Mathematics Teacher Education and Development 1999, Vol. 1, 50-63

Thinking about Teacher Effectiveness: How do Teacher Appraisers make 'Gut Feeling' Rankings?[1]

Julia Horring, Judy Paterson and Bill Barton

The University of Auckland, New Zealand.

As part of a larger research project, seven trainee mathematics teachers completing a one year post-graduate secondary teaching diploma were observed teaching two lessons each. The researchers, who were experienced mathematics teachers, ranked the pre-service teachers in terms of their future effectiveness using two methods of ranking. One involved an observation schedule derived from behaviours associated with effective teacher research. The other required the two researchers to make a 'gut feeling' ranking of the pre-service teachers. An investigation of the 'gut feeling' ranking showed that it could be explained using standard teacher effectiveness models. The correlations between the two rankings were high, raising some interesting questions to be investigated using larger samples.

Introduction

Teacher expertise is a field of study that has generated much interest in educational research. The reality of school life, and the resource restrictions on pre-service teacher education programmes, mean that it is usually not possible to observe an extended series of lessons with each beginning teacher. So, is it possible for teacher appraisers to make assessments of teacher effectiveness using a few classroom observations?

A wider study (Paterson & Barton, 1998), which investigated the contributions of content knowledge and pedagogical knowledge to pre-service teacher education, also provided an opportunity to compare two methods of assessing effective teaching from limited observations. This paper reports on just that aspect of the study. The two methods of ranking the teachers for their potential effectiveness were: a classroom observation schedule based on behaviours predicted by theory; and a holistic impression or 'gut feeling' made by two observers who were experienced mathematics teachers.

A recent trend in educational circles has been to base assessments on 'outcomes', usually behavioural ones, or those that can be easily observed. The observational schedule was developed from behaviours that might be expected given theoretical models of effective teaching, although practical limitations affected the final schedules used.

Professional judgment is often used in schools by teachers and teacher educators when making an assessment of pre-service teachers. Their judgment is not usually guided by research findings but is based on a 'gut feeling' or instinct. The thinking process that the observers used while making their ranking provides an account of the sorts of things that experienced teachers

[1]The research reported in this paper was part of a project funded by The New Zealand Ministry of Education.

take into account when evaluating whether a teacher will become an effective mathematics teacher. This process was analysed and compared with a model of teacher effectiveness based on models created by Jaworski (1992) and Sullivan and Mousley (1994).

This research is an initial attempt to compare quantitative and qualitative assessments. It could be seen as questioning the empirical paradigm. The conclusions drawn are necessarily tentative, both because the sample was small, but also because the study was done within the context of a wider study, and therefore included some less than perfect design features. We believe, however, that it is important indicative work.

The paper first reviews aspects of the literature which contribute to this study. Next, details of the study are given, including the way in which the observation schedule was developed. The result of comparing the methods of ranking are examined and implications drawn, followed by a discussion of the validity of the comparison. Finally, the process of 'gut feeling' ranking is discussed and compared with the theoretical model.

Effective Teaching

The study approached the question of what constitutes effective mathematics teaching from two perspectives: a theoretical one and a research-based one. Using the theoretical literature, three models which viewed effective teaching as promoting good mathematics learning were drawn together into a 'teaching triad' which helped to describe quality teaching in the mathematics classroom. This model was used to explain how experienced mathematics teachers made professional judgments on the future effectiveness of teachers. The research-based literature was used to identify the characteristics of effective teaching. These different behaviours were compiled to help construct the observation schedule for evaluating the future effectiveness of pre-service teachers.

Effective Mathematics Teaching Promotes Learning

Jaworski (1992) investigated qualities of mathematics pedagogy based on the constructivist view of knowledge and learning, in which the teacher's role is to gain access to a student's thinking and to influence his/her construction of knowledge. From observations of the practice of teaching she has identified aspects of practice which seem significant in terms of this theoretical standpoint. She noted that effective mathematical classrooms are those which (a) provide a supportive learning environment, (b) offer appropriate mathematical challenge, and (c) nurture processes and strategies which foster learning.

Jaworski incorporates the above three elements into *the teaching triad* to clarify the teaching process in mathematics. She describes the essence of mathematics teaching as lying in three domains: *management of learning, sensitivity to students,* and *mathematical challenge*. Any teaching situation is likely to have elements of each of these three domains.

The *management of learning* vertex of the triad includes classroom organisation and curriculum decisions that involve establishing ways of working, and creating classroom values and expectations. *Sensitivity to students* involves developing an approach for working with students which is consistent with the individual characteristics and needs of the students. The third vertex of the triad, *mathematical challenge,* involves stimulating mathematical thought and inquiry and motivating students to become engaged in mathematical thinking.

Sullivan and Mousley (1994) used another approach to establish the features of quality mathematics lessons and reached similar conclusions to the Jaworski (1992) study. They sought the views of experienced mathematics educators in an open survey and attempted to find some commonality among the diverse responses. Six key components were identified. The *building understanding* component stands out as being an overarching feature. *Organising for learning, nurturing, engaging, communicating* and *problem solving* are seen as being part of a quality lesson if they contribute to building the students' understanding.

The statements that experienced mathematics educators use to describe quality lessons are closely linked to the *teaching triad* that Jaworski (1992) developed. Horring (1998) has suggested the following synthesis of Jaworski's triad with Sullivan and Mousley's six key components (see Figure 1).

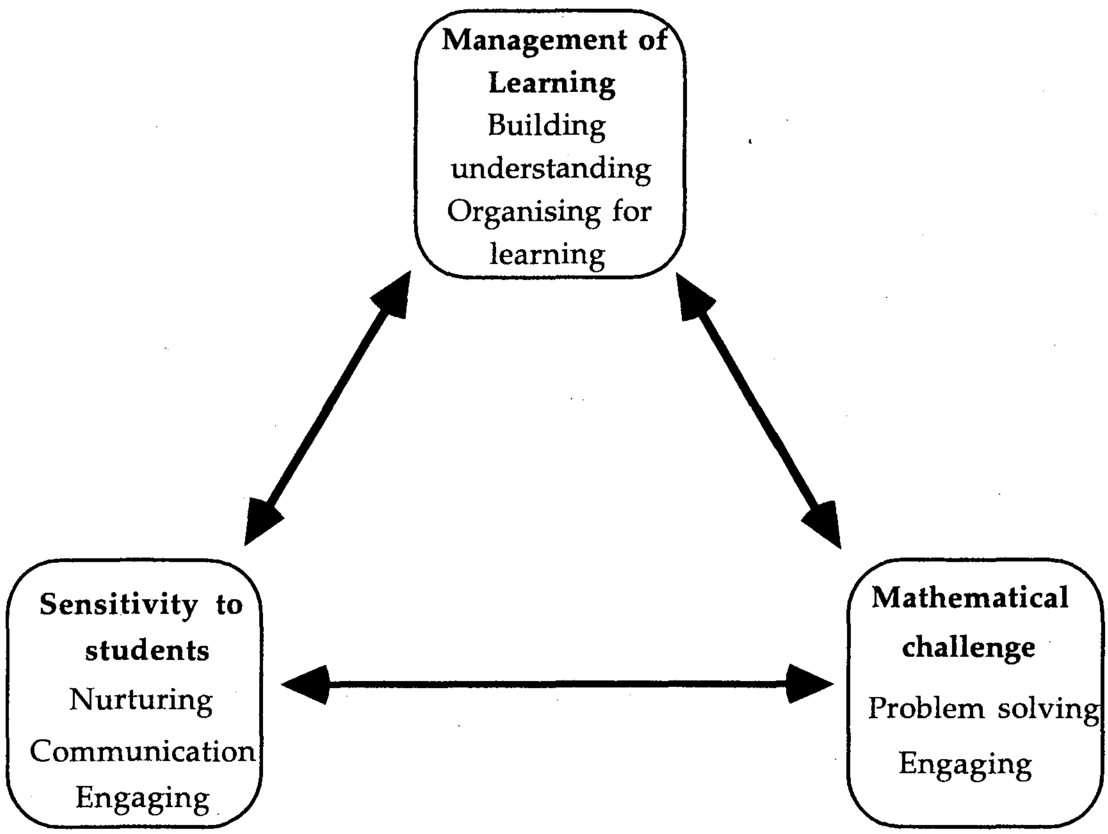

Figure 1. Horring's synthesis of the teaching triad.

The *organising for learning* and *building understanding* components of the Sullivan and Mousley model closely match the *management of learning* vertex of the Jaworski teaching triad. All three components refer to the teacher setting out an organised and structured lesson to provide a learning situation for the students.

The *problem solving* and *engaging* components can be related to the *mathematical challenge* vertex of the teaching triad. Mathematical challenge can be created through students working out problems for themselves. The activity needs to be appropriate to the students' interests and ability level for it to be engaging.

The final part of the triad, *sensitivity to students*, is closely related to the *nurturing* aspect of the Sullivan and Mousley (1994) diagram. Parts of the *engaging* component also fit in with this vertex since the work can only be engaging if the teacher is sensitive to the needs of the students.

The communication aspect was the most difficult to place in the teaching triad. Sullivan and Mousley's (1994) *communication* aspect included the idea that a quality lesson has opportunities for students to discuss, write, share, and cooperate. This can be thought of as a teaching strategy that a teacher uses to manage the learning. However, providing students with the opportunity to communicate can also be considered as the teacher being sensitive to the needs of the students. Since Sullivan and Mousley (1994) have already established that the *building understanding* component is the overarching feature to which the other components contribute, it was decided to place *communication* in the *sensitivity to students* vertex of the triad.

This extended teaching triad provides a model that helps to define key aspects of effective mathematics lessons. In the analysis of the Sullivan and Mousley (1994) model the *building understanding* component needed to be the key factor in the organisation of the other components. Although the three aspects are interconnected and depend on each other, the top vertex appears pivotal in ensuring that mathematics learning is the focus of the lesson.

A third approach to thinking about effective teaching in relation to learning is offered by Berliner (1987) who proposes a simple definition of effective teaching based on the idea of maximising *academic learning time*. This approach, Berliner claims, helps to explain effective teaching practices. Teaching behaviour and classroom processes that affect achievement probably affect *academic learning time* and therefore influence the students' opportunity to learn. It is suggested that Berliner's conception of effective teaching involving the opportunity to learn and academic learning time helps to reinforce the idea that the top vertex of the triad, *management of learning*, has the most significant role in understanding effective teaching.

Research into the Characteristics of Effective Mathematics Teaching

A second perspective on the nature of effective mathematics teaching focuses on teacher behaviours that are identified from relevant research. This provides the basis of the observation schedule that was developed for this study.

Gagne, Yekovich and Yekovich (1993) focus on teachers' cognitive processes and teacher knowledge in order to understand more fully how teachers organise the learning environment to provide meaningful and effective learning. Three areas of skill and knowledge make up expertise in teaching in their account. First, an expert teacher has a highly organised and elaborated declarative knowledge base. Second, the expert uses automated sets of basic skills. Third, the expert teacher has well developed but adaptable sets of teaching strategies. These are discussed in more detail below.

Highly organised and elaborated declarative knowledge base. Shulman (1986) proposed a theoretical framework for analysing teacher knowledge that helps to understand more fully the complexities of teacher understanding and teaching. He distinguished three categories of content knowledge: (a) subject matter knowledge (including facts, concepts and structures); (b) pedagogical content knowledge (including effective representations of the subject); and (c) curriculum knowledge.

Studies by Leinhardt and Smith (1985), McDiarmid and Wilson (1991), and Stein, Baxter and Leinhardt (1990) reinforce Shulman's idea that a teacher's subject matter knowledge must include more than just the basic facts and concepts in the subject domain. Also necessary is a rich interconnection of facts and processes within the subject domain. Effective and expert teachers have a much fuller understanding of the semantic and syntactic structure of the mathematics domain.

Pedagogical content knowledge is the way in which the teacher represents and formulates the subject to make it comprehensible to others, including the most useful representations of the key ideas, powerful analogies, illustrations, examples, explanations, and demonstrations (Wilson, Shulman & Richert, 1987). A teacher needs to be able to account for differences in student abilities, prior knowledge and learning styles. Even and Tirosh (1995) add to this an understanding of what makes learning specific topics easy or hard. The teacher needs to know what preconceptions and misconceptions students are likely to bring to the lesson, and to have strategies to help sort out the misconceptions. Such knowledge also helps to clarify and connect the new topic with previously learnt concepts. Effective teachers would display this knowledge during their explanations of mathematical concepts.

Automated sets of basic skills. Automated basic skills enable the expert teacher to execute basic routines in the classroom smoothly and efficiently without apparent effort. Expert teachers have routines for such things as setting student norms for their classrooms, checking homework, summarising class performance, marking the roll, and so on (Gagne et al., 1993). These routines help to keep the students engaged in the instructional process since the students do not have to keep track of a changing environment. When classroom management routines have been automated the teacher is better able to attend to the teaching and learning of the students (Leinhardt, 1989; Borko, 1989).

Flexible and adaptable sets of teaching strategies. Through the use of well-developed but flexible and adaptable sets of teaching strategies the expert teacher is able to plan, teach and evaluate lessons in successful and imaginative ways (Shulman, 1986). Expert teachers are able to call on a variety of instructional strategies as appropriate. In particular, they adapt instruction to the students' interests, ability levels, and attention spans.

Tomic (1989) and Leinhardt (1989) outline teaching strategies that effective teachers use in the classroom. Questioning and explaining are two areas that they highlight. Borko and Livingston (1989) emphasise the importance of teaching strategies being flexible and responsive to the needs of the students. This flexibility seems to be related to the expert teachers' connected and easily accessible knowledge structures.

Details of this Study

During 1998 a class of pre-service post-graduate secondary mathematics teachers included some who had just graduated in mathematics but who had no teaching experience, and some who were experienced teachers but were still learning mathematics. The New Zealand Ministry of Education agreed to fund a research study to compare and contrast these preparations for mathematics teaching (Paterson & Barton, 1998). Part of that study involved assessing teacher effectiveness within the context of the limited observations possible in a pre-service programme. A researcher employed on the study used this aspect as the basis of a dissertation, comparing the different modes of assessment (Horring, 1998).

The study involved seven pre-service mathematics teachers, four from a mathematical background and three from a teaching background; three women and four men. They varied widely in age (early twenties to mid-forties) and classroom experience (none to several years in various countries).

Data Collection

The data collection involved observing the seven pre-service teachers for two lessons each. Each pre-service teacher in this study taught a mathematics lesson to a year eleven class and, later in the year, to a year twelve or year thirteen class. They had not previously taught those classes. Attempts were made to make these encounters as similar as possible between the teachers, for example, through content, familiarity with the school, and timing. Both researchers watched every lesson. During each lesson they recorded items on an observation schedule of half the observation items (see below) and made notes to aid their ranking process.

After each lesson the researchers independently ranked the pre-service teachers that they had seen so far. This ranking was based on their holistic impression, free from predetermined criteria and from the scores on the observation schedule ('gut feeling'). (The extent to which this objectivity was achieved is discussed below). The observations were not discussed by the researchers until the ranking of all the teachers was complete for each cycle. The researchers each wrote up an account of what they considered as they made their rankings. This writing consisted of a list of positive and negative factors in the lessons in relation both to the general 'flavour' of the lesson (e.g., that the teacher related in a warm manner to students), and to key events (e.g., when a teacher made a mathematical error but corrected it with the aid of the students).

The data collected from the observation of the lessons therefore included: Researcher 1's 'gut feeling' ranking and notes on the process of ranking, Researcher 2's 'gut feeling' ranking and notes on the process of ranking, Researcher 1's observation schedule, and Researcher 2's observation schedule. The two observation schedules were combined to give a third ranking. Note that none of these rankings involved a measurement, they were simply an ordering of the pre-service teachers in order of effectiveness. Rankings were done after the observations in June, and again after observations in September, thus there were six rankings altogether.

The Development of the Observation Schedule

The study assumed that if a novice teacher is going to be an effective mathematics teacher she/he will exhibit some behaviours that characterise effective or expert teaching. Hence, the observation schedule was developed using the three areas identified by Gagne et al. (1993) above. Lists of teaching behaviours based on these areas were developed, and then restricted to those items which could be observed in a single lesson. The following additional criteria were used. The teaching behaviour should: (a) occur with sufficient frequency in over fifty per cent of the lessons observed; (b) be well defined so that the observer can easily distinguish an example of the behaviour; (c) be recordable and countable; (d) relate significantly to the theoretical models; (e) be reliable, that is, there is to be high agreement between two observers. The schedule also needed to include items that would be demonstrated by pre-service teachers with either mathematical or teaching backgrounds. The total schedule had to be of reasonable size. Items initially chosen were trialed by experienced teachers in real classroom situations, and the list modified. The final list of items is below.

1. Any link between mathematical ideas outside today's agenda.
2. Adapting to student input, for example, changing tack in response to student.
3. Spontaneous generation of examples /metaphors /counter-examples.
4. Connections with other domains, for example, history /real life /other subjects.
5. High level questions: Explain, prove, ...
6. Long request or question time lag.
7. Prepared examples /metaphors /counter-examples.
8. Number of verbal personalised connections.
9. Mathematical objective punctuation.
10. Management punctuation.
11. Any positive acknowledgment.
12. Any negative acknowledgment.

The items were mixed and distributed across two observation schedules, one for each researcher. Each researcher also identified one student from the front centre of the class and kept a record of whether the student was engaged in the lesson or not.

Results and Discussion of Ranking Comparisons

Table 1 shows the correlation coefficients between the three rankings (each of the two observers 'gut feelings' and that generated from the scores on the observation schedules).

Table 1
Spearman Rank-Order Correlation Coefficients

Comparison of Rankings	June	Sept
Two observers' 'gut feelings'	0.75	0.94
Schedule and Observer 1 'gut feeling'	0.61	0.71
Schedule and Observer 2 'gut feeling'	0.61	0.37

Assuming these results are properly validated and generalised, do they indicate that we can throw away observation schedules and rely on subjective assessments? Or do they give us confidence that 'objective' (and possibly inexperienced) observers can be used to judge good teaching (in preference to experienced mathematics teachers)? Or do the results simply validate the theoretical constructs from which the observations were drawn?

We regard the main implication to be continued confidence in subjective judgements of experienced observers over mechanical means of assessment. The reasons for this are as follows.

1. Not only did the whole process of devising discrete 'observables' and then counting them feel like a farce because many important factors had to be discarded as impractical or unobservable in one lesson, but also it failed to take into account differences in the situation: group work, class attendance and behaviour, time of day, etcetera. We would argue that these are relevant. Furthermore, the cut-down schedule still needed two observers to complete. The reality is that pre-service observations are done by one lecturer who must also attend to other matters such as content, behaviour management, and class context.
2. Devising and recording the observation schedule required experienced judgment. Despite trialing, judging whether, for example, a teacher was spontaneously generating examples (item 3), required experience and was not always certain.
3. The post-event analysis of the 'gut feeling' process revealed close links with the theoretical model (see below), implying a rational basis for this subjective assessment.

The research does go some way to validating the theory as a useful model for thinking about effective mathematics teaching. An implication is that some experience with such models is likely to lead to better judgments, and could usefully be part of teacher educator development.

Validity of the Comparison

A serious design drawback of this study (necessitated by the wider study from which it was taken) is the validity of a comparison between two methods which were carried out by the same researchers. It is acknowledged that a study using different observers and a larger sample is required to draw firm conclusions, however there are both subjective and objective reasons for believing that the comparison carries weight.

1. The two researchers came away from the research surprised at the high correlations because the two processes did not 'feel' linked. No totals were counted before all observations were complete and the 'gut feeling' ranking had been made. A large number of ticks on a schedule did not necessarily translate to higher ranking because different items had different mean numbers of ticks, and some were inversely rated. In any case, remembering totals on six items between seven lessons spread over a two-week period would have been impossible.
2. The schedule recorded observable behaviours that occurred in a particular lesson only, whereas the 'gut feeling' was based on an assessment of potential and took into account the context (e.g., lesson content, class size, use of groups, or behaviour). Accounts of the 'gut feeling' process did not mention any of the observation items. This was backed up statistically in two ways. First, the two 'gut feeling' rankings had a higher correlation than that between either 'gut feeling' and the schedule. Second, the correlation between 'gut feeling' and the student engagement (0.71) was higher than between 'gut feeling' and the schedule (0.54). General student attention is acknowledged to be part of a subjective impression.
3. Each observer completed only half of the schedule and thus did not have a full picture of what was, in any case, a restricted set of observations. The correlations between each researcher's 'gut feeling' and his/her half-set of observations was lower than between 'gut feeling' and the full set of observations.

Analysis of the 'Gut Feeling' Ranking Process

The researchers' 'gut feeling' ranking was based on their impression of how effective the pre-service teacher would be in the future using the observation of one lesson. This was different from a ranking based on the effectiveness of the lesson itself. Attempting to view the lesson as an indicator of future effectiveness allowed more flexibility so that, (even though the lesson may not have been effective), actions, attitudes or styles could indicate that the teacher would become an effective mathematics teacher.

After each lesson the researchers made written accounts of the things which they noticed and took into account. An analysis of these accounts was done by grouping the positive and negative statements into key themes. Six themes emerged: (a) presentation of material, (b) connections with students, (c) promotion of student thinking, (d) mathematics content, (e) classroom/lesson organisation, and (f) nature of the class being taught. Each theme is discussed in relation to how it influenced the researcher's decision.

Presentation of material. The researchers made comments about the clarity and structure of the presentation of material within the lesson. Examples of the observers' statements in this theme referred, for example, to the way the teacher demonstrated a concept with visual aids, explained an example, or utilised the chalk-board. Warmth and confidence in teaching was also mentioned. Teachers who were described as having a clear, organised structure were generally rated more highly.

Connections with students. How the teacher related to the students was often referred to in the comments about the lessons. Eye contact, allowing student input, and offering individual help, all helped to promote good connections with the students. A number of the teachers discussed their own thinking and learning with the class as a way to 'get on side' with the students. Using students' names promoted good connections with students, helped by using strategies such as seating plans or paper nameplates.

Promotion of student thinking. How the teachers promoted student thinking had a major influence on how the researchers made decisions about teacher effectiveness. Some teachers showed an awareness of how the students were thinking and used techniques that helped to improve the students' understanding. The researchers also noted times when the teachers had missed an opportunity to further the students' understanding. Also included in this theme were how the teachers monitored student thinking. All the teachers moved around the room but the difference lay in what they did during this time. The teachers who were ranked as more effective were able to monitor the whole class while working with an individual or group.

Mathematics content. The researchers described the way in which the mathematics was integrated into the lesson. The teacher's ability to break down the mathematics to make it understandable to the students was a key factor. The researchers valued lessons where the mathematics was presented in reference to the students' thinking rather than in a predetermined way. The appropriateness of the level of the work, encouragement of student participation and activities which involved students thinking for themselves, were considered important. Real world examples or applications were also noted. A number of the pre-service teachers made mathematical errors during the lesson. Some errors were minor and were sorted out without difficulty during the lesson. Other errors showed a lack of understanding of the mathematical concept and were not corrected during the lesson. If the teacher became aware of the error, and acknowledged the need to review understanding, the error tended to be less of a worry to the researchers.

Classroom/lesson organisation. The pre-service teachers tended to use the same classroom organisation that their associate teacher used, although two pre-service teachers used group work and changed the seating in the classroom. In these cases it was clear from the teachers' planning that these changes in classroom environment were linked to the objectives of the lesson. Making such a change brings with it a certain amount of risk and this was positively judged by the researchers. Transitions between different parts of the lessons were also mentioned, especially when the researchers felt that they were not handled very well.

Nature of the class being taught. Several factors that the researchers wrote about concerned the difficulty in making rankings when the classroom situations were different. Two of the lesson observations were greatly affected by unforeseen changes. In one case half the class was away on an outing, whilst in another, the observed lesson was the second of a double-period. The researchers needed to make their comparisons while keeping these situations in mind. The type of students in the class and the behaviour of the students affect the learning opportunities. The researchers therefore noted how well the teacher coped with these difficulties and took this into account when ranking. The top stream classes provided fewer opportunities to show classroom management skills, while in a few of the mixed ability classes, class management skills were an essential part of the teacher's being able to create learning opportunities.

Overview of the ranking process

The themes identified provide an overview of the factors that the researchers felt were important when ranking the student teachers. The variety and differences in the lessons provided a rich array of impressions which the researchers attempted to take into account when making their 'gut feeling' ranking. It is difficult to establish exactly how each of the themes influenced the final ranking.

After the ranking process was complete, discussions revealed other ways of thinking about the lessons. One way in which the researchers coped with the complexities of the lesson was to use their own teaching practice on a good day, as a measuring stick. If the pre-service teacher missed an opportunity to extend an idea, to motivate a student, or to incorporate student thinking into the lesson that the researcher might have taken up, the effectiveness of that teacher tended to be marked down. If the pre-service teacher surprised the researchers with something that they would have missed, or demonstrated some new technique, then they tended to mark the effectiveness up. The observers' experience as mathematics teachers and their beliefs about teaching and learning therefore influenced the way in which effectiveness was judged. Berliner (1987, p. 95) comments about the influence of the observer's experience and knowledge:

> a knowledgeable judge evaluating teachers' effectiveness is expected to possess well-grounded beliefs about what might be reasonable levels of performance under the conditions that exist in a particular instructional setting.

A second way in which the researchers viewed the lesson was in terms of its *thickness*. Some lessons had, by virtue of both their 'flavour' and the positive events that occurred, a sense of depth and substance, whilst others felt 'thin' and insubstantial. As an analogy for television viewers, thickness may be equated with the difference between an insubstantial 'soap' and a well-constructed drama series.

The process of making a 'gut feeling' ranking was much more than just counting up the number of positives and taking away the number of negatives. The interaction of many factors affected the overall impression of the lessons and the ranking of the pre-service teachers' future effectiveness.

'Gut feeling' Ranking and the Extended Teaching Triad

A comparison between the factors considered by the observers and the theoretical model (see Figure 1.) shows a good match. This validation of the model can also be viewed as evidence that the researchers were ranking the effectiveness of how well the lessons promoted good mathematics learning.

The top vertex of the teaching triad, *management of learning*, includes the ways that the classroom and lesson are organised to establish learning opportunities for the students to build understanding. Most of the factors that the researchers considered when they made the rankings, fit into the top vertex of the triad. This indicates that the researchers were concerned with the way in which learning was managed during the lesson, and reflected the prime importance of this vertex in the theoretical model.

The *mathematical challenge* vertex involves stimulating mathematical thought through problem solving at an appropriate level to engage students. The researchers identified opportunities for the learners to be involved in solving problems that were at an appropriate level for the students. Whether the explanations and examples were mathematically correct was an issue that the researchers considered. A key factor in deciding the teachers' future effectiveness was whether they showed the ability to review their own thinking and understanding.

The third vertex involves creating a nurturing, encouraging environment that incorporates the needs of individual students. The researchers looked at the ways the teachers could establish communication in the classroom through personal *connections with students*. The way in which the lesson was adapted to the students' ability level and interest, and monitoring of student thinking, are both strategies that showed that the teacher was sensitive to the needs of students.

Conclusion

The observed lessons were different in many ways. This richness of interactions and activities made comparisons between the lessons and the teachers very difficult. Despite all the differences and complexities, however, the analysis of the researchers' 'gut feeling' rankings shows how they were able to monitor many different aspects of the lessons and take them into account when ranking the pre-service teachers. Each lesson raised new considerations which needed to be weighed up when comparing it to a different lesson. The two 'gut feeling' rankings were strongly correlated with each other, indicating that there was a high level of agreement between the two researchers.

The observation schedule, in contrast to the 'gut feeling' ranking, considered the teacher behaviours in a very different way. Each item had an equal weighting and was considered in the same way for each observed lesson. The individual characteristics and differences in the lessons were, therefore, not taken into account. Furthermore, the schedule was limited by practical considerations from incorporating all aspects of effective teaching behaviours.

The two methods of attempting to assess the future effectiveness of pre-service teachers showed high correlation. Subjective judgments of experienced mathematics teachers (the researchers) were positively related to the results of an observation schedule. It is concluded that subjective assessments by classroom-experienced and theoretically literate observers are probably preferable to mechanical observations of behavioural items, both on practical grounds and because of the greater context flexibility of such judgments. These are particularly important considerations during pre-service teacher development. However it is noted that this is indicative research only, and needs to be confirmed by a wider study with more objective design features.

The analysis of the researchers' 'gut feeling' ranking in relation to the theoretical model was able to show that the ranking was based on how well the pre-service teachers were able to provide an environment which promotes good mathematics learning. The teachers needed to engage the students in appropriate mathematics activities and take into account how the students were thinking and learning. Therefore, evidence of the pre-service teacher's subject matter knowledge and pedagogical content knowledge during the lesson was an important part in how s/he was ranked.

This small study attempted to evaluate the future effectiveness of only seven pre-service teachers using observations of just two lessons for each teacher. Although it would clearly have been preferable to use more lesson observations to make such evaluations, this study attempts to work within the realities that teacher educators often face. It would be interesting to repeat this study with other groups of pre-service teachers and other experienced teachers as researchers to see if similar, strong correlations are found.

References

Berliner, D. (1987). Simple views of effective teaching and simple theory of classroom instruction. In D. Berliner & B. Rosenshine (Eds.), *Talks to teachers*. New York: Random House.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*(4), 473-498.

Even, R., & Tirosh, D. (1995). Subject-matter knowledge and knowledge about students as sources of teacher presentations of the subject-matter. *Educational Studies in Mathematics, 29*, 1-20.

Gagne, E., Yekovich, C., & Yekovich, F. (1993). *The cognitive psychology of school learning*. (2nd ed.). New York: Harper Collins.

Horring, J.G. (1998). *The assessment of student teachers as effective teachers of mathematics using limited classroom observations*. Unpublished Masters Dissertation, The University of Auckland.

Jaworski, B. (1992). Mathematics teaching: What is it? *For the Learning of Mathematics, 12*(1), 8-14.

Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal of Research in Mathematics Education, 20*(1), 52-75.

Leinhardt, G., & Smith, D. (1985). Expertise in mathematics: Subject matter knowledge. *Journal of Educational Psychology, 77*(3), 247-271.

McDiarmid, G. W., & Wilson, S. (1991). An exploration of the subject matter knowledge of alternative route teachers: Can we assume they know their subject? *Journal of Teacher Education*, 42(2), 93-103.

Paterson, J., & Barton, B. (1998). *Investigating mathematics pre-service teacher education: Research towards optimising processes in teacher education*. Report for The New Zealand Ministry of Education.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*, 4-14.

Stein, M., Baxter, J., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal, 27*(4), 639-663.

Sullivan, P., & Mousley, J. (1994). Quality mathematics teaching: Describing some key components. *Mathematics Education Research Journal, 6*(1), 4-22.

Tomic, W. (1989). Teaching behaviour and student learning outcomes in the mathematics classroom. *Journal of Educational Research, 82*(6), 339-347.

Wilson, S., Shulman, L., & Richert, E. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking*. London, UK: Cassell Educational.

Authors

Julia Horring, Mathematics Education Unit, Department of Mathematics, The University of Auckland, Auckland, NZ. E-mail: <juliahorring@hotmail.com>

Judy Paterson, Mathematics Education Unit, Department of Mathematics, The University of Auckland, Auckland, NZ. E-mail <paterson@math.auckland.ac.nz>

Bill Barton, Mathematics Education Unit, Department of Mathematics, The University of Auckland, Auckland, NZ. E-mail: <barton@math.auckland.ac.nz>