TEACHERS DIFFER IN THEIR EFFECTIVENESS

Peter Sullivan
La Trobe University

Andrea McDonough
Australian Catholic University

This article reports an unanticipated result from a large scale project that sought to examine ways of improving numeracy learning. We found that there were marked differences in achievement between classes, irrespective of geographic or socioeconomic variables, and we argue here that these differences are attributable to the teachers. In other words, teachers influence the learning of their students and there are marked differences between teachers in their effectiveness.

INTRODUCTION

One sometimes gets a sense that the community at large, and even some education bureaucrats, believe that anyone can teach, or that teachers are born, or that new teachers can learn to teach just by watching others. The assumptions behind those beliefs seem to be that effective learning is the result of curriculum policy statements, resource provision, school leadership, or other factors amenable to central policy decisions. We argue the opposite: children from similar backgrounds have markedly different experiences at school, and these differences are attributable to their teachers.

Of course teachers have responsibility for more than the learning of the students, but this learning is clearly fundamental to the teacher's role, and the following argument is based on an assumption that a measure of the effectiveness of teachers is the growth in learning of their students.

There is a tradition that connects learning to teaching effectiveness. For example, Darling-Hammond (1997), in proposing a case for teacher education, summarised research on data from 900 school districts in Texas that found that 40% of the measured variance in student achievement across grades 1 to 11 was due to teacher expertise. She argued that, even after controlling for socio-economic status, the large differences in achievement between "black and white were almost entirely accounted for by differences in the qualifications of their teachers" p 8.

Such research can be unconvincing because of the way that achievement is sometimes measured using multiple choice items on tests of low level skills. The following argument draws on results from a large scale numeracy project to examine this issue using carefully collected, ample and representative data, based on a research informed view of a particular domain of mathematics, measurement of length. We conclude that teachers make a difference to the learning of the children in this domain, and we suggest that such differences may be evident in other domains as well.

THE ROLE OF THE TEACHER IN IMPROVING LEARNING

The object of educational reform is improved learning of the children. Clearly the instruments of any improvement in learning are teachers. Yet we suspect that there is an assumption that teachers are more or less equally skilled, or that reforms are to some extent independent of the teacher. Certainly the structure of terms and conditions of employment, and of professional development programs themselves, do not seem to make assumptions about differences between teachers.

One possible explanation for this lack of acknowledgement of the importance of the teacher is connected to the way that learning is conceptualised. It is common for reforms to be informed by a social constructivist perspective that was summarised by Ernest (1994) as recognising that knowing is active, "individual and personal, and that it is based on previously constructed knowledge" (p. 2), and that the knowledge is not fixed, rather it is socially negotiated, and is sought and expressed through language. It is possible to infer that learning is independent of the teaching. Yet Ernest saw the teacher as central listing among the pedagogical implications that teachers need to be sensitive to learners' previous constructions, to seek to identify errors and misconceptions, to foster metacognitive techniques, and to acknowledge social contexts of learners and content. Likewise, Cobb and McClain (1999) argued that teachers should have a clear impression of the direction that the learning of the individuals and the class will take. They proposed that the teacher should form for the class an "instructional sequence (that) takes the form of a conjectured learning trajectory that culminates with the mathematical ideas that constitute our overall instructional intent" (p. 24).

Another possible explanation for the underemphasis of the teachers' role might be a belief that the primary determinant of students' learning is their family or cultural background. Tschannen-Moran, Hoy and Hoy (1998) reviewed research on teacher efficacy that contradicted this. Teacher efficacy refers to the extent that teachers believe that they can influence how well students learn, independent of their motivation, their background, their prior learning or other factors. Tschannen-Moran et al. cite a range of studies that connect high scores of efficacy by the teachers with higher achievement of their students.

In other words, teachers have an active role in promoting learning, and this is connected to their beliefs about the nature of learning and the nature of learners, but this active role is not necessarily acknowledged in educational policy and practice.

SEEKING A MEASURE OF THE EFFECTIVENESS OF TEACHERS: SOME RESULTS FROM THE ENRP

To explore the impact of teachers on student learning, we draw on results from the Early Numeracy Research Project[1] (ENRP) that investigated mathematics teaching and learning in the first three years of schooling, involving teachers and children in

35 project ("trial") schools and 35 control ("reference") schools (for details see Clarke, 2000; Sullivan, Clarke, Cheeseman, & Mulligan, 2001). One source of data was a one-to-one interview over a 30 to 40 minute period with every student at the beginning and end of the school year (Feb/March and November respectively). Interviews were conducted by the classroom teachers, who were trained in all aspects of interviewing and recording. The processes for assuring reliability of scoring and coding are outlined in Rowley and Horne (2000). The data reported in this paper were collected in the year 2000, the second year of the project.

The data from this project arise from intensive interviews with large numbers of children, with trained interviewers, and experienced coders, with double data entry, and using a framework for learning based on interpretation of research. We argue that these data provide a reliable measure of learning, and a further perspective on previous research on the ways that teachers influence student learning.

Learning of Length Measurement

The data, on which the argument proposed below relies, is based on growth in students' learning of aspects of length, one of nine domains on which data were sought within the ENRP. Growth points for the key aspects of learning length were proposed, and assessment tasks administered individually were developed, after consideration of some common findings in length research.

Much research on learning of measurement is influenced by work of Piaget and his colleagues, who identified stages of development in coming to understand measurement concepts such as conservation, the idea of a unit, transitivity, and iteration, with focus on development of cognitive abilities within individuals. Nunes, Light, and Mason (1993) present an alternative Vygotskian perspective in which competence is not fixed within individuals but cultural representations play a mediating role in development of understandings. In measurement, such cultural representations are often conventional rather than conceptual. In developing the assessment framework, we were concerned that Piagetian stages did not lead to clear teaching and assessment guidelines (see Carpenter, 1976; Kamii & Clarke, 1997).

Through the framework and interview for the domains of Length, the ENRP gives emphasis first to whether children show awareness of length through tasks that facilitate use of language. Measurement by direct comparison is then considered. We intended that both the growth point and the task prompting direct comparison be inclusive and suggestive of conservation as well.

The next growth point requires the use of a non standard but consistent unit to quantify a length measurement, including some of the requirements for iteration. The final two points relate to the use of standard units and their application. We did not include transitivity directly in the framework.

The six proposed growth points for length are as follows:

No apparent awareness of the attribute of length and its descriptive language. (*Not apparent*)
Awareness of the attribute of length and use of descriptive language (*Awareness of the attribute*)
Compares, orders, & matches objects by length (*Comparing lengths*)
Uses uniform units appropriately, assigning number and unit to the measure (*Quantifying lengths*)
Uses standard units for estimating and measuring length, with accuracy (*Using standard units*)
Can solve a range of problems involving key concepts of length (*Applying*).

These were developed as a conjectured sequence. It was assumed that students will follow different pathways in their learning, but nevertheless the intention was to describe the learning trajectory of the majority of students. To illustrate the style of the assessment, the following were the first two tasks posed (with italics indicating what to do, and the normal text indicating what to say):

The string and the stick
Drop the string and the skewer onto the table.
A) by just looking (without touching), which is longer: the string or the stick?
B) how could you check? *(touching is fine now)*
C) so, . . . , which is longer?
The straw and the paper clips
Get the straw and show the child the eight (5 cm) paper clips.
Here are some paper clips. Here is a straw.
A) measure how long the straw is with the paper clips. . . . *(if child hesitates)* use the paper clips to measure the straw.
B) what did you find? *(no prompting)*
If correct number is given (e.g., 4), but no units, ask "4 what?"

The interviewers proceeded through the interview in order, but moved directly to the next domain (e.g., Mass) if the student answered a question incorrectly. A coding rubric was used to score the students' responses.

Entry Level Students and the Length Growth Points

The issues of interest here are how the students responded to the tasks, and how they improved over the year. Table 1 presents the percentage of entry level students (commencing at age 5) in project schools rated at each growth point in both March and November, near the start and end of the school year respectively.

Table 1: Entry Level Students (%) on Length Growth Points (March and November)

	March (n=1488)	November (n=1484)
Not apparent	18	3
Awareness of attribute	6	1
Comparing lengths	62	50
Quantifying lengths	13	43
Using standard units	0	3
Applying	0	0

The students are spread over the growth points, even at the start of their schooling. It is suspected that these differences are due to home or other specific experiences prior to school, recognising that language would be a contributing factor in some cases.

Given the low number of students who have reached the growth point, *Awareness of the Attribute,* the first two points are combined for the subsequent discussion, and termed *Not yet comparing*. To examine further the improvement over the year, individual responses in March are compared with those in November. Table 2 presents these as numbers of students.

Table 2: Comparisons of Student Growth from Respective Growth Points (n=1369)

	November		
March	Not yet comparing	Comparing lengths	Quantifying lengths or beyond
Not yet comparing	37	187	92
Comparing lengths	11	453	400
Quantifying lengths		40	149

Under 4% of the students went backwards, 47% stayed at the same growth point and 50% improved. The argument here is about growth, and to explore this further, we examine the largest set of students whose rating did not improve over the year.

The 453 students who were at *Comparing lengths* in March who were still at that point in November represent 33% of this group. These "*Comparing length*" students have not had whatever experiences may have been necessary to grow to *Quantifying length*. To explore further the nature of the development and potential of such students, some other aspects of their learning of mathematics were explored. Table 3, for example, compares their responses on the *Counting* domain in November to those of the group overall (n=1369).

Table 3: The *Comparing length* students (%) on *Counting* in November

Growth point descriptor	The *Comparing Length* students (n=453)	The whole group (n=1369)
Not yet able to count to 20	7	6
Can say number sequence to 20	4	4
Can count a collection of 20 objects	62	57
Counts forward & back by 1s from *x*	13	15
Can count from 0 by 2, 5, 10	13	17
Can count from *x* by 2, 5, 10	1	1

The *Comparing length* students have a similar profile on the *Counting* domain to the group overall. This suggests that there is little direct connection between *Counting* and *Comparing length*, at least at this level. For example, over one quarter of these *Comparing length* students were able to count forward and backwards from various starting points. Only 11% were not able to count a collection of 20 teddies. That these students could not progress to *Quantifying length* over the year is not due to inability to count.

Similarly, in the other domains of *Place Value*, *Addition and Subtraction*, *Multiplication and Division*, and *Time*, the profile of these *Comparing length* students was similar to the group overall. This suggests that to develop to the next growth point the students need particular experiences associated with the learning of

length, rather than general mathematical development. In other words, the mathematics experiences that these students have had has not resulted in across the board changes but rather that improvements are a result of domain specific experiences.

To explore this further, we examined the improvement by classroom on the *Length* domain for Entry level students. As Table 4 shows, there are marked differences between the classrooms. Note that the table presents results for Entry level students who are in the 54 classrooms with only entry level students, and the results of students who are in multi-age classes are not included in the analysis.

Table 4: Percentage of students per class in Entry level classes who improved (n=54)

Percentages of students per classroom improving	Number of Entry level classes
Up to 20% of the students per classroom improving on Length	5
More than 20 up to 40% of the students per classroom improving on Length	22
More than 40 up to 60% of the students per classroom improving on Length	19
More than 60 up to 80% of the students per classroom improving on Length	6
More than 80 up to 100% of the students per classroom improving on Length	2

To be more specific, there was a class in which 24 out of 27 children improved, another where 21 out of 24 improved, whereas there was a class where 2 out of 25 improved, and others where 3 out of 24, 5 out of 28, and 2 out of 21 improved respectively. These are very noticeable differences and are unlikely to have occurred due to chance factors. An examination of the schools and other factors indicated that neither being effective nor being less effective teachers in terms of promoting improvement was dependent on school size, socio-economic community, student language background, or years of experience of the teacher.

This suggests it is possible for students at this level to move through the growth points but it is teacher dependent. A similar analysis of results of Grade 2 teachers on Length showed similar although less striking differences. This result is also similar to that reported by Sullivan, Clarke, Cheeseman and Mulligan (2001) with respect to differences in effectiveness in the teaching of multiplication and division at Grade 2.

Characteristics of teachers who made a difference

To explore the nature of these differences between the results of the classrooms, six teachers who had higher proportions of students improving on length in each of the first two years of the project were interviewed. Among other items, each of the teachers was asked to describe an example of an activity they used in their teaching of Length. A particular feature of the teachers was that they seemed able to describe rich experiences for the students and the purpose of those experiences. For example, the following is an extract from the response of one of the effective teachers:

> Well my favourite one ... "The Long Red Scarf" and I based the series of lessons on that covering the different growth points ... I had a whole lot of teddies that the children made scarves for and we compared lengths and then we actually taught them how to measure

> using blocks and bears and things and we measured our scarves and ... language because a lot of our children do not have the language so even simple things like longer and shorter ...
>
> I started with reading the book and we talked about scarves, then I brought in scarves and we put them on the floor in the middle of the big circle and I spread them out haphazardly and I said well "which scarf here is the longest?" and the children said "have a guess at anything sort of thing" and I have got a very bright boy who said "no you can't do it like that, you have to line them up" and he lined them up and then someone else said "no but you've got to match them at the end" so they matched them at the end ... so we got lots of language.
>
> The second lesson ... we actually said we were going to make scarves so I gave them paper. ... they had to make a scarf long enough to go around their teddy ...they proceeded to make their scarves and some of them even decorated them and then we actually compared lengths again so that was all one lesson, they came back and they put their scarves down and we talked about who had the longest and some of them had very long scarves because they had bigger teddies and some had short scarves.

It seems that this teacher had a clear vision of the experiences that were needed, was able to engage the students, and was not deterred from such a rich experience by the unfamiliarity of some of her students with the language demands. Other effective teachers gave similarly rich examples. Another common theme, was that these teachers were prepared to probe the thinking and understanding of the children. For example, in response to the same prompt another teacher said:

> I always try and make sure that there's a sharing of findings at the end of each session ...and I always ask the kids "how did you obtain such a result?" or "how did you get your answer?". So there's that constant reflection ... "if you measured your foot and you found out that it was 22" ... also I try and challenge the kids by asking them "if we've all measured our feet and we've all measured the length of a basketball court and we've all got a different response, why is this?" so I'm actually getting them to think a little bit beyond just obtaining a result.

In other words, the teachers seemed to be aware of characteristics of rich experiences and how to use those experiences to extend the students' thinking.

CONCLUSION

In a detailed study of a small aspect of a large project, an unanticipated result emerged. Teachers who were given extensive professional development including teaching advice some of which focussed on teaching of measurement, participated in structured planning teams, and released from teaching to interview all of their students, differed substantially in the extent to which their students improved in defined growth points in Length. That teachers make a difference is supported by other studies. The data presented here suggested that the differences between the most effective and least effective teachers are substantial. Effective teachers seemed able to articulate focused, developmentally appropriate and engaging activities for their students, and engage them actively in interrogating those experiences.

Before substantial policy decisions are taken more research is required. If further research confirms these results then there may be some important and somewhat challenging implications. If there are teachers who are substantially more effective than others, presumably steps could be taken to find out who those teachers are and acknowledge their effectiveness in some way. If there are teachers who are substantially less effective than others, we could seek to find out who those teachers are, and find ways to assist them to understand the impact of their teaching, and to examine strategies that might assist them to become more effective. We fully recognise the dangers of simplistic solutions to such issues, but suspect that it is possible to devise strategies that avoid undesirable side effects.

REFERENCES

Carpenter, T. (1976). Analysis and synthesis of existing research on measurement. In R. Lesh (Ed.), *Number and Measurement* (pp. 47–83). Athens, Georgia: ERIC/SMEAC, University of Georgia.

Clarke, D. M. (2000). The early numeracy research project: Some insights from an exciting first year. In Department of Education, Employment and Training. (Eds.), *High expectations: Outstanding achievement* (Proceedings of the Early Years of Schooling P-4 conference, CD-ROM). Melbourne: DEET.

Cobb, P., & McClain, K. (1999). Supporting teachers' learning in social and institutional contexts. In Fou Lai Lin (Ed.), *Proceedings of the 1999 International Conference on Mathematics Teacher Education (pp. 7–77)*. Taipei: National Taiwan Normal University.

Darling-Hammond, L. (1997). Doing what matters most: Investing in quality teaching. National Commission on Teaching and America's Future. New York

Ernest, P. (1994). Varieties of constructivism: Their metaphors, epistemologies and pedagogical implications. *Hiroshima Journal of Mathematics Education, 2*, 1–14.

Kamii, C., & Clarke, F. B. (1997). Measurement of length: The need for a better approach to teaching. *School Science and Mathematics, 97*(3), 116-121.

Nunes, T., Light, P., & Mason, J. (1993). Tools for thought: The measurement of length and area. *Learning and Instruction, 3*, 39-54.

Rowley, G., & Horne, M. (2000, December). Validation of an interview schedule for identifying growth points in early numeracy. Paper presented to the Australian Association for Research in Education Annual Conference, University of Sydney, New South Wales.

Sulllivan, P., Clarke, D., Cheeseman, J., & Mulligan, J. (2001). Moving beyond physical models in learning multiplicative reasoning. In M. van den Heuvel-Panhuizen (Ed.). *Proceedings of the 25th annual conference of the International Group for the Psychology of Mathematics Education (Vol. 4 pp.233-240)*.

Tschannen-Moran, M., Hoy, A, & Hoy, W. (1998). Teaching efficacy: Its meaning and measure. *Review of Educational Research, 68* (2), 202-248.

[1] The *Early Numeracy Research Project* (ENRP) is a collaborative venture between Australian Catholic University, Monash University, the Victorian Department of Employment, Education and Training, the Catholic Education Office (Melbourne), and the Association of Independent Schools Victoria, and is directed by Doug Clarke, and the team includes Barbara Clarke, Jill Cheeseman, Ann Gervasoni, Donna Gronn, Marj Horne, Andrea McDonough, Pam Montgomery, Anne Roche, Glenn Rowley and Peter Sullivan (who was at Australian Catholic University for most of the project).