Video Study of Mathematics Teaching in Chile

Roberto Araya, Center for Advanced Research in Education, Universidad de Chile,
roberto.araya.schulz@gmail.com
Pablo Dartnell, Center for Advanced Research in Education,
Department of Mathematical Engineering and Center for Mathematical Modeling,
Universidad de Chile, dartnell@dim.uchile.cl

ABSTRACT

A detailed characterization of school mathematics teaching patterns using the videos that teachers recorded themselves as part of the 2005 Chilean national teacher assessment program is presented. This is a new compulsory teacher assessment program, but in 2005 teachers volunteered to participate in a first version of the program. This research work includes the analysis of more than 700 forty minute video recordings, both from elementary and high school classes, from different regions of the country, containing lessons on the algebra, geometry, data and numbers strands. One 4 minute slice of each video was studied. The slice was randomly selected from the following moments of the class: first 4 minutes, from minute 10 to 14, from minute 20 to 24, from minute 30 to 34 and the last 4 minutes. More than one hundred variables were codified by independent coders. The codification methodology was successful and stable across different coders, making the analysis huge amounts of data possible. The main findings were: very little autonomous student participation (only one mathematical question made by students per 40 minutes class), teachers neither present nor discuss any proofs, no use of information technology, almost no use of textbooks, almost no explicit use of metaphors.

INTRODUCTION

This research work intended to study mathematics teaching practices in primary and secondary school education in Chile: how mathematics is actually being taught in the classrooms. With that information at hand, comparisons with other countries could be made, but also, and more important, recommendations for public policy towards improving the quality of pre-service and in-service teacher training could be made in order to improve the teachers classroom practices and in turn, their students achievements.

Considerable amounts of data were already available, consisting of hundreds of video recordings made for the 2005 Chilean Teachers Professional Assessment (Manzi, 2007). Mathematics teachers (as well as teachers of other disciplines) that chose to be evaluated in 2005 had to have a video of one of their classes recorded as part of the assessment requirements. These videos had been stored by the Ministry of Education after having been used for their original purpose. We could get a hold of a total of 720 of those videos, so that they could be analyzed in search for a characterization or description of the Chilean teaching practices in Mathematics.

The analysis was chosen to focus on didactic aspects such as the modality of student working arrangement (whole class work, small group work, or individual work), degree of student participation in class and level in which the teacher encourage it (Mathematics questions asked, for instance), motivational strategies and didactic strategies particular to Mathematics, such that proofs and use of mathematical metaphors (Lakoff and Nuñez, 2000; Richland, Holyoak and Stigler, 2004; National Mathematics Advisory Panel for multiple representations, 2008). The class dynamics was also quantified, attempting to quantify and find characterizations of the beginning, middle and ending of the lectures. Finally, the use of different types of technological aids, such as board, computers, text books, paper and scissors, etc, was quantified.

Theoretical and methodologically, only objective measurements, that cold be consistently repeated by a second observer, were made to account for the different aspects to be measured. Also, to be able to work with the large amount of videos we had at hand, for each one only a 4 minute (or sometimes a 2 minute) slice was randomly selected viewed, corresponding to either the first 4 minutes of the class, minute 10 through minute 14, minute

20 through minute 24, minute 30 through minute 34, or minute 40 through minute 44, given that in Chile there are 45 minute lectures.

The working hypothesis was essentially that there would be rather scarce use of proofs and metaphors, and that the percentage of use of modern technology such as computers would be low. On the other hand, polls conducted with teachers report that they, in a great percentage, use the textbooks that the Ministry gives for free to all children in public schools (Ministerio de Educación de Chile, 2002; Universidad de Chile, 2006). However as we will see, the results were appalling: no evidence of proofs, metaphors or computers, and the textbooks were almost not used at all. Also sadly surprising was the fact that student participation, in the way of (mathematics) questions asked, was extremely low: an average of one student question in the whole lecture time.

As expected also, some patterns emerged in the way that there was a clear characterization of the three thirds of the lecture, and the primary and highs school teachers have very distinctive didactic differences.

The fact that the short class slices would allow us to gather relevant and statistically significant information about the pedagogical practices of mathematics teachers was also a working hypothesis that we could confirm within the study. Examples in the literature that that encouraged this approach are, for instance, the SPAFF methodology that is able to predict accurately future behavior of married couples by watching 3 minute videos of them before their marriage (Gottman, 2000; Coan and Gottman, 2007) and the Ambady and Rosenthal study of college teacher assessment by students (Ambady and Rosenthal, 1993; Gladwell, 2005).

THEORETICAL FRAMEWORK

There are various theoretical approaches that can be adopted to design and analyze a study of the teaching/learning situation in the school classroom. The one chosen for this study had into consideration that the material to be studied consisted on hundreds of video recordings already made. And these videos consisted of one 45 minute class per teacher, and were made at the teacher's request, without any research consideration in mind. Having that sort of material does not really allow for questions involving why some teaching strategy is used, which would probably involve interviews with the teachers after the lecture, or studying in depth the strategy to teach some topics, which might require having whole sequences of lectures, for instance. Instead, the material forced us to search for and choose a very pragmatic theoretical framework, one in which one uses very objective information to give an account of what is globally happening in the classrooms over the country, without really being able to describe or analyse each separate class.

Such a framework had already been used by previous video studies, such as the TIMSS Video Studies (TIMSS, 1999; LessonLab, 2003; Hiebert, Gallimore, Garnier, Givvin, Hollingsworth and Jacobs, 2003; Stigler and Hiebert, 2004) . We have adopted it, and, essentially, within it we try to draw conclusions about what is going on in the classroom situation based only on very objective evidence. Evidence that does not depends on anyone's opinion, but on objective measurements such as number of mathematic questions that the teacher asks, or the amount of time the teacher spends writing on the black (ore white) board. The variables or indicators to be used also enjoy the quality of being repeatable, that is to say, independent of the particular observer. The idea is that if patterns are obtained using such variables, then the conclusions drawn can become very solid evidence of what is happening countrywide, even though probably no judgement or assessment can be made of individual teacher practices, because relevant though more subjective factors can be important when you come to judge an individual class. For such a judgement a different epistemological lens might very likely be necessary (Tuminaro and Redish, 2007; Redish, 2003; Díaz, 2006), and the material available for this study would probably be non suitable for the study.

RESEARCH METHODOLOGY

As described earlier, the class videos had already been recorded without a research study in mind. Therefore our research methodology did not start by dealing with the design of the recordings, as it usually happens with this type of study, but rather with the design of the means for obtaining the most information we could from the already available material.

The videos were available on tape, therefore the first step was to encode them into a digital format that would be both, compact and easy to reproduce so that more than one observer could view the same video at the same time, in different computers. The digital video format XviD was chosen, both for being open source and an efficient equivalent of the DivX standard. Furthermore there is plenty of computer software that is able to play this format, and under a variety of platforms (Windows, Mac, Linux, for instance).

As the videos were encoded, they were given a filename that consisted of the national ID number of the teacher of the class, which was among the few extra information we had available for each recording. It must be pointed out that the ID number is somehow correlated to the age of people, younger people typically has higher ID numbers, therefore this number gave us indications of the approximate age group of the teachers. The videos were classified also by the educational level they belonged to, that is to say, middle school or high school. 78.8% of the tapes were of middle school level and only 21.2% from high school. There were no tapes from elementary school.

In parallel to this encoding, research was performed to find out the variables and different classifications already used in other video studies, and a large number of variables (around 200) were chosen as candidates to be measured in the videos. Later, in an iteration process which used an increasing number of randomly selected videos, from 10 initially to finish with 100, the suitability of variables was tested, and in parallel assistants were trained on the video coding task. Those variables which did not fit the repeatability criterion (no statistical difference was to be found between the values measured by both researchers who recorded those variables), or those that were not able to be measured because they belonged to categories simply absent in the group of videos (for instance, metaphors were not found, therefore variables that distinguishing different types of metaphors did not make any sense for this sample) were dropped, and we were left with a group of 120 variables.

We saw no practical need for using specialized software, such as "Transana", to record the data, because, as we will mention later, only 4 minutes or less were coded for each video, and for those lengths it is sufficiently easy to find the sought information without the need of special software. Instead, the set of variables were divided into two Excel forms, and two researchers were in charge of collecting the data to be filled in each of the forms. Thus, each variable was recorded by two people for each video.

The number of videos to be analyzed was considerably large: 720. Viewing them all in full was impracticable considering that it was estimated that for accurately recording each of the variables to be measured the material would have to be replayed about 20 times, which would give about 15 viewing hours per video, per form and per reviewer. The decision was then made that only a short sample of each video would be processed.

There is plenty of encouraging experience on such time sampling of videos. For instance, Gottman (Gottman, 2000; Coan and Gottman, 2007) has developed a methodology named SPAFF with which he has shown that by recording a 3 minute video of an engaged couple he can predict with great accuracy future (several decades, in fact) behavior of the married couple, including things like if they will stay together or will get divorced. There is also the classic study by Ambady and Rosenthal (Ambady and Rosenthal, 1993), which shows that the students evaluation after viewing a 10 second video of a university professor class proved to be statistically equivalent to what they would write after a semester attending the class. On the other hand, there are studies that claim that much longer observation periods are needed to judge teaching practices. For instance, Shimizu and Yoshinori (Shimizu and Yoshinori, 2003) advocate for a whole sequence of 10 lectures as a minimum unit to study patterns of pedagogical practices. They state that, for instance, homework assignments play a linking role between lectures, and would have a significant educational role.

In our case the videos were pre recorded and we did not have access to any sequence of classes, and we chose to review 4 minute slices from all the recordings available. In fact, we chose to do some checking of the hypothesis that even smaller times might be good enough, and in several videos, slices only 2 minutes long were viewed.

The segments were chosen so that they would be well distributed over the lecture time: one starting at the beginning, one at minute 10, one at minute 20, one at minute 30 and one at minute 40.

Results and analysis

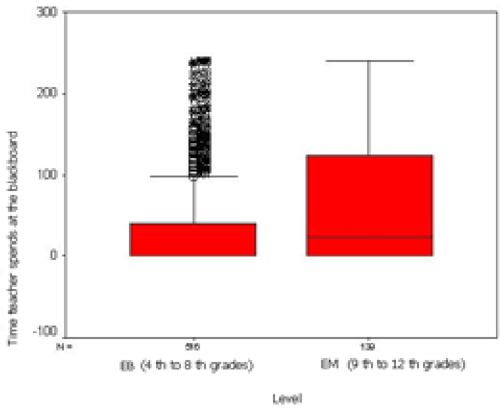
About the general statistical validity of the choices made, both of variables and of short slices, the data obtained is encouraging. From both, the preliminary tests and the final coding, the data shows that the different researchers coding the same variables obtained statistically similar patterns, thus validating the stability and repeatability hypothesis.

Furthermore, there was no statistical difference between the outcomes of the variables measuring total time of an event, for the 2 minute slices and the 4 minute slices (the 2 minute slice times properly scaled, naturally).

However variables counting number of events did not always show the same behavior independent on the slice size. In fact, when the events counted were rather short ones, the variables behaved the same, but for events essentially longer, the behaviors differed significantly.

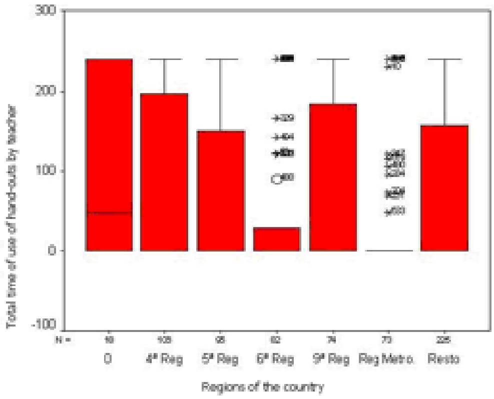
Let us now review the data gathered and the findings that seem to derive from it. Let us first concentrate in variables or groups of them which give us information that one could perhaps think that "common sense" would tell us that the results could not be different from the ones suggested by the data. We think that even if that is the case, having hard empirical evidence of the facts, justifies including this information as relevant results of the study.

The following graph shows the amount of time in seconds, within the 4 minute slice reviewed, that teachers from middle school (EB) and high school (EM) spent at the blackboard. It can be seen that high school teachers spent considerably more time at the board.



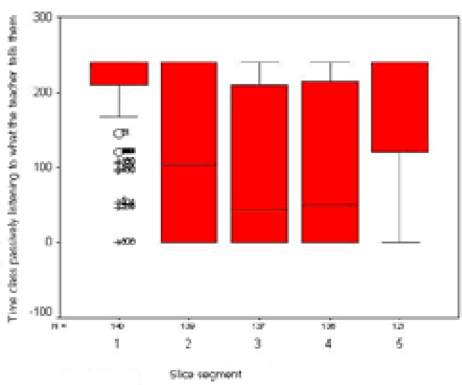
We also gathered similar data regarding the fact that high school teachers spend more time than their middle school peers writing mathematics, that they have their students spend less time in activities involving paper, scissors, cartons, etc, that their students spend more time solving mathematics problems, and that they make less eye contact with the students. All this tells us that there are clear differences between didactic strategies of middle school and high school teachers.

Looking for regional differences in didactic strategies, we found that there is a significant difference between the teachers from the "Región Metropolitana", which is the region where Santiago, the Capital City, is located (and which, incidentally, concentrates more than 1/3 of Chile's total population) and the rest of the country. For instance, the following figure shows the amount of time (within the 4 minute slices) spent working on teacher hand-outs. It is obvious from the graph that the use of hand-outs by teachers at Región Metropolitana is neglectible compared with teachers in other regions.



Similar data shows that teachers from Región Metropolitana approach the students a lot less than the rest (to check and supervise their work), but they ask their students considerably more mathematical questions (and therefore the number of answers to mathematical questions by the students is higher).

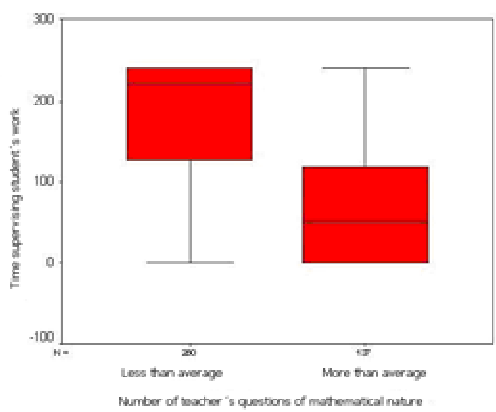
Regarding the different sections of the class, meaning beginning, middle and ending, as could be guessed in advance, there are measurements that differentiate them, for instance, next graph shows that in segment 1 (0-4 minutes) and 5 (from minute 40) there is no independent student work, and most of the time the whole class is passively listening to what the teacher tells them.



Some more "very obvious" information of a similar sort as the previous one can also be extracted from the data. Given the obvious nature, we only mention it here, without giving the associated numbers. Teachers walk more around the students seats supervising their work when they are not lecturing to the whole class. Teachers approach more the students to check their work when they are working with concrete materials (cutting and pasting papers, or drawing pictures). There is a different didactic pattern in the geometry strand: as opposed to all the rest of the strands (algebra, chance, numbers), the teacher shows more objects, or draw more pictures, and the students spend more time cutting and pasting papers.

When looking at what happens in the intermediate slices (number 2: from minute 10, number 3: from minute 20, and number 4: from minute 30) some predictable, although not necessarily desirable, correlations appear.

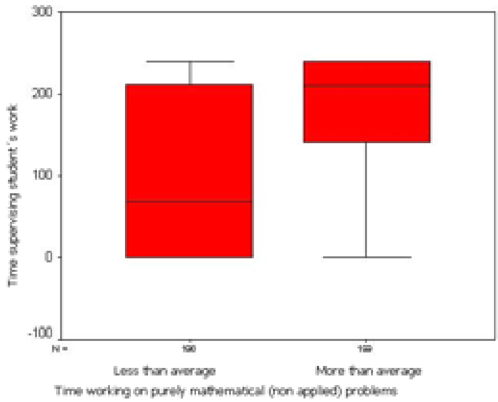
The following graph shows that when a teacher asks fewer questions of mathematical nature (fewer than the mean) to their students, he o she spends considerably more time supervising student work than the teachers that ask more (mathematics) question than the average:



Thus, teachers seem to be divided in two disjoint groups marking clearly differentiated didactic patterns: the ones who ask math questions and the ones that supervise student work.

It can also be seen from the data that, for instance, when teachers asked more than the average mathematical questions to the students, they also called more students to work on the blackboard.

Another phenomenon found when observing slices 2, 3 and 4, is that when students spend more time than the mean working on purely mathematical (non applied) problems, they spend less time doing activities with paper and scissors (not really surprising), but the teachers in turn spend more time supervising their work. The following graph illustrates this last statement, by presenting teacher's time supervising students for the cases of less and more than average student work on purely mathematical questions.



From segments 2, 3 and 4, it can also be observed that younger teachers seem to use some different strategies than older ones. Teachers with ID less than the average (less than 6.915.638) ask more questions of mathematical nature, walk around the students more and bring the students to the blackboard more than the ones with ID number greater than the mean. It must be recalled that ID number is correlated with age (higher ID numbers usually mean younger people). Of course, it must be pointed out that 82% of the teachers with smaller ID number were middle school teachers, and only 68% of the ones with higher ID numbers taught in middle schools. This might then be tied to teaching level rather than age.

Now, let us look at what is probably the most important, but also most disappointing, information that can be extracted from the data in this study:

The students ask very few question of mathematical nature in class. Indeed there is, as a group (not each student individually), at most one math question per class.

Other sources of bad news:

- There was very scarce use of textbooks in class.
- There was no evidence of ICT usage in class. No computers or educational software at all in the observed segment of any class.
- There were no mathematical proofs, or evidence of deductive reasoning in the observed videos.
- There were no mathematical metaphors to be observed in class.

One encouraging one, to finish:

- Teachers did not make mathematical mistakes. There were actually a few conceptual errors that were corrected right away, but that was.

DISCUSSION AND CONCLUSIONS

The conclusions of this study are of two types. One has to do with the validation of some methodological aspects, and the other with findings coming from the observed variable.

About methodology, we could conclude that our coding was stable and reliable. Since different coders watching the same slices of the same videos produced statistically similar patterns, then the information we are getting from the variables does not depend on the observer and our results can in a way be called objective, or at least repeatable.

It is also a conclusion belonging to the methodology realm that the short segments we chose to be viewed in the videos, actually provide relevant information. It has already been pointed out that the information provided by variables measuring total time duration of events does not degrade considerably when we consider 2 minutes slices instead of the longer 4 minute slices. And the same is true with counting variables, provided that the episodes being counted are short. For variables counting appearances of longer episodes, the quality of the information deteriorates when halving the temporal size of the slice. It easy to make up an explanation for the phenomenon, longer events have higher probability not to happen entirely during shorter observation period.

In terms of the information provided by the measured variables, as it has been mentioned before, there are some conclusions that might seem rather obvious, like the one saying that there is no independent student work at the beginning or ending of the class. Obviously we too share the view that those are not pioneering findings, but their inclusion here has the virtue of documenting that this is happening, and also quantifying it.

In many other cases, apart from the reasons given above for including some "common sense" conclusions, it is also true that those facts do not have to be that obvious. For instance, the fact that high school teachers spend more time at the blackboard than their middle school colleagues, might be something we would expect, but not necessarily something we desire. It might be beneficial for high school students to solve problems by themselves in class (a considerable part of the lecture time, at least) and have their teachers walking about the classroom checking on the student's work and helping them with it, instead of spending that much time lecturing from the board. Also, even though we expected to see that emerge from the data, there is no real reason why the use of concrete material in high school lectures should be scarce. Leaving aside the obvious example of the geometry strand, the teaching of probability, strand which is present in high school curriculum, can greatly benefit from playing al sort of chance games which usually involve concrete materials (under the form of cards or dice, for instance). It could also be arguable that it should be natural to spend less time writing mathematics while teaching in middle school than in high school.

In any case, justifiable or not, it is clear from the data that there are didactic differences between middle school and high school.

Somehow puzzling is the data showing that teachers from Región Metropolitana (which, as we said, concentrates a large amount of Chile's population, more than one third of it, and by far larger than any other region of the country) seem to have a more conservative teaching style. They seem not to give hand-outs to the students, then approach the kids a lot less than their colleagues in other regions to check on their work. However, they ask more questions of mathematical nature, and in turn, get more answers from the students.

Among the data describing slices 2, 3 and 4, there are also some obvious things, for which the major importance of measuring them is to have quantified evidence that these things happen, but here there are also some others where it could be argued that even though they were to be expected, they are not necessarily desirable.

For instance, there is no reason why the supervision of student work should mainly be only when they are using concrete material, as it was found.

Also, there is no need for teachers not to be able to supervise student's work and at the same address the whole class, or even ask math questions to the students. The use of technology, and/or games can make it easy for the teacher to do both things at the same time.

There is also no obvious reason why the concrete material should be used more frequently, as found, while working in applied problems. Such materials, and later the mental representation of them, can often help greatly the solving of problems of more mathematical or theoretical nature (National Mathematics Advisory Panel, 2008, Chapter 4: Report of the Task Group on Learning Processes), especially if the right metaphors are used to approach them.

Clearly, the most alarming finding of this study is that relevant student participation, seem to be neglectible. An average of a little less of a question of mathematical nature coming from the students per class is obviously much too little. The origin of this must be investigated in some other way. Being so few, our data does not allow us to correlate the number of student math questions to any other variable that we might think is a cause, such as time spent by the teacher lecturing from the front of the classroom.

To explain the scarce evidence of textbook usage, international evidence might have the answer. According to a NSF study in the United States (Banilower, Boyd, Pasley and Weiss, 2006), the probability of a teacher using a textbook is extremely sensitive to the hours spent on professional development training for the use of the materials. And even though Chilean state makes the great effort of giving every child from public schools a textbook, no teacher training for the book usage is offered. Also, the state policy of constantly opening competitive biddings for these textbooks, makes it very likely that a textbook for a given level change every 2 or 3 years, and it is hard for teachers to be constantly readjusting their lectures to ever-changing textbooks.

For the absence of ICT material, and of proofs in the lectures, some people argue that the videos we used were for teacher assessment, and the teachers probably wanted to record a class that would be evaluated as an excellent one, and they probably did not feel confident enough either with the use of technology, or discussing proofs. If that is true, and in "normal" classes there might be a bit more proofs and use of technology, the argument also would say that teachers are far from feeling confident with these two things, and so there is still a big problem with them. Similar arguments apply to lack of textbook usage and this sort of argumentation.

This line of argument tells us that we should not feel too happy about the absence of conceptual flaws in the classes that were reviewed. They were carefully prepared, and it might be expected that each teacher taught in a subject with which he or she felt rather confident.

Finally, some other international studies (Richland, Holyoak and Stigler, 2004) have found some limited use of them. Perhaps our criteria for defining and finding metaphors have to be reviewed and a search for them have to be performed again.

REFERENCES

Ambady, Nalini & Rosenthal, Robert (1993) Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*. Mar Vol 64(3) 431-441

Banilower, Eric R.; Boyd, Sally E.; Pasley, Joan D. & Weiss, Iris R. (2006) *Lessons from a Decade of Mathematics and Science Reform. A Capstone Report for the Local Systemic Change through Teacher Enhancement Initiative*. Report for the National Science Foundation.

Coan & Gottman (2007) The Specific Affect Coding System (SPAFF). In Coan & Allen (Eds) *Handbook of Emotion Elicitation and Asssessment*. Oxford University Press.

Díaz, L. (2006) Profundizando en los Entendimientos Estudiantiles de Variación. In *Investigaciones sobre Enseñanza y Aprendizaje de las Matemáticas: un Reporte Iberoamericano* . Comité Latinoamericano de Matemática Educativa, A.C.

Gladwell, Malcolm. (2005). *Blink*. Back Bay Books.

Gottman, John. (2000) *The Seven Principles for Making Marriage Work: A Practical Guide from the Country's Foremost Relationship*. Three Rivers Press. Trade Paperback, 288 pages. ISBN: 978-0-609-80579-4 (0-609-80579-7)

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMMS 1999 video study*. U.S. Department of Education, National Centre for Educational Statistics.

LessonLab, Inc. (2003). *TIMSS-R Video Math Coding Manual*.

Lakoff G. & Nuñez F. (2000) *Where Mathematics comes from?*, Basic Books,New York.

Manzi, Jorge (2007) Evaluación Docente : Antecedentes, resultados y proyecciones, en Seminario *Evaluación Docente en Chile: Fundamentos, Experiencias y Resultados*. Seminar organizad by Mide UC on Octuber 2nd.

National Mathematics Advisory Panel, 2008. http://www.ed.gov/about/bdscomm/list/mathpanel/index.html

Redish, E. (2003) A Theoretical Framework for Physics Education Research: Modeling Student Thinking. *The Proceedings of the Enrico Fermi Summer School in Physics*, Course CLVI . Italian Physical Society

Richland, L; Holyoak, K & Stigler, J. (2004) Analogy Use in Eighth-Grade

Mathematics Classrooms *Cognition and Instruction*, 22(1), 37–60

Stigler & Hiebert (2004) Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study.

Shimizu, Yoshinori (2003). Capturing the Structure of Japanese Mathematics Lessons as Embedded in the Teaching Unit . Symposium "Mathematics Lessons in Germany, Japan, the USA and Australia: Structure in Diversity and Diversity in Structure" at the *Annual Meeting of the American Educational Research Association*, Chicago, April 21-25.

TIMSS (1999) *Video Study of Eighth-Grade, Mathematics Teaching*. National Center for Statistics. Department of Education. Institute of Educational Sciences.

Ministerio de Educación de Chile (2002). Estudio sobre uso de textos escolares en el primer ciclo básico. Subcomponente Seguimiento, Unidad de Curriculum y Evaluación, http://www.textosescolares.cl/documentos/Resumen_Ejecutivo_Estudio_Uso_de_Textos_1.pdf

Universidad de Chile (2006) Encuesta sobre Uso y Calidad de Textos Escolares. Centro MICRO-DATOS, Departamento de Economía, http://www.textosescolares.cl/documentos/admdocs/docs/200801311222580.Informe_Encuesta_Textos_2006.pdf

Tuminaro, J. & Redish, E. (2007) Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics - Physics Education Research* 3, 020101