Third Misconceptions Seminar Proceedings (1993)

Paper Title: **COMBINING QUALITATIVE AND QUANTITATIVE RESEARCH METHODS IN TUTORING EXPERIMENTS**

Author: Zietsman, Aletta & Clement, John

Abstract: Many researchers have suggested that statistical/hypotheses testing type studies are not satisfactory when a study is aimed at the fine-grained analysis of learning successes and difficulties of students. There is often a gap between what the reseacher puts forward as a hypothesis about learning processes and the hypothesis that may emerge from research data. For one, there are the obvious differences between teacher and student worlds - even for very sensitive and caring teachers: students simply see situations so differently from us.

Keywords: Research Methodology,Concept Formation,,Qualitative Research,Learning Processes,Misconceptions,,,
General School Subject: Physical Science
Specific School Subject: Physics
Students: Junior High

Macintosh File Name: Zietsman - Tutoring Experiments
Release Date: 5-27-1994 H, 11-10-1994 I

Publisher: Misconceptions Trust
Publisher Location: Ithaca, NY
Volume Name: The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics
Publication Year: 1993
Conference Date: August 1-4, 1993
Contact Information (correct as of 12-23-2010):
Web: www.mlrg.org
Email: info@mlrg.org

A Correct Reference Format: Author, Paper Title in The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Misconceptions Trust: Ithaca, NY (1993).

Note Bene: This paper is part of a collection that pioneered the electronic distribution of conference proceedings. Academic livelihood depends upon each person extending integrity beyond self-interest. If you pass this paper on to a colleague, please make sure you pass it on intact. A great deal of effort has been invested in bringing you this proceedings, on the part of the many authors and conference organizers. The original publication of this proceedings was supported by a grant from the National Science Foundation, and the

transformation of this collection into a modern format was supported by the Novak-Golton Fund, which is administered by the Department of Education at Cornell University. If you have found this collection to be of value in your work, consider supporting our ability to support you by purchasing a subscription to the collection or joining the Meaningful Learning Research Group.

----- -----

COMBINING QUALITATITIVE AND

QUANTITATIVE RESEARCH METHODS

IN TUTORING EXPERIMENTS

1. Gaps in Current Methodolgies

2. Goals Analysis for Learning Studies
* combine qualitative clinical interviewing data and quantitative summative data
* allows one to focus on and analyze key learning episodes
* additionally: combine summative data that show large significant gains on transfer of learning problems (despite small sample sizes)

3. Structured, Formative Research and Evaluation Cycles
* Major Stages:
1. Diagnostic testing and interviewing
2. Exploratory teaching
3. Standardized teaching cycles

* Detailed Steps
1. Diagnostic analyses (culminates in fixed dagnostic test)
2. Exploratory tutoring
3. Lesson design
4. Standardized teaching (cycle 1)
5. Summative evaluation
6. Qualitative analysis
7. Developing general models of learning and teaching strategies
8. Re-desing lesson
9. Standardized teaching (cycle 2)
10. Experimental comparisons
3. Advantages
a. Qual. analysis allows one to evaluate potential of

particular teaching strategies without controlling all other variables. This allows one to change many teaching strategies on each cycle, speeding up the development process. The evidence is corroborated by experimental studies when feasible, but gives good guidance on its own.

1>   Example of support for strategy: Analogy fails and why it fails, extreme case succeeds and why it succeeds
2>   Example of critique of strategy- done above?]

3>   References to other findings and papers

6.   General Applicability
Use Exploration first
Then Cycles
Then Comparison

COMBINING QUALITATIVE AND QUANTITATIVE RESEARCH METHODS IN TUTORING EXPERIMENTS

Aletta Zietsman
Physics Department
University of the Witwatersrand, Johannesburg
South Africa

and

John Clement
Scientific Reasoning Research Institute
University of Massachusetts, Amherst

GAPS IN CURRENT RESEARCH METHODS

Many researchers have suggested that statistical/hypotheses testing type studies are not satisfactory when a study is aimed at the fine-grained analysis of learning successes and difficulties of students. There is often a gap between what the reseacher puts forward as a hypothesis about learning processes and the hypothesis that may emerge from research data. For one, there are the obvious differences between teacher and student worlds - even for very sensitive and caring teachers: students simply see situations so differently from us.

What seems to be needed is a kind of action research, but on a much smaller scale than what is traditionally defined as action research. In this paper we are proposing a research method that makes possible the detailed analysis of learning and reasoning processes that one can obtain from individual interviews, as well as formative evaluation cycles of instructional interventions typically obtained from action research studies carried out in classrooms. We will describe the method in general and then present some examples from a specific study.

EXPLORATORY TUTORING EXPERIMENTS

Other learning studies have initiated the development of an approach that combines the use of qualitative clinical interviewing data and quantitative summative data (e.g. Brown, 1987). These methodologies provide what we hope will be a widely applicable model for graduate student theses and smaller studies that can identify learning processes and general teaching strategies in science. The methodologies have allowed us to focus on and analyse key learning episodes where protocols document the effect of a specific tutoring strategy (such as the use of anchoring examples) on the student's conceptions, in combination with summative data showing large significant gains on transfer problems - this despite sample sizes of six to eight in each group.

An example of one of the methodologies we are refining was used in a learning study of seventh graders learning about levers. Here the narrower goal was to criticize and improve a lesson design, while the broader goal was to develop a grounded theory of the critical learning processes and general teaching strategies involved.

The methodology included four major stages: diagnostic interviewing and testing, exploratory tutoring, standardized tutoring cycles and quantitative, summative evaluation phases.

1. The diagnostic interviewing and testing phase that culminates in a fixed pre and post test.
2. Within the exploratory tutoring, instruction is not standardized and several tutoring strategies are explored in individual case studies where creative adaptations to student responses are made by the researcher/tutor. Attempts to improve the overall lesson design can be made after each

interview, until one converges on a preferred instructional sequence.

3. Within the standardized tutoring there is a pretest given to six to eight students, a fixed instructional sequence given to each student, and a post test. One then uses the test to do a summative evaluation of whether learning occurred and a detailed, qualitative, formative evaluation of why learning processes did or did not occur, and criticize and modify the lesson design. Then a new standardized tutoring cycle is repeated with another matched group of six to eight students in order to evaluate the effect of the modifications made.

4. Quantitative comparisons of gains in the two groups are made on performance, clinical evidence for alternative conceptions and target conceptions, scores on transfer questions and confidence scores.

We will use the phrase Structured, Formative Research and Evaluation Cycles for this design. As will be described later, the lever study is yielding both new qualitative descriptions of learning processes and significant quantitative differences between participating groups.

STRUCTURED, FORMATIVE RESEARCH AND EVALUATION CYCLES

Detailed steps within the four major phases outlined above, are:

1. Diagnostic analysis. Group testing and interviews are used to map misconceptions and anchoring conceptions in the area being studied.

This cycle is open to changes but culminates in a fixed diagnostic test.

2. Exploratory tutoring. Open-ended, exploratory tutoring sessions are conducted in order to develop a teaching approach. Here instruction is not standardized and several

teaching strategies are explored in individual case studies where creative adaptations to student responses can be made by the tutor. Attempts to improve the overall lesson design can be made after each interview, until one has enough data to design a promising instructional sequence.

The cycle remains open to any on-line changes by the researcher.

3. Lesson design. A lesson sequence is designed incorporating the most promising teaching strategies developed in the exploratory tutoring.

4. Standardized tutoring, cycle 1. This consists of a pretest given to six to eight students, a fixed instructional sequence given to each student in the same way (or with prescribed branching), and a post test. Instructional situations remain fixed for a particular cycle.

5. Summative evaluation. The extent to which learning occurred is judged using pre-post test gains in performance, occurrence of alternative conceptions and target conceptions, and confidence scores. The procedures are fixed for the particular cycle.

6. Qualitative analysis. One then does a detailed qualitative analysis of why learning processes did or did not occur. Two outcomes are:

(a) A formative evaluation: specific criticisms of, and recommendations for, changes in the lesson. The cycle is open to input all the time.

(b) Process models: key episodes from all of the protocols are classified and analyzed qualitatively in order to develop general models of learning and teaching strategies, as well as barriers to understanding.

The process is open to changes.

7. Re-design lesson. The lesson design is modified or rewritten on the basis of the formative evaluation. The design is open but within the constraints suggested by cycle six to eight.

8. Standardized tutoring, cycle 2. A new cycle (steps 4 -

6) is repeated with another matched group of six to eight students in order to evaluate the effect of the modifications made.

9. Experimental comparisons. Quantitative comparisons of the data from the two cycles are made on pre-post test gains, and qualitative comparisons are made on learning processes.

Some specific research questions within a structured, formative research and evaluation type study may be:

1. What are the important preconceptions (e.g. anchoring intuitions) in the area?
2. What are some promising ideas for teaching strategies in this area?
3. Did some learning of new conceptual understandings occur?
4. What elements of the lesson were important to learning?
5. What elements of the lesson were detrimental to learning? What are the important barriers to learning?
6. What modifications could be made to the lesson?
7. (After a second version has been tested.) Did the modifications made in the lesson allow more learning to occur? (Also repeat questions 3, 4, 5, and 6.)
8. What hypotheses concerning general models of learning processes and teaching strategies can be proposed and supported on the basis of the transcripts?

Addressing the first question depends on interviewing and diagnostic testing, while addressing question 2 depends on exploratory tutoring. Addressing questions 4 and 5 depend heavily on descriptive clinical analysis, while questions 3 and 7 depend on quantitative measures from pre and post tests given in the same way to all students in the study. Questions 6 and 8 depend on all the questions preceding them and therefore on both quantitative and qualitative analysis. Question 8 concerns the new preconceptions, anchoring conceptions, reasoning processes, learning processes, and barriers to learning identified in the protocols. When these occur in different contexts, they are candidates for components of general teaching strategies.

THE LEVERS STUDY

A stratified sample of six to eight students was interviewed in each cycle of the study of seventh graders learning about levers. Participants were selected to represent three levels of ability and were half males and half females. Post tests included performance measures, transfer problems, and confidence scales. Since there were grounds to suspect a possible practice effect from the test, a control group was interviewed on the pre and post test.

Quantitative Analysis

The most important value of the quantitative analysis was in the overview it provided of the effectiveness of the lesson. This was obtained by calculating participants' scores on pre and post test interviews. Scores were obtained from the students' answers and their confidence in the answers, and were computed by: one, assigning positive and negative values to corrrect and wrong answers respectively; two, assigning a number (1 to 4) to the confidence level (rated from a "guess" to "sure I'm right" on a four point scale); and three, multiplying the confidence level number with the appropriate symbol to indicate a correct (or not) answer. Thus, a student who guessed a wrong answer would score -1 on a question, while a student who was sure that he was right about a wrong answer, would score -4. In Table 1 the experimental and control group scores on the pre- post test questions are given.

The Mann-Whitney test was used to test hypotheses (1) that the control and experimental group 1 were identical with respect to their performances on the pre and post tests; and (2) that the experimental group 1 and experimental group 2 were identical with respect to their performances on the pre and post tests. In (1) the null-hypothesis was accepted: for $p < 0.05$, $U_{calculated} = 16$ has a probability of occurence under H_O of $p = 0.41$.

Using a criterion of $p < 0.050$, the difference between experimetal groups 1 and 2 is significant ($U_{calculated} = 7$, $p < 0.047$).

Table 1

Experimental and Control Groups: Pre and Post Test Scores

Experimental 1		Experimental 2		Control (for 1)	
Student	Summed Score	Student	Summed Score	Student	Summed Score
E1	-4	EE1	+12	C1	+4
E2	+14	EE2	+21	C2	0
E3	-10	EE3	+16	C3	0
E4	-6	EE4	+1	C4	+14
E5	+3	EE5	+18	C5	-6
E6	+13	EE6	+9	C6	-7
Group Total	+10	Group Total	+77	Group Total	+5

Research Cycles and Qualitative Analysis

All through the study, the reasoning power of the students and their conceptions surprised us greatly -despite years of experience in dealing with them in other areas. At each level of the research process there were unexpected findings and results, and even in the final cycle (standardized teaching, cycle 2) new hypotheses about the students' learning emerged. This illustrates the need for research and evaluation cycles in areas of teaching and learning where significant understanding is sought. We have numerous examples of such unexpected findings and the resulting changes in the various stages, and present here some striking episodes from the diagnostic tutoring and exploratory tutoring cycle, and from the first and second standardized tutoring cycles.

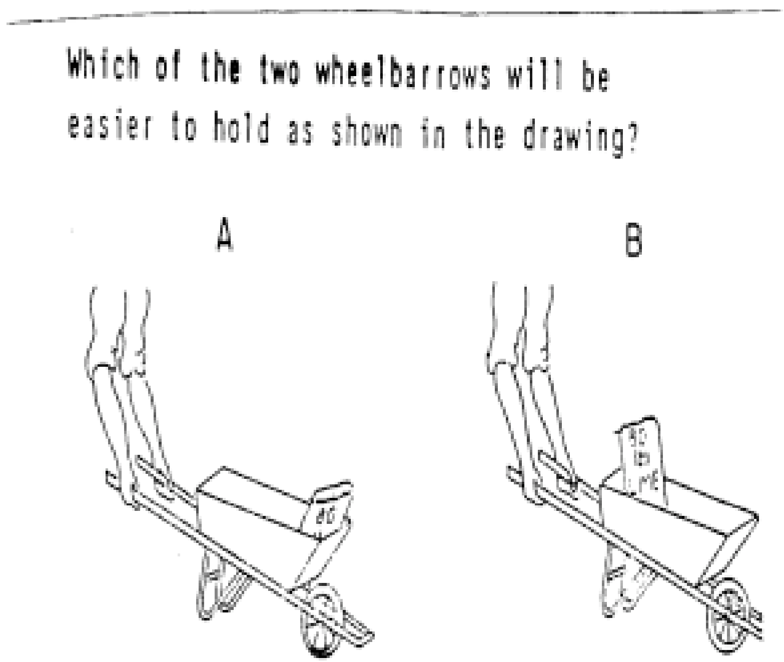
Diagnostic Interviews and Exploratory Tutoring

The most alarming finding from the first cycle was that not one person involved in the research process till then (researcher and teachers who were consulted) had any inkling of lever features or actions that were important for the children. Two levels where such unexpected findings emerged from the qualitative analysis are: (1) examples of lever situations that triggered intuitive conceptions, and (2) students' use of nonformal reasoning strategies that became the focal point of the lessons in the final tutoring cycle.

A striking example of the difference between our expectations and those of the children of important or useful physical situations and those lever situations, emerged from the combination of diagnostic interview analyses data and exploratory tutoring data. We had assumed that a seesaw (teeter-totter) would be the lever most children would understand intuitively. One could, for example, use their intuitions about the seesaw to bridge to the use of a spoon to open lids of tins (in the exploratory tutoring cycle). However, it was clear from the protocol analysis that these situations were not intuitively understood by the children: they gave poor explanations about physical situations involving the seesaw and the spoon as levers. At the same time, the data showed that an example that generated the most discussion and the emergence of the use of extreme case reasoning was around a diagnostic question depicting two wheelbarrows (Figure 1).

Figure 1
<u>Comparing Wheelbarrows</u>

Figure 1
<u>Comparing Wheelbarrows</u>



The wheelbarrows example was not used in the standardized teaching cycles as a learning situation. What is important is that the children's reasoning about the wheelbarrows suggested that we concentrate on extreme case reasoning involving the fulcrum, rather than on semi-quantitative reasoning strategies involving balance beams or seesaws in the development of teaching strategies. Examples of children's reasoning about the wheelbarrows are:

Ken 051 In A you don't have to lift up as much...the wheel will push up on the stuff inside...

Dani 034 The wheel holds it [the load] up in A, but in B **you** have to hold it up.

Phil 069 A is easier to lift because in A the load is nearest to the wheel, it is right on the axle.

In addition to such important findings, a great bonus of the exploratory tutoring cycle was that the on-going qualitative analysis and immediate feedback allowed one to evaluate teaching strategies "on line", to create new learning situations, thus speeding up the development process.

Standardized Tutoring Cycles

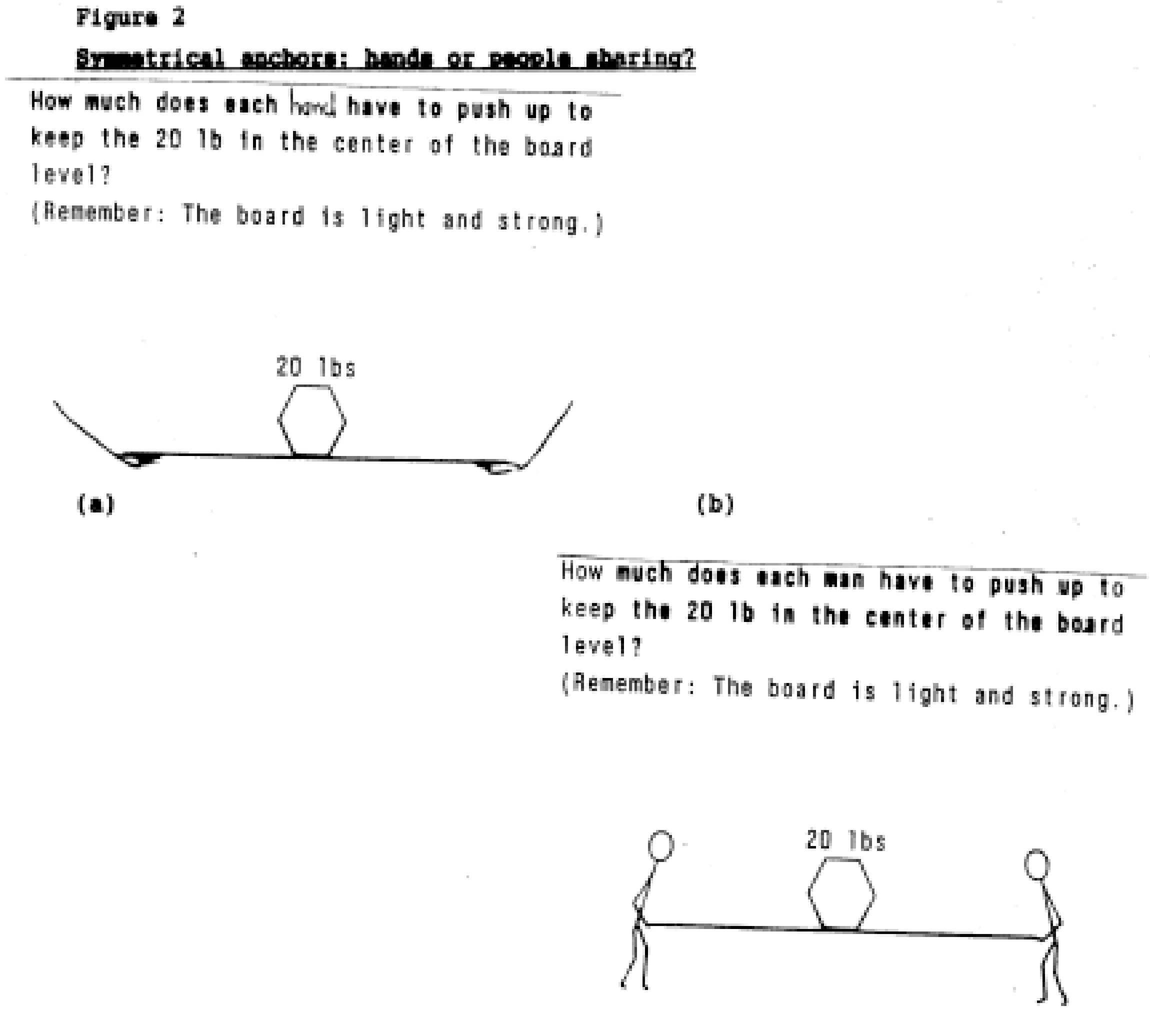
Within each of the standardized tutoring cycles the instructional sequence was fixed. Again, the first cycle of six participants produced some unexpected findings. Even after the considerable diagnostic and tutoring efforts, although one of the three sections of the lesson seemed to work fairly well, another section produced no gain differences, and a third sequence actually produced negative gains. However, the qualitative analysis of the first cycle was used to redesign the lessons with very positive results in the second cycle (see Table 1).

A good example of the power of the students and their conceptions to surprise us greatly, is the findings about the anchoring situation for the first tutoring cycle and how these impacted on the second cycle's instructional design. We have defined anchoring situations as student intuitions that are compatible with accepted physical theory (Zietsman & Clement, 1993). The idea is that one grounds instruction in such elemental, primitive physical intuitions - thus: building instruction on what is intuitive to the student. One of the strongest group anchors suggested by the diagnostic cycle data, was the symmetrical anchoring example in Figure 2(a): a person holding a load level on a light, strong board. A typical example of students' reasoning about this situation is:

Seth 132 The person is holding it [the load] up with his hands, each hand is supporting 10 lbs, because it's [the load] in the middle and therefore it would be the same on both hands.

Figure 2

Symmetrical anchors: hands or people sharing?



(a) (b)

In the first tutoring cycle interviews a student suggested two people sharing a load instead of two hands (Figure 2(b)):

Emi 029 I don't know. It just seems easier to see if two people are holding a board level when they have to be sharing the same weight.

This example may seem trivial, yet when incorporated in instructional sequence 2 the students' confidence in their reasoning about the second example was higher and may have had an impact on their evidently better understanding. What is important is that the changes were generated by the students: we simply incorporated ideas that were self-evaluated by the students.

CONCLUSION

We find the research design interesting for the following reasons: First, the small number of participants allows one to make a detailed analysis of the students' learning and reasoning processes, as well as of the teaching strategies. Why did things work and why did they not? In addition, one may also obtain significant differences in the performance of groups on pre and post test scores. The latter is an added bonus - not the main goal of the research. Much more significant is that we gain insights into what "works" in instruction for different students. Second, it is also interesting that the method combines basic research into students' learning and reasoning as well as evaluation. In the basic research we gained insights on general learning processes and many interesting new hypotheses about learning emerged from the qualitative analysis. In the evaluation cycles we were able to find specific results about teaching strategies.

Some of the most important results from the lever study are: One, several new and widespread insufficient conceptions about levers have been discovered that are less accurate or general than conventional conceptions. With this knowledge one can classify problem solutions as being based on a particular conventional conception or an insufficient conception. After instruction, more of these involve conventional conceptions, and more involve a model of hidden mechanisms (pushes or forces) explaining how the lever works as opposed to merely a rule about the lever's behavior. Furthermore, many students were able to transfer the conceptions they learned to fairly difficult transfer problems involving compound levers. Thus a methodology is being developed for supporting a claim for increased understanding from several different sources.

Two, important counter-examples have been found to a previous hypothesis that symmetrical anchoring cases in physics are brittle, i.e. that cases tapping a students' intuitions about symmetrical situations are not plausibly extendable by analogy to non-symmetrical cases (Clement, Brown & Zietsman, 1989). This is an important modification to our theory of anchoring intuitions.

Three, in the lessons the use of extreme cases and the use of models of hidden mechanisms have been much more powerful teaching strategies than we had expected for this younger age group. Extreme cases may seem like a trivial trick used by experts as non-essential short cuts, but in this case they were central to a good lesson design (Zietsman & Clement, in press). We are surprised by these results because the extreme case reasoning strategies often seem more powerful than the analogical reasoning techniques we were counting on originally. Surprises like these testify to the power of learning studies to change and improve teaching strategies.

REFERENCES

Brown, D. (1987). Use of analogical reasoning and bridging strategies in remediating misconceptions on mechanics. Unpublished doctoral dissertation, university of Massachusetts, Amherst MA.

Clement, J., Brown, D.E. & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding anchoring conceptions for grounding instruction on students' intuitions. International Journal of Science Education,

Zietsman, A. & Clement, J. (1993). Explanatory models for conceptual understanding: A case study in model evolution. Paper presented at the American Educational Research Association's annual conference, April, Atlanta GA.

Zietsman, A & Clement, J. (in press). Extreme cases as mechanisms of learning. Instructional Science.