The mechanization of teaching: Teachers' metaphors and evaluation in Japanese tertiary education

Keywords
student evaluation, metaphors, teacher improvement, effective teaching

Twelve ELT university teachers reflected, through using metaphors, in interviews about the use of Student Evaluation of Teaching surveys (SETs) in their respective universities. Studying teachers' metaphor reveals their first-hand experience of how they were affected in their teaching by SETs. Metaphors suggest that SETs do not match teachers' conceptions of teaching as an art. Such evaluation has caused relations between teachers, administrators, and students to fracture due to competitive ranking. While participants accept formative evaluation as a necessary process to give insights to teachers, they wish for a more open, improvement-focused, cooperative, specific evaluation. They want more teacher involvement and more dialogue between teachers to discuss the results of SETs to aid the reflective process for change.

各大学での学生による授業評価（SETs）について、12名のELT担当の大学教師へのインタビュー調査を実施し、回答に用いられたメタファー（比喩）を分析した。その結果、SETsで各教師の実際の教え方にどのような影響があったかが明らかになり、SETsと教師側の「教える」という概念とは一致しないことが示唆された。このような評価は、競争的な順位付けをすることで教師側・大学当局側・学生側の関係を壊している。被験者の大学教師達は、教師の自己洞察のために必要な過程として形成的評価を受け入れる一方で、よりオープンで改善を目的とした、連携的で具体的な評価を望んでいる。さらに、SETsの結果に教師がもっと関わり、教師間で意見交換することで授業改善を進めることを希望している。

Peter Burden
Okayama Shoka University

The introduction of student evaluation of teaching

The Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan has made the implementation of self-evaluation in tertiary education compulsory since 1999 (MEXT, 2004). Reflecting the popularization of higher education, end-of-semester student evaluation of teaching surveys (SETs) have been encouraged in the belief that popular teachers and courses offer student satisfaction, will attract potential students and, for private institutions dependent on fees for income, will make them more able to retain students once they have entered.

In this study, twelve ELT university teachers reflected, through using metaphors, in interviews about the use of SETs in their respective universities. The paper will first outline how SETs are administered in tertiary education and then briefly focuses on contentious areas that have led the author to question the use of SETs from an ELT perspective. After outlining the research methodology in which details are given about the interview style and the participants, the importance of metaphorical expressions which teachers employ when talking about their professional beliefs about evaluation is discussed. Teachers' spontaneous use of metaphors during interviews revealed participants' perceptions of their roles in tertiary education, and the following discussion offers some implications for improving the use of evaluation. These include a greater need for clarity of the evaluation purpose, more 'horizontal' dialogue between

the parties involved in evaluation, and the use of multiple data sources so that evaluation becomes more personally meaningful for teachers.

The administration of SET surveys

SETs in tertiary education in Japan usually utilize paper and pencil questionnaires containing Likert-type 1-5 scales anchored from "Very poor (1)" to "Very good (5)." These questions are coupled usually, but not always, with a final global characteristic of 'overall satisfaction' of the course and 'effectiveness' of the teacher. Many schools require the students to anonymously fill in closed-item questions which are subsequently used for data analysis by the school administration and are the basis for summative scores. Many writers, for example Feldman (1988, p.291), note that if faculty and students do not agree as to what constitutes effective teaching, then faculty members may well be "leery" of students' overall ratings of them. Often, there is not any explicit statement of purpose delivered either to schools or to teachers, or any indication of a remedial path for teachers who receive poor evaluations. While many may see the introduction of SETs ultimately as a benign attempt to encourage teachers to somehow improve or innovate their teaching, for many teachers the lack of any remedial path, the delay in feedback, and the actual timing of the administration suggest a summative decision-making perspective.

Rationale for the study

Gorsuch (2000) argues that knowledge in Japan is traditionally seen in terms of immutable truths so, there is a danger of dissonance through oversimplifying the conditions required for language learning to a set of discrete points instead of recognizing that the "whole is more than the sum of the parts" (Crabbe, 2003, p.27). While Darling-Hammond and Snyder (2000, p.523) suggest that "teaching is becoming more complex in response to increasingly challenging curriculum expectations and growing diversity among students," the emphasis seems to be one of controlling behavior and learning in such a way that they will conform to pre-determined ends or an "identical path to understanding" (Darling-Hammond & Snyder, 2000, p.523).

Recognizing that effective teaching is contextual, if definitions of the constituents of effectiveness are not in place, teachers and administrators may have conflicting expectations (Stronge & Tucker, 1999). This researcher started to hear concerns among English language teaching colleagues when SET surveys began to be administered at the end of a single semester of English education. Is it possible for 'communicative' language teachers who encourage functional language proficiency involving the expression, interpretation, and negotiation of meaning to be evaluated after just a single fifteen-week semester by first-year undergraduates who may not previously have experienced such a teaching approach during six years of junior high and high school English education? Teaching is too important an activity to be conducted without critical inquiry and as there have been insufficient explorations of teachers' perceptions into the introduction of SETs, research focusing on faculty perceptions and how evaluation affects teaching is clearly warranted. To understand teachers' personal understandings of the introduction of teaching evaluation, and whether the use of SETs matches their conceptions of teaching, data from teachers' spontaneous use of metaphors during interviews were collected.

If evaluation through one tool, SETs, is to encourage improvement, the key element of receptivity to this form of evaluation from teachers cannot be ignored, as feeding back useful, diagnostic information creates energy, which can then be directed through reflection into an action plan which leads to development.

SETs and the use of metaphor

Reform in Japanese education has been described as top-down (Gorsuch, 2000), but made opaque through the "extraordinary reluctance to clarify, define, and articulate policy" by MEXT (Miyoshi, 2000, p.681). While evaluation should be seen as "an agent of supportive program enlightenment and change" (Norris, 2006, p.578), it can be argued that if evaluation is left to the end of a course, it loses any opportunity to inform and influence teaching. The longevity of SETs use in America may suggest presumptive 'evidence' for the benefits, but studies considering the institutional effects on teachers are "scarce or non-existent." (Kulik, 2001, p.15).

It is fruitful to consider what sorts of metaphors teachers use to refer to evaluation, how the metaphors are used, and to discuss what implications can be drawn from teachers' metaphor. They serve as "pattern making devices" (Miles & Huberman, 1994, p.225), placing the metaphors into the larger context of evaluation and the teachers' position within the current evaluation method. Metaphor also "captures the thinking of teachers in their own language, rather than in the language of the researcher" (Munby, 1986, p.198), while De Guerrero and Villamil (2000) suggest that teachers employ metaphorical expressions when talking about their professional beliefs, which reflect how teachers understand their world.

As metaphors reveal "tensions, surprises, confusion, challenges and dilemmas" (Louie, Drevdahl, Purdy, & Stackman, 2003, p.143), an examination of metaphor use can encourage reflection on the relationships teachers have with other stakeholders—students, colleagues, parents, and administrators.

Method

Twelve tertiary English language teaching (ELT) faculty were asked to outline their perceptions of the introduction of SETs in their tertiary institution through interviews. The interview questions were flexible and encouraged teachers to reflect on their first-hand experience of how they were affected in their daily teaching by the introduction of SETs. A range of perspectives from both male and female ELT teachers was sought to enhance credibility (Rubin & Rubin, 2005).

Seven male and five female teachers from five different universities—one national and four private universities—in one city in Western Japan participated. Their ages ranged from early 30s to late 50s, while their teaching experience in the tertiary sector ranged from one year to close to thirty years. The two Japanese teachers of English in this study were full-time tenured faculty. Seven of the ten expatriates had lower-status, limited term contracts and the remaining three were tenured. As evaluation is inherently political, anonymity and confidentiality procedures were outlined, and participants understood that the tape-recorded interviews would be transcribed verbatim.

The interviews took place approximately two months after teachers had administered evaluation during the final weeks of the second semester ending in early February. It was assumed that the university administration had had sufficient time to analyze and send the data back to teachers in anticipation of the new school year starting in mid-April. However, none of the teachers had received feedback despite the two-month gap.

Findings

Findings suggest that teachers feel threatened by the introduction of SETs and are concerned about the purpose and consequences of this form of evaluation. Participants' metaphors reveal their lack of involvement, voice, and feelings of distance from power holders, which often encourages an absence of trust in accepting organizational change.

Metaphors to describe those who devised SETs items

The participants feel threatened by the opaque evaluation purpose and use uninformed speculations while disparaging others they have not met. Participants have little confidence in the ability of administrators who wrote the questions. Administrators are seen as "powers that be," "big cheeses," or "old farts" and "groups of little men" who form "nameless committees" and "get together" in "darkened rooms" and whose views are not consonant with teachers' educational goals and conceptions of teaching. Questions are seen as "outdated" and "irrelevant," being written "about a million years ago" by some "Japanese statistician type" or by "someone in the hard sciences a long time ago." One teacher compared the questionnaires to dictionaries which build on the original corpus and only slowly change over passing years.

Participants' feelings of unease about the role of the administration reflect findings in Ryan, Anderson, and Birchler (1980), which suggested that SETs usage increased the distance between faculty and administration. In the current administrative climate, participants fear they are evaluated unfairly because consequences of SETs are often unknown as stakeholders hold different purposes for evaluation, and so considerations

of what or who the evaluation serves is far from clear. While the developmental formative nature of evaluation is often recognized in English language teaching literature (see Hedge, 2000), without clear description, teacher understanding is incomplete and so teachers do not understand which behaviors to improve, which to retain, and what the likely consequences of this form of evaluation are.

Metaphor to describe evaluation as a form of consumer satisfaction

For participants, another focus of evaluation is to directly improve the quality of student satisfaction so "the goal is getting more students and keeping them in business" so they become "cash cows" and should not be "let go for four years." Therefore, participants suggest that evaluation has become a "popularity contest" and, while those teachers whose "little numbers and charts" look "good" are safe, universities can say to "poor" teachers in the face of declining admissions: "You've had consistently low evaluations and we don't need your services any more." Evaluation is seen as a "marketing tool" to "sell" the school and if teachers are "not jumping up and down in class" the students may not perceive it as enthusiasm and so give a poor overall global evaluation. One participant suggests that "popular teachers" can get a "good reputation" and can "make the school money" in "fun" classes. Schools' survival is addressed through evaluation - "because the kids basically walked in doesn't mean that they're going to stay" as students may drop out due to a lack of immediate "satisfaction." This caused one teacher to ponder:

> I know I shouldn't feel scared or uncomfortable by doing this because teachers should be evaluated. I think students have to be satisfied but at the same time they don't know how to study, they don't know what the good education is so we have to make them do things they don't want to do. Even though they hate it, it doesn't mean that the teacher is a bad teacher. This is the difficulty. One teacher said, "Of course I get the bad scores because they don't want to study." So he knows that he isn't popular.

The issue of "popularity" is a fundamental issue for another participant who says the degree of preparation, or "hidden labor" is not addressed through evaluation while he hears students complaining of workloads. He says:

> You could be a real, quote, "strict" teacher. I tend to give a lot of homework and the comments are, "You make us work too hard." But I don't think that it's too hard. It depends on your interpretation. I think they can handle it. I think the work they do outside the classroom is just as important as in it. They've got to bring English into their daily lives so I have them doing things outside and then I get complaints.

However, he worries that "if student complaints are reflected on here [evaluation forms] then I'm a bad teacher." Participants suggested that classes where content is not emphasized will lead to "dumbing down" because teachers will need students to have "a good time" so that "appropriate" education becomes secondary to an education the student "wants," which is problematic when students enter school with little initial academic interest. Participants suggest the competition for students means that teachers need to be a "draw" to attract students through word of mouth, which may promote speculation and tension among contracted teachers with regards to their future employment.

Metaphors to describe fracturing relationships

Similarly, participants are "wary" of ranking teachers in "league tables" which emphasize "winning and losing" as they can can lead to "a competitive win-lose situation" (Braskamp & Ory, 1994, p.7) where faculty learn little about "*how* to improve, only that they *should*" (p.6). This decline in collaboration and dialogue has led to harboring bitter feelings expressed through metaphor towards colleagues, especially teachers of "conversation." These classes are seen as "fun," "non-challenging classes" with colleagues who "play games," "jump around" and "act like a jack-in-the-box." This resentment may well stem from a belief that the evaluation "playing field" is not even, with evaluation being unfair as it is only used to judge part-timers. One

participant has heard of tenured faculty with "poor" evaluations being retained at the expense of part-timers with better scores. Participants are suspicious of others' teaching methods, the ability of students to appreciate and evaluate "academic" classes, and whether teachers manipulate evaluation data to inflate their scores. As the parameters are unclear, teachers question, "When is 'good' good enough?"

Perhaps paradoxically, while many teachers seem to oppose the use of SETs for summative purposes, they lament the teaching performance of those around them. Most participants implement their own evaluation to aid reflection on their own practice, but point to a lack of professionalism of those around them. Participants talk of "dead wood," suggest that tenured, full-time university teachers "go through the motions" or "fall into ruts" or "comfortable routines," and "devalue teaching because it gets repetitive." Comments above may reflect different levels of evaluative scrutiny for tenured or non-tenured faculty, similar to Nasser and Fresko's (2002) findings where few tenured faculty reported changing their teaching as a result of course evaluations.

Metaphor as an expression of conceptions of teaching

Participants saw their teaching through metaphors of "art," which suggests "a unique set of personal skills" (Freeman & Richards, 1993, p.206). As one participant says:

> I can feel when the kids are tired or preoccupied. But I'm sure there are teachers who wouldn't feel anything. Teaching is not a craft or a skill you can learn, or a set of techniques. Art is something that is inside that I can develop. Other teachers are more mechanical; it's more like they've studied techniques and things. I feel I pick it up as I go; I develop it and can see it working and feel when something worked or didn't work.

He feels evaluation reinforces a view of teaching as a set of techniques which can be learned but which do not form a "complete teacher." He suggests teachers need to have the "space" to "develop" ideas and to experiment even at the risk of failure. However, SETs surveys reinforce specific faculty teaching behaviors, and "may constrict teaching styles rather than encouraging a diversity of classroom strategies" (Braskamp & Ory, 1994, p.182). Another participant commented:

> The questions are predetermined by administrators who know little about teaching, and who actually determine what techniques should be used. In the same sense that a textbook assumes a certain method or approach, evaluations show techniques a teacher is required to use. Evaluation is not responding to the humanity of the teachers or students. Knowledge for me is something that they can discover for themselves, but as it is a foreign language it's not something inside them; to discover from examples by themselves is a good way but just to sit and tell them this is what we do here- I don't think that's an effective way - getting them to reach answers for themselves is the best way.

For a third participant, rather than behaviors or "techniques," teaching is a "creative process" which requires constant reflection leading to "refinement" and "development." While teaching can be "learned" like mathematics so that "there are practices you can follow so that anyone can carry out a teaching job," unreflective teachers are "unempathetic," while "good teachers" can "know when [they've] caught the audience and can lead them to tears or laughter." Other participants suggest similar metaphors, seeing their roles as a "magician" or "a creator" who "creates the sequence or order to best fit the students in different classes," or, again, as an artist being creative in order to hold onto, and encourage, interpersonal relations and positive attitudes.

Another teacher illustrates the irrelevance of the evaluation drawing a distinction between teachers' concerns with the day-to-day running of classes—"the small details and things like atmosphere"—and the university interest in the "framework" or the "published, visible side" of what teachers do inside the classroom. Therefore, participants have little confidence in the ability of power holders whose views are not consonant with teachers' educational goals and conceptions of teaching.

Metaphor as an expression of teacher programming

Teachers see a "robotic" or "cloning" metaphor implicit in SETs and its representation of teaching as "teacher programming." One teacher observed that:

> It makes clones out of everybody; do this and this and this and you'll be an acceptable teacher. Yet every teacher has a different personality…you have to watch what other teachers do and listen to the students and if you want to know whether a teacher is effective or not you need to know a lot more than the answers to a few questions.

"A robot could do that" [the teaching implied by the evaluation], while "it could be programmed," with the questions seen as "limiting" because they emphasize the "little aspects of teaching" and so "diminish the trust of teachers." It is suggested that the "Ministry" is trying to project an image of a "correct institution" which "squashes the teaching style." There is a lack of a shared sense that SETs reflect important aspects of teaching, and the use is not consonant with teachers' educational goals and conceptions of teaching. An extended quote from one of the participants serves as a useful summary:

> SETs evaluation is based on the concept of the class as a lecture and somewhere in here maybe the bureaucratic control the belief is that there is a good way to teach…these questions are a good way to teach. If you can do XYZ then you're a good teacher and breaking down teaching into these nice little categories that are numerically controlled.

Another laments: "I would like to say my job is a profession but it's just a job." Giroux's (1988) school-as-factory metaphor comes to mind as SETs reduce teaching to basic, predetermined skills to quantify and make tangible figures out of teaching. Teachers learn to understand and change their work behavior by continually examining, analyzing, hypothesizing, theorizing and reflecting as they work (Schön, 1983). Teachers' valuing evaluation and using feedback depends on how the teaching act is construed, and there is little in evaluation which considers the 'thought' behind teaching.

Discussion

Increasingly, the introduction of student evaluation of teaching is seen to "focus on the abilities of teachers" (MEXT, 2001), but the underlying conception of what *good* teaching entails and how it can be encouraged has not been made clear.

All of the participants accept that *formative* evaluation is necessary as a process to give insights to teachers. The participants suggested they often administer self-generated student evaluations which offer students opportunities to provide additional, qualitative comments about the course, the teaching and the teacher, as well as to evaluate their own course performance. However, they all wished for a more open, improvement-focused, cooperative—but specific—institutional evaluation. They want more teacher involvement, more dialogue between teachers to discuss the results to aid the reflective process for change, and the removal of the pervasive atmosphere of secrecy that surrounds data results.

Openness about the process encourages knowledge of both the purpose and what happens to the surveys after they leave the classroom. It should also be made clear how important each student's opinion is, how the opinions impact on non-tenured teachers and on elective classes. If the university evaluating body has criteria by which the evaluations are reviewed these should be made known; if there is an overall objective to which teachers are supposed to be working it would be useful to know what that is so that classes might be adjusted. While teachers do not wish to take a lot of student time, more specific questions would push students to think more about answers. Also underpinning SETs are judgments from an accountability perspective whereby there is an assumption that all students pursue an identical path to understanding. This view erodes individual teacher's artistic and intuitive knowledge. There is a loss of a "sense of involvement of teachers" (Prabhu, 1990, p.172) as the participants distanced themselves from mechanical SETs. One participant sees evaluation as personally irrelevant to his notions of improvement as he sees teaching as a personal, sharing act, from which knowledge grows. He does not see education in terms of "concrete" improvement.

Teachers also suggest that using other evaluation methods would create more of a balance and useful feedback. One participant says:

> Well, I would want that decision not to be based solely on one thing. Any kind of assessment needs to have multiple sources. However these universities are understaffed and overworked and they don't have any money and they don't care.

All of the teachers suggest that using SETs as the sole criterion for evaluating teachers is flawed. As another participant says:

> Students should be given every opportunity to give feedback to teachers about their teaching. If they cannot, then the teacher is missing a vital perspective on the effectiveness of lessons taught. However, this should be balanced with the views of one's colleagues. I feel that if the survey were balanced with some form of peer review, such as classroom observation and feedback, then it would be a more valuable exercise. Evaluation only by one's students seems a dangerous path for education and educators and worrying for the future development of Japanese education.

Using other evaluation methods would create more balanced, useful feedback. Instead of easy to administer SETs, peer review would enable teachers to learn from each other, while self-evaluation would encourage deeper reflection, without "condemning" teachers. Rating teachers on low-inference, observable behavior as the sole basis for judgments is still widespread, contradicting the recommended use of multiple sources (Seldin, 1993). Even *if* SETs are intended for formative development, many teachers do not gain any new knowledge as they question the value of the source of information. Utilizing focus groups may be one way forward for authentic teaching improvement through representatives of students, teachers, parents, and administrators discussing evaluation in a peer-group context.

Conclusion

The relation between MEXT who impose evaluation, school administrators who introduce individual school evaluation mechanisms, and the teachers who carry out evaluation, is problematic. Teachers have legitimate concerns over the use of data and everyone involved—faculty, administrators, and students—need to discuss how the data should be collected, who should receive the data before any SETs are collected, and how those results are used. Also, participants' metaphors suggest the need for more teacher involvement and ownership and more dialogue between teachers to discuss the results. This would aid the reflective process for change and remove both competitive feelings and the pervasive atmosphere of secrecy that surrounds data results.

References

Braskamp, L., & Ory, J. (1994). *Assessing faculty effectiveness.* San Francisco: Jossey Bass.

Crabbe, D. (2003). The quality of language learning opportunities. *TESOL Quarterly*, *37*(1), 9-34.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, *16*, 523-545.

De Guerrero, M., & Villamil, O. (2000). Exploring teachers' roles through metaphor analysis. *TESOL Quarterly, 34*(2), 341-351.

Feldman, K. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, *28*(4), 291-344.

Freeman, D., & Richards, J. (1993). Conceptions of teaching and the education of second language teachers. *TESOL Quarterly, 27*(2), 193-216.

Giroux, H. (1988). *Teachers as intellectuals: Towards a critical pedagogy of learning.* Westport: Bergin and Garvey.

Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, *34*(4), 675-710.

Hedge, T. (2000). *Teaching and learning in the language classroom.* Oxford: Oxford University Press.

Kulik, J. (2001). Student ratings: Validity, utility, and controversy. In M. Theall, P. Abrami, & L. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them?* (pp. 9-25) San Francisco: Jossey Bass.

Louie, B., Drevdahl, D., Purdy, J., & Stackman, R. (2003). Advancing the scholarship of teaching through collaborative self study. *The Journal of Higher Education*, *74*(2), 150-172.

MEXT. (2001). *Educational reform plan for the 21st Century. The Rainbow Plan*. Retrieved from <www.mext.go.jp/english/topics/21plan/010301.htm>

MEXT. (2004). *FY2003 white paper on Education, Culture, sports, science and technology*. Retrieved from <www.mext.go.jp/english/news/2004/05/04052401. htm>

Miles, M., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.

Miyoshi, M. (2000). The university and the "global" economy: The cases of the United States and Japan. *The South Atlantic Quarterly*, *99*(4), 668-697.

Munby, H. (1986). Metaphor in the thinking of teachers: An exploratory study. *Journal of Curriculum Studies*, *18*(2), 197-209.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment and Evaluation in Higher Education*, *27*(2), 187-196.

Norris, J. (2006). The why (and how) of assessing student learning outcomes in college foreign language programs. *The Modern Language Journal*, *90*(4), 576-583.

Prabhu, N. (1990). There is no best method—Why? *TESOL Quarterly*, *24*(2), 161-176.

Rubin, H., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing*. Thousand Oaks: Sage.

Ryan, J., Anderson, J., & Birchler, A. (1980). Student evaluation: The faculty responds. *Research in Higher Education*, *12*(4), 317-333.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education*, *39*, 40-43.

Stronge, J., & Tucker, P. (1999). The politics of teacher evaluation: A case study of new system design and implementation. *Journal of Personnel Evaluation in Education*, *13*(4), 339-359.

Peter Burden is a Professor at Okayama Shoka University, where he has taught for nearly twenty years. His recent research area has been the introduction of student evaluation of teaching and he has published and presented widely in this area, at home in Japan and also internationally.