every teacher would receive the survey and all reminder notices. The response rate may have been affected by the change in survey administration, however, since the paper/pencil fall survey was completed as part of a beginning-of-the-year structured professional development session that was attended by nearly all teachers, whereas the online spring survey was completed when teachers found time in their individual schedules at the end of the school year.

Teacher Categorization Procedures

Grade-level bands. For analysis purposes, the teacher survey respondents were categorized according to six mutually exclusive grade-level bands that referred to the grade level(s) of their teaching assignments: grades K-2, grades 3-5, grades K-5, grades 6-8, grades 6-12, and grades 9-12. These bands were chosen primarily for the purpose of isolating the responses of teachers in the K-2 band since the student survey developed for the MDE evaluation model would not be administered to students in grades K-2. In addition, we wanted to see if our results would replicate those of previous research regarding differences between elementary and secondary teachers (Kauchak, et al., 1985; Schwab & Iwanicki, 1988), making it necessary for us to avoid

broad categories that spanned elementary and middle school grades (i.e., K-8) and elementary and middle/high school grades (i.e., K-12). The categories were mutually exclusive, so that a teacher could be included in only one grade-level band. For example, a teacher who taught only kindergarten was included in the K-2 band, a teacher who taught grades 3 and 4 was included in the grades 3-5 band, a teacher who taught grades 1, 2, 3, and 4 was included in the K-5 band, and so on. The six grade-level bands were further collapsed into three mutually exclusive grade-level bands: K-2, K-5, and 6-12. In this paper, these three grade-level bands are referred to as primary, elementary, and secondary, respectively.

Academic subject categories. Three categories were used for the academic subject(s) that the classroom teacher respondents taught: core/generalist, elective/specialist, and special education. These categories were mutually exclusive so that no teacher was placed in more than one category. “Core” was used primarily for middle and high school teachers and included literary arts, mathematics, science, and social studies. “Generalist” was used only for elementary teachers who taught all courses in the general curriculum. “Elective” included mostly middle and high school teachers who taught subjects such as agriculture, business, and world languages. “Specialist” was used for elementary teachers who taught visual arts, music, or physical education. The special education category was comprised of teachers at any level who indicated that their subject area was special education.

Fall Teacher Survey Results

A total of 22 out of 37 principals completed the survey for a response rate of 59%. A total of 791 out of 920 K-12 teachers responded to the fall survey, resulting in an overall response rate of approximately 86%. As described earlier, the teacher respondents were categorized according to the grade level(s) of their teaching assignments and the subjects that they taught. Respondents who did not provide grade level or subject information were excluded from the analyses, as were respondents whose teaching assignments fell into more than one of the grade-level bands. Following this procedure, 72 of the 791 cases were excluded from the data analysis. A total of 466 K-12 teachers responded to the spring survey, resulting in an overall response rate of 51%. Forty-four of these respondents were excluded because they provided answers only to items requesting demographic information (e.g., district, subjects taught) and omitted all other items. After excluding these and an additional 65 responses that did not fall into any of the mutually exclusive grade-level bands, the spring data set contained 357 cases.

Four different analyses were carried out on responses given on the fall teacher survey. First, agreement rates were calculated across grade-level bands and academic subjects. Second, pairwise comparisons were carried out to test differences among the mean ratings of the three grade-level bands. Third, within each grade-level band, the mean ratings of the three academic subject areas were tested for significance via pairwise comparisons. Fourth, the rating given for each item was correlated with years of teaching experience.

Agreement Ratings Across Grade-Level Bands and Academic Subject Areas

The 12 questionnaire items were grouped according to whether they expressed support for

the use of student surveys or expressed a concern. Agreement rates for items in these two groupings are reported in Tables 1 and 2.

Support for student surveys. Six of the 12 items expressed support for the inclusion of student surveys in a teacher evaluation system. These items and their agreement rates are displayed in Table 1. Between 50% and 78% of teachers agreed with the statements by selecting one of the three response options that indicated agreement (i.e., *somewhat agree, agree, strongly agree*). It should be pointed out that the most frequently chosen response option for five of the six items in Table 1 was *somewhat agree*. In addition, only slightly more than half of the teachers (53%) indicated that they were in favor of using feedback from student surveys to evaluate their teaching. These results suggest that, at the beginning of the school year before the student surveys had been administered, the teachers were expressing tentative rather than strong support.

Table 1

Agreement Rates for Questionnaire Items Expressing Support for Student Surveys

| Questionnaire item | Agreement rate |
|----------------------------------------------------------------------------------------------------------------------------------|----------------|
| • Student surveys will provide me with feedback that will enable me to identify strengths and weaknesses in my teaching methods. | 78% |
| • Student surveys will provide me with information that will enable me to improve my effectiveness as a teacher. | 78% |
| • Feedback from student surveys can help my school identify areas where teachers need professional development. | 59% |
| • Feedback from student surveys can be used to track changes in my teaching performance over time. | 53% |
| • I am in favor of using feedback from student surveys to evaluate my teaching. | 53% |
| • Surveying my students is worth the time it takes to administer the survey. | 50% |

Concerns about student surveys. Six of the 12 items expressed a concern regarding the use of student surveys in a teacher evaluation system. Four of the items expressed a concern in a negative manner (e.g., *I am concerned that students will give low ratings in a subject that is difficult for them*), and two of the items expressed the concern in a positive manner (e.g., *Students in the grades and subjects that I teach can discriminate between effective and ineffective teaching behaviors*). The six items and their corresponding agreement rates are displayed in Table 2. Overall, between 26% and 70% of teachers indicated agreement (*somewhat agree, agree, or strongly agree*) with the six statements, and the most common response selected for four of the six items was *somewhat agree*. Based on the levels of agreement associated with

these items, it appears that teachers' support for the use of student surveys was mixed with reservations related to uncertainty about the students' ability to provide valid feedback and the time it would take to administer the student survey.

Table 2

Agreement Rates for Questionnaire Items Expressing Concerns about Student Surveys

| Questionnaire item | Agreement rate |
|-----------------------------------------------------------------------------------------------------------------------------------|----------------|
| • I am concerned that students will give low ratings in a subject that is difficult for them. | 70% |
| • I am concerned that students will use the survey to get back at teachers. | 56% |
| • Students in the grades and subjects that I teach can discriminate between effective and ineffective teaching behaviors. | 42% |
| • It will take a lot of time for me to review the results of my students' surveys in order to make them useful. | 41% |
| • Responses that my students give on the student survey can accurately predict how well they do on learning performance measures. | 37% |
| • Student surveys are not worth administering because I can easily influence students so they give positive ratings. | 26% |

Grade-Level Band Comparisons

Means were calculated for the 12 questionnaire items based on the following rating scale and numerical codes: *strongly disagree* = 1, *disagree* = 2, *somewhat disagree* = 3, *neither disagree nor agree* = 4, *somewhat agree* = 5, *agree* = 6, and *strongly agree* = 7. For each questionnaire item, Bonferroni nondirectional pairwise comparisons were carried out on the item means of the three grade-level bands. The type 1 error probability was set at .05 for each comparison.

Support for student surveys. Statistically significant grade level differences were obtained on three of the six items that expressed support for student surveys. These items mentioned that feedback from student surveys would help identify strengths and weaknesses in teaching, would enable teachers to improve their effectiveness, and that surveying students was worth the time it took to administer the surveys. For all three items, the mean rating of primary teachers was significantly lower than that of both elementary teachers and secondary teachers (see Table 3). These results indicate that the primary teachers were more tentative in their support of the use of student surveys in teacher evaluation than either elementary or secondary teachers.

Table 3

Rating Means and Standard Deviations of Items Expressing Support for Student Surveys by Grade-Level Band

| Questionnaire item | Rating means and standard deviations by grade-level band | | | Statistically significant pairwise comparisons |
|----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|---------------------|---------------------|---------------------------------------------------|
| | Primary (K-2) | Elementary (K-5) | Secondary (6-12) | |
| • Student surveys will provide me with feedback that will enable me to identify strengths and weaknesses in my teaching methods. | 4.43 (1.48) | 5.05 (1.26) | 5.21 (1.17) | Primary < Elementary Primary < Secondary |
| • Student surveys will provide me with information that will enable me to improve my effectiveness as a teacher. | 4.52 (1.51) | 5.04 (1.26) | 5.18 (1.14) | Primary < Elementary Primary < Secondary |
| • Surveying my students is worth the time it takes to administer the survey. | 3.75 (1.44) | 4.32 (1.52) | 4.52 (1.41) | Primary < Elementary Primary < Secondary |

Concerns about student surveys. Statistically significant differences were obtained on four of the six items that expressed concerns about using student surveys for teacher evaluation (see Table 4). The concerns mentioned in these items were students using the survey to get back at teachers, students giving low ratings in subjects that are difficult for them, students not being able to discriminate between effective and ineffective teaching behaviors, and teachers influencing their students to give positive ratings. The results indicate that the possibility of students using the survey to get back at teachers or giving low ratings in difficult subjects was less a concern for primary teachers than for teachers of older students. In addition, although the means for all three groups were fairly low regarding teachers being able to influence their students to give positive ratings, this area was also of significantly less concern to primary teachers than teachers in the other two grade-level bands. On the other hand, primary teachers appeared to be more doubtful than teachers of older students that their students had the ability to tell whether or not teaching behaviors were effective.

Table 4

Rating Means and Standard Deviations of Items Expressing Concerns about Student Surveys by Grade-Level Band

| Questionnaire item | Rating means and standard deviations by grade-level band | | | Statistically significant pairwise comparisons |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|------------------|------------------|-----------------------------------------------------------------------|
| | Primary (K-2) | Elementary (K-5) | Secondary (6-12) | |
| <ul style="list-style-type: none"> I am concerned that students will use the survey to get back at teachers. | 3.69 (1.61) | 4.08 (1.45) | 4.93 (1.42) | Primary < Elementary Primary < Secondary Elementary < Secondary |
| <ul style="list-style-type: none"> Students in the grades and subjects that I teach can discriminate between effective and ineffective teaching behaviors. | 2.83 (1.43) | 3.72 (1.48) | 4.26 (1.46) | Primary < Elementary Primary < Secondary Elementary < Secondary |
| <ul style="list-style-type: none"> I am concerned that students will give low ratings in a subject that is difficult for them. | 4.58 (1.43) | 4.87 (1.40) | 5.13 (1.37) | Primary < Secondary |
| <ul style="list-style-type: none"> Student surveys are not worth administering because I can easily influence students so they give positive ratings. | 3.88 (1.39) | 3.43 (1.37) | 3.40 (1.38) | Elementary < Primary Secondary < Primary |

Academic Subject Comparisons

The same Bonferroni pairwise comparison procedure utilized for grade-level band comparisons was also used for comparing academic subject areas within each grade-level band. Only one item within each grade-level band was associated with statistically significant pairwise differences (see Table 5). Results for each grade-level band are described in the following paragraphs.

Table 5
Rating Means and Standard Deviation of Items with Statistically Significant Differences by Teachers' Academic Subject within Grade-Level Band

| Grade-level band | Questionnaire item | Rating means and standard deviations by teachers' academic subject within grade-level band | | | Statistically significant comparisons |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|---------------------|-------------------|----------------------------------------------------------------------------------|
| | | Core/generalist | Elective/specialist | Special education | |
| Primary | <ul style="list-style-type: none"> It will take a lot of time for me to review the results of my students' surveys in order to make them useful. | 3.95 (1.39) | 5.33 (1.00) | 3.42 (1.56) | Core/generalist < Elective/specialist Special education < Elective/specialist |
| Elementary | <ul style="list-style-type: none"> Student surveys are not worth administering because I can easily influence students so they give positive ratings. | 3.56 (1.39) | 3.27 (1.19) | 2.84 (1.21) | Special education < Core/generalist |
| Secondary | <ul style="list-style-type: none"> Students in the grades and subjects that I teach can discriminate between effective and ineffective teaching behaviors. | 4.32 (1.36) | 4.44 (1.39) | 3.61 (1.82) | Special education < core/generalist Special education < elective/specialist |

Primary. The item where significant differences were obtained for teachers in the primary grade-level band concerned the amount of time it would take to review the results of student surveys in order to make them useful. For this item, both the core/generalist and special education teacher means were significantly lower (indicating less agreement and less time) than the elective/specialist teacher means. This result is consistent with the large student load that specialist teachers often have. It is not uncommon for visual arts, music, and physical education teachers at this level to provide instruction to all students in a school, whereas generalists and

special education teachers typically provide instruction for single classrooms comprised of 25 students or less. Consequently, it is understandable that the primary elective/specialist teachers would anticipate that the review of their students' survey feedback would take a substantial amount of time.

Elementary. The questionnaire item that was associated with a significant pairwise difference for teachers in the elementary grade-level band concerned the influencing of students to give positive ratings. Although the means of all three academic subject groups were relatively low, the mean of special education teachers was significantly lower than that of core/generalists. Thus, it appears that special education teachers perceive themselves as being less able than core/generalists to influence the responses their students give on a perception survey.

Secondary. For secondary teachers, statistically significant pairwise differences were obtained on only the item concerning the students' ability to discriminate between effective and ineffective teaching behaviors. For this item, the mean rating of the special education teachers was significantly lower than the mean rating of both core/generalists and elective/specialists. These results indicate that special education teachers at the secondary level are more doubtful about their students' ability to provide valid teacher performance ratings than either teachers of core subjects or teachers of elective subjects.

Years of Teaching Experience

Pearson correlation coefficients were calculated to test for significant relationships between years of teaching experience and item rating. The correlations associated with seven items were statistically significant; however, all seven correlations were very weak, ranging in absolute value from 0.10 to 0.16 (see Table 6). Taken together, the results of the correlational analyses carried out on item agreement rating and years of experience indicate that teachers with more experience have a slight tendency to view the use of student surveys in teacher evaluations with more skepticism than teachers with less experience.

Table 6

Statistically Significant Correlations Between Item Ratings and Years of Teaching Experience

| Questionnaire item | Pearson correlation |
|----------------------------------------------------------------------------------------------------------------------------------|---------------------|
| • Student surveys will provide me with feedback that will enable me to identify strengths and weaknesses in my teaching methods. | – 0.16 |
| • Student surveys will provide me with information that will enable me to improve my effectiveness as a teacher. | – 0.12 |
| • I am in favor of using feedback from student surveys to evaluate my teaching. | – 0.16 |
| • Feedback from student surveys can help my school identify areas where teachers need professional development. | – 0.12 |
| • Feedback from student surveys can be used to track changes in my teaching performance over time. | – 0.10 |
| • Surveying my students is worth the time it takes to administer the survey | – 0.12 |
| • Student surveys are not worth administering because I can easily influence students so they give positive ratings. | 0.11 |

Changes in Teachers' Responses from Fall to Spring

The teacher survey was administered two times during the school year. The first time was in the fall, prior to the start of the school year, and before teachers were introduced to the specific components of the teacher evaluation model and the use of student surveys in their performance ratings. The second time was near the end of the year, in May, after the students had taken the student perception surveys and summary feedback reports had been made available to the teachers. The fall and spring survey instruments contained the same 12 items, designed to solicit teachers' opinions about using student surveys in a teacher evaluation system. Item ratings given by teachers at the beginning of the school year were compared to the ratings given by those same teachers at the end of the school year. Wilcoxon rank sum tests were used to test the significance of the pre-post mean differences. The type 1 error probability was set at .05 for each test. Results are displayed in Tables 7 and 8.

Table 7

Fall to Spring Change in Teachers' Ratings of Statements That Express Support for the Use of Student Surveys in Teacher Evaluation

| Questionnaire Item | <i>n</i> | Fall to Spring Change | | <i>z</i> | <i>p</i> | |
|--------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------|------------------------------------------------|----------|----------|-------------------------------------------------|
| | | % Agreeing ^a | % Selecting Lower Rating ^b | | | % Selecting Higher Rating ^b |
| I am in favor of using feedback from student surveys to evaluate my teaching. | 218 | Fall = 55% Spring = 45% | 49% | 28% | 4.16 | .000 |
| Student surveys will provide me with feedback that will enable me to identify strengths and weaknesses in my teaching methods. | 219 | Fall = 77% Spring = 66% | 46% | 22% | 4.42 | .000 |
| Student surveys will provide me with information that will enable me to improve my effectiveness as a teacher. | 219 | Fall = 78% Spring = 67% | 46% | 16% | 5.66 | .000 |
| Feedback from student surveys can be used to track changes in my teaching performance over time. | 219 | Fall = 51% Spring = 39% | 47% | 23% | 3.67 | .000 |
| Feedback from student surveys can help my school identify areas where teachers need professional development. | 219 | Fall = 60% Spring = 52% | 49% | 25% | 4.47 | .000 |
| Surveying my students is worth the time it takes to administer the survey. | 217 | Fall = 55% Spring = 48% | 47% | 26% | 3.29 | .001 |

^aPercent agreeing is aggregated across ratings of *somewhat agree*, *agree*, and *strongly agree*.

^bCalculations are based on the seven-point rating scale of *strongly disagree*, *disagree*, *somewhat disagree*, *neither disagree nor agree*, *somewhat agree*, *agree*, and *strongly agree*.

Table 8

Fall to Spring Change in Teachers' Ratings of Statements That Express Concerns Regarding The Use of Student Surveys in Teacher Evaluation

| Questionnaire Item | n | Fall to Spring Change | | z | p | |
|--------------------------------------------------------------------------------------------------------------------------------|-----|----------------------------|------------------------------------------------|-----|------|-------------------------------------------------|
| | | % Agreeing ^a | % Selecting Lower Rating ^b | | | % Selecting Higher Rating ^b |
| I am concerned that students will give low ratings in a subject that is difficult for them | 217 | Fall = 69% Spring = 75% | 26% | 38% | 2.21 | .027 |
| I am concerned that students will use the survey to get back at teachers | 217 | Fall = 60% Spring = 67% | 24% | 37% | 2.51 | .012 |
| Student surveys are not worth administering because I can easily influence students so they give positive ratings | 219 | Fall = 30% Spring = 32% | 32% | 42% | 2.23 | .026 |
| Students in the grades and subjects that I teach can discriminate between effective and ineffective teaching behaviors | 215 | Fall = 44% Spring = 35% | 44% | 27% | 3.32 | .001 |
| Responses that my students give on the student survey can accurately predict how well they do on learning performance measures | 217 | Fall = 35% Spring = 25% | 42% | 27% | 3.22 | .001 |
| It will take a lot of time for me to review the results of my students' surveys in order to make them useful | 217 | Fall = 39% Spring = 25% | 45% | 25% | 3.36 | .001 |

^aPercent agreeing is aggregated across ratings of *somewhat agree*, *agree*, and *strongly agree*.

^bCalculations are based on the seven-point rating scale of *strongly disagree*, *disagree*, *somewhat disagree*, *neither disagree nor agree*, *somewhat agree*, *agree*, and *strongly agree*.

Change in Support for Student Surveys

The six items displayed in Table 7 are statements that express support for the use of student surveys. The fall-spring differences for all six items were statistically significant, and all six reflected decreases in approval of using student surveys in teacher evaluations. The fall-spring differences between grade-level bands were also tested for each item, but none were statistically significant. These results suggest that after teachers acquired experience with student perception surveys, regardless of the grade levels they instructed, they became more skeptical regarding the usefulness of student survey feedback in a teacher evaluation system. It should be noted, however, that although the changes were statistically significant, the fall aggregated agreement rates ranged from 51% to 78% and the spring aggregated agreement rates ranged from 39% to 67%. Therefore, in general, the change in teacher opinion from fall to spring would best be described as becoming more skeptical and uncertain rather than becoming more strongly opposed to the use of student surveys.

Change in Concerns about Student Surveys

The six items displayed in Table 8 are statements that express concerns about student surveys. The fall-spring differences for all six of these items were statistically significant. The fall to spring difference for the first three items in Table 8 was an increase in agreement, indicating that, after the student surveys had been administered in their schools, the teachers became more concerned about students giving low ratings in difficult subjects and using the survey to get back at teachers, and more concerned about teachers being able to influence their students to give positive ratings. The change for the last three items in Table 6 was a decrease in agreement. The content of two of these items indicates that, from the beginning to the end of the school year, the teachers became less likely to believe that their students could discriminate between effective and ineffective teaching behaviors or that the student survey responses could accurately predict their academic achievement. The final item in Table 8 is concerned with the time it would take to review the results of the student survey. The statistically significant decrease associated with this item indicates that in the spring after they had received the student survey results, the teachers felt the review took less time than they initially thought it would. Grade-level band comparisons were also carried out, but similar to results regarding change in support for the use of student surveys, none of these were statistically significant.

Comparisons of Teachers' and Principals' Opinions

MDE contracted with a national nonprofit organization to collaborate with a group of Minnesota educators and survey design experts to develop two student perception surveys specifically for use by the Minnesota districts participating in the pilot. Two different versions of the survey were created, one for students in grades 3 through 6 and the other for students in grades 6 through 12. The surveys contained items designed to obtain students' perceptions of their classroom experiences regarding academic, behavioral, cognitive, and affective domains (e.g., *Our class stays busy and does not waste time* and *This teacher asks questions to be sure we understand*). The student perception surveys were administered in the spring to students in the districts participating in the pilot of the state model.

On the teacher and principal surveys administered in the spring, the respondents were asked their opinion about the state model student perception survey with respect to two potential purposes of the surveys: a) generating information that would enable a principal to assign a valid rating to a teacher’s performance and b) identify areas where a teacher might benefit from professional development. Teachers’ and principals’ responses to these two items were compared.

Figure 1 presents the distribution of responses given by teachers and principals regarding their opinion of the state model student survey for generating information that would enable a principal to assign a valid rating to a teacher’s performance. The results of a statistical test of the difference between the mean ratings indicated that the principals’ mean ($M = 2.50, SD = 0.94$) was significantly higher than that of the teachers ($M = 1.84, SD = 0.90$) with respect to the use of student surveys for this purpose [$t(263) = 2.65, p < 0.01$].

Figure 1. *Distributions of ratings given by teachers and principals to the state model student survey with respect to its generating information that enables a principal to assign a valid rating to a teacher’s performance.*

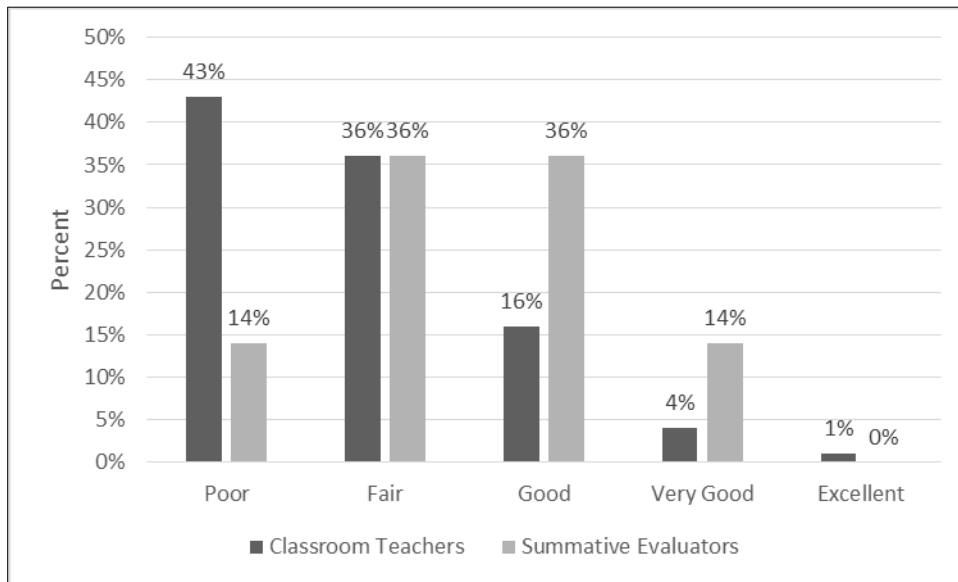
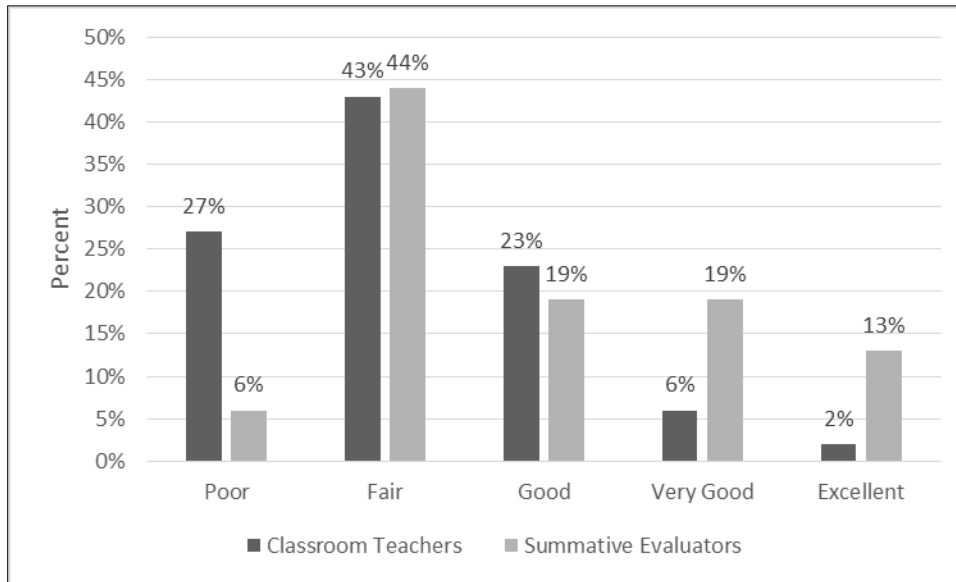


Figure 2 presents the distribution of responses given by teachers and principals regarding their opinion of the state model student survey for generating information that would enable a principal to identify areas where a teacher could benefit from professional development. Again, the principals’ mean ($M = 3.33, SD = 1.75$) was significantly higher than the teachers’ mean ($M = 2.14, SD = 0.94$), [$t(265) = 4.84, p < 0.0001$]. Taken together, these results indicate that principals see more value in the use of the state model student survey data than teachers with respect to assigning performance ratings to teachers and identifying areas where teachers might benefit from professional development.

Figure 2. *Distributions of ratings given by teachers and principals to the state model student survey with respect to its generating information that enables a summative evaluator identify areas for a teacher's professional development.*



Discussion

K-12 teachers in Minnesota districts participating in the pilot of a teacher evaluation system were surveyed in fall at the beginning of the school year and again in late spring at the end of the school year regarding the use of student survey feedback as a component of an evaluation system that would be used to rate their teaching performance. The purpose of the fall survey was to gather information regarding the teachers' initial opinions about the student survey before administration of the student survey had taken place and feedback reports had been distributed. The purpose of the spring survey was to investigate changes in opinion that might have occurred.

Overall, teachers' responses to the survey administered at the beginning of the year indicated that they perceived student feedback to be potentially valuable for providing information that would help them improve their effectiveness. However, their responses indicated that they were uncertain about their students' ability to provide valid feedback.

With respect to the grade level(s) of a teacher's position, the results of the fall survey analyses replicate the results of two previous investigations (Kauchak et al., 1985; Schwab & Iwanicki, 1988) that found higher levels of teacher support for student surveys as the teacher's grade level(s) increased. We found, for example, that teachers of students in primary grades gave significantly lower agreement ratings than secondary teachers to questionnaire items stating that student survey feedback would enable them to identify their teaching strengths and weaknesses or that student feedback would provide them with information that would help them to improve their teaching effectiveness. In addition, the fall survey responses indicated that teachers believed younger students were less able than older students to differentiate between more and less competent teachers. For example, the ratings provided by primary teachers were significantly lower than those given by teachers of elementary and secondary teachers with respect to agreeing

that their students had the ability to discriminate between effective and ineffective teaching behaviors. Although these findings provide research support for Minnesota's decision not to administer a survey to children until they were in third grade, it should be noted that teachers of older students were not completely without reservations. More specifically, secondary teachers were more likely than primary and elementary teachers to express concerns related to the possibility that students would use the evaluations as a way to get back at teachers or that students might give lower ratings in courses that were difficult for them.

The subject areas taught by the teachers also appeared to be related to their opinions about student surveys as a component in a teacher evaluation system. Fall survey responses of primary teachers (K-2) indicated that time was of more concern to specialists (i.e., visual arts, music, physical education) than it was to primary generalists or special education teachers. Admittedly, the student load for specialists can be very large and it is not surprising that specialists at lower grade levels would exhibit higher levels of concern regarding the time it would take to review and interpret student feedback and to develop strategies to address areas needing improvement. The responses of teachers of students in elementary grades (K-5) indicated that generalists were more concerned than special education teachers about a teacher being able to influence students so that they would give positive responses. At the middle and high school levels, special education teachers were more doubtful than teachers of core or elective subjects regarding their students' ability to distinguish between effective and ineffective teaching behaviors.

Years of experience appeared to have only a slight relationship with the teachers' responses. The fall results indicated that as years of experience increased, teachers tended to perceive student surveys as being somewhat less helpful for informing their instructional practice.

The results of analyses comparing responses given to surveys administered in the fall and then again in the spring were very consistent in showing a decrease in teachers' support of including student perception surveys in the evaluation system that would be used to rate their performance. Although their support had decreased, however, the teachers had not become strongly opposed to student surveys. Rather, regardless of the grade levels that they taught, teachers had in general become more skeptical and doubtful of the validity of student survey data and the usefulness of student feedback for improving their teaching performance. When compared to teachers, principals seemed to place more value on feedback from student surveys, both for the purpose of assigning a rating to a teacher's performance and for identifying areas where a teacher could benefit from professional development.

Because of the potential role that student surveys play in a teacher's performance rating, it is essential that the instruments demonstrate good psychometric properties. Accordingly, MDE is currently investigating its student perception surveys with respect to their reliability and the quality of individual items. In addition, survey validation is needed to provide evidence that the student surveys measure the construct of effective teaching behaviors. A validation study might include the same components included by Balch (2012) in his investigation of a student survey instrument that was utilized in a pilot teacher evaluation project in Georgia. Balch's investigation included analyses addressing content validity, convergent validity, and predictive validity.

As was stated in the introduction to this paper, teacher buy-in is critical for the success of a teacher evaluation system. Research on teacher evaluation reform suggests that buy-in is cultivated by allowing teachers to play an active role in the development of the evaluation system, by soliciting their input for a shared vision of quality instruction, and by communicating with them early and continuously about the system (White et al., 2012). The present study has given the teachers participating in Minnesota's pilot a systematic means of sharing their opinions and concerns regarding student surveys. The next step, of course, is for MDE to respond with a thorough investigation of the psychometric properties of its student surveys, identification of needed revisions, and the development of modified, improved instruments. The results and subsequent modifications should then be communicated to teachers.

The generalizability of this study is limited because it was conducted in only one state and school districts were not randomly selected. The study was carried out with the 16 districts participating in the student engagement component of that state's teacher development and evaluation model. However, the use of student perception surveys in teacher evaluation systems is of concern to many states, several of which might be in the process of refining their evaluation tools. Therefore, a number of important questions can be addressed by future research carried out in Minnesota as well as in other states. Do the student perception instruments and subsequent revisions demonstrate adequate reliability and validity? Do the teachers feel that the student survey feedback is accurate and meaningful? Do teachers' opinions about student surveys become more positive after their concerns and suggestions are addressed in revised instruments and feedback reports? How do teachers and principals use the survey feedback to improve instructional practice? Finally, and very importantly, do student perception surveys lead to enhanced teacher performance and improved student learning outcomes?

Author Notes

Beverly J. Dretzke is a Research Associate at the Center for Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota.

Timothy D. Sheldon is a Research Associate at the Center for Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota.

Alicia Lim is an Undergraduate Research Assistant at the Center for Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota.

Acknowledgements: Support for this research was provided by a grant from the Joyce Foundation awarded to the Center for Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota. We are grateful to the teachers and principals who participated in this study and to Tyler Livingston and Greg Keith at the Minnesota Department of Education for their assistance. We also wish to thank the three anonymous reviewers of the manuscript for their helpful suggestions.

Correspondence concerning this article should be addressed to Beverly J. Dretzke at dretz001@umn.edu

References

- Balch, R. (2012). *The validation of a student survey on teacher practice*. (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.
- Burden, P. (2010). Creating confusion or creative evaluation? The use of student evaluation of teaching surveys in Japanese tertiary education. *Educational Assessment, Evaluation and Accountability*, 22(2), 97-117.
- Burniske, J., & Meibaum, D. (2012). *The use of student perceptual data as a measure of teaching effectiveness*. Austin, TX: Texas Comprehensive Center. Retrieved from http://txcc.sedl.org/resources/briefs/number_8/bp_teacher_eval.pdf
- Doherty, K. M., & Jacobs, S. (2013). *State of the states 2013 connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Washington, DC: National Council on Teacher Quality.
- Dretzke, B. J., Ingram, D., Peterson, K., Sheldon, T., Wahlstrom, K., Baker, J., . . . Yap, S. (2015). *Minnesota State Teacher Development, Evaluation, and Peer Support Model Evaluation Report*. St. Paul, MN: University of Minnesota, College of Education and Human Development, Center for Applied Research and Educational Improvement.
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: Brown Center on Education Policy at Brookings. Retrieved from <http://www.brookings.edu/research/reports/2011/04/26-evaluating-teachers>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Hanover Research. (2013). *Student perception surveys and teacher assessments*. Washington, DC: Author. Retrieved from <http://dese.mo.gov/sites/default/files/Hanover-Research-Student-Surveys.pdf>
- Hull, J. (2011). *Building a better evaluation system: At a glance*. Alexandria, VA: Center for Public Education. Retrieved from <http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Building-A-Better-Evaluation-System>
- Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance*. Alexandria, VA: Center for Public Education. Retrieved from <http://www.centerforpubliceducation.org/teacherevalreview>
- Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers' attitudes toward teacher evaluation practices. *Journal of Research and Development in Education*, 19(1), 32-37.

- MET Project. (2013). *Ensuring fair and reliable measures: Culminating findings from the MET Project's three-year study*. Seattle, WA: Bill & Melinda Gates Foundation.
Retrieved from http://metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf
- Minnesota Statutes Education Code: Prekindergarten-Grade 12, Chapter 122A, §§ 40-41 (2014).
- Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin*, 88(639), 60-79.
- Peterson, K. S., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 135-153.
- Peterson, K. D., Wahlquist, C., Bone, K., Thompson, J., & Chatterton, K. (2001). Using more data sources to evaluate teachers. *Educational Leadership*, 58(5), 40-43.
- Schwab, R. L., & Iwanicki, E. F. (1988). *Merit pay: Stressor or motivator?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Schulz, J., Sud, G., & Crowe. (2014). *Lessons from the field: The role of student surveys in teacher evaluation and development*. Sudbury, MA: Bellwether Education Partners.
Retrieved from http://bellwethereducation.org/publication/Lessons_from_the_Field
- U.S. Department of Education. (2014). *State requirements for teacher evaluation policies promoted by Race to the Top: NCEE evaluation brief*. Washington, DC: U.S. Department of Education. Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York: The New Teacher Project. Retrieved from <http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>
- White, B. R., Cowhy, J., Stevens, W. D., & Spote, S. E. (2012). *Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts*. Chicago: The University of Chicago Consortium on Chicago School Research.