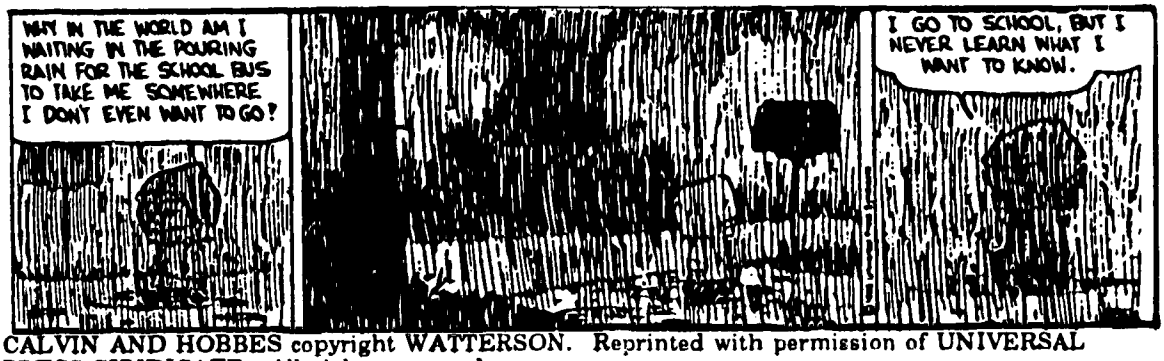
Apparently Calvin couldn't agree more about the interest level of his instruction. Here he laments:

Figure 11

Cartoon: Uninteresting Instruction



Normally an inquisitive little boy and enthusiastic about life in general, he is turned off by school. He speaks for many young children who look so eagerly to school but never learn what they want to know.

Unfortunately, to make learning "fun" for Calvin, some teachers expend untold amounts of energy, dressing up the classroom, and having students engage in "fun "activities. Learning is indoctrinated with "fun" mascarades and façade. However, for Calvin, real learning is interesting in its own right. When Calvin is investigating insects, he is completely absorbed in the wonder of them just as they are in their own "living color." Caterpillars, for example, do not need grinning faces to make them appealing or "fun" for Calvin, and yet how many classrooms do we find in the spring with green caterpillars with silly big grins plastered over the classroom walls--clone after tedious clone, carefully cut out from duplicated worksheets? Early childhood researcher, L. Katz (personal communication, April, 1987) once made that very same observation. She noticed children in an "art" class diligently following step-by-step direction from the teacher in order to construct the prescribed green caterpillars as they did each spring, out of the usual egg cartons, which they finally painted with the standard green paint. At the time, Katz made the observation to the teacher that caterpillars do come in many colors, many sizes, with many different kinds of exterior coverings. Yet many students leave school believing that caterpillars are green. Shallow instruction like this cannot be helpful for deep assessment of learning. We need to have both rich instruction for real learning so that we can have real, or authentic assessment.

When I think of the most "fun" I have had in learning, it was not doctored up with frippery. It was real. Learning to play the piano was real in its own right, as was learning how to make sense of, and work the computer. How the hours flew by as I was completely absorbed in learning! I did not need special programs or activities to coerce me into believing it was "fun"--and for the most part, nor do children. Teachers have been taught to believe that if it is not fun for students it is not good. Learning in itself is enjoyable, intriguing,

interesting, and yes, sometimes it does take tenacity and steady hard work to succeed. Often it is fun--but it doesn't have to be for the learner to feel a deep sense of satisfaction and accomplishment.

The business world is aware of the powerful message that "fun" sends to the uninformed public. Many commercial companies in computer technology, for example, have specifically focused their wares on making learning "fun" for children, particularly in education. Does that mean that learning, then, in its own right, is not fun? F. Smith (1986) takes exception with the "fun" focus as well. He cites the following computer programs that insult students' intelligence: His examples include, a Scott Foresman software program called "Reading Fun"; Miliken Publishing Company's entire program called "Edu-fun" producing such programs as "Mathfun" and "Wordfun," and Barnell Loft of Balwin, with a series of vocabulary exercises called "Fun With New Words," and "Reading for Fun." And it seems that "fun" is gaining in momentum. The 1993 Mid-Winter issue of MacSelect catalogue features more "fun" programs, not just for young children but for older ones as well. For example, "Fun Physics" from Knowledge Software, to programs that promise "Interactive learning fun for ages three and up" from T-Maker, and from MECC: "Math basics have never been more fun for kids!: A fun grid format makes learning fun for children eight and over." Parents are now targeted as well, for example, Davidson & Assoc., "New Math Blaster Plus:--a fun way for your child to learn math!" Finally, grandmothers aren't spared either: From Broderbund, we have "Just Grandma and Me CD--animation and music make learning fun!" (pp. 60-61)

F. Smith (1986) poses an interesting question about fun and learning:

> What's wrong with all this concern with fun? Does learning have to be miserable? Of course not. But the underlying implication of "learning should be fun" is that learning will be a painful and tedious activity unless it is primped up as entertainment. Learning is never aversive--usually we are not aware of it at all. It is failure to learn that is frustrating and boring, and so is having to attend to nonsensical activities. Children do not learn things because they are fun, but because they enable them to accomplish ends, and they learn in the process of accomplishing those ends. (p. 82)

What Meaningful Teaching Is

I have understood that what students believe is meaningful teaching and what they believe it is not. Students prefer to be engaged in meaningful, interesting topics; topics and activities that are relevant and have real-world application. For example, I gave my class of university students an assignment which they felt was one of the most relevant ones that they had done at university to date. It was challenging in that they had to find a way to use an abstract concept such as statistics in a real-world application. The following is the assignment as it was given to the students.

APPLICATION ASSIGNMENT: 10 MARKS

The purpose of the Application Assignment is to have you experience student engagement in learning in the context of descriptive statistics. Too often material such as this is learned by rote for a test instead of learning to understand it for its real-world application. Class time will be available for you to work on this project and I will be available for individual/group help.

Specifically, you are to demonstrate what you know and can do with central tendency and variance. You may choose to demonstrate this in a real-world situation of your choice.

Data Gathering: Data may be obtained from unlimited sources such as the sports world, medical fields, businesses, research station, entertainment world, agriculture, etc.

Procedure: You may wish to work collaboratively with others, in pairs, or individually, whichever is the best way for you to demonstrate your knowledge and which will be the most useful, meaningful, and interesting for you. There really is no limit to the number of ways to proceed.

Presentation/Representation of Knowledge: Again there are numerous ways to do this, for example by use of charts, video, photo album, slides, "filmstrip," dramatization/role play--as an alderman/woman presenting findings, board chairperson, coach, teacher, principal, superintendent, doctor, dentist, caretaker, etc.

GRADING CRITERIA:

5 MARKS: You will accurately apply mean, median, mode, range, frequency, and standard deviation to real-world data.

5 MARKS: You will present this knowledge clearly in an interesting, creative, and meaningful form that would be understood by the general public. If you are working with others, you must indicate how your group wishes to be graded to ensure fairness to all.

The students could choose how they wished to apply this concept, and the range of the different ways they did this was exceptional, from traffic statistics to the race track statistics. Students found they spent more time and energy on this assignment than they had intended, and yet they enjoyed it. Some students who previously "learned" standard deviation for their math test in other courses said that this was the first time that they actually understood it. Now when they read an article that referred to standard deviation, they could make sense of it because they had internalized it (Doll, 1989). As F. Smith (1986) says, "When learning is successful, it is totally inconspicuous" (p. 27).

This was an example of an engaging activity, or as Wiggins (1993b) calls it, an "ill-structured task"--the students' representation and presentation of their understanding were completely up to them. The assessment was authentic because the instructional activity was authentic; "the methods and criteria are quite clear to all students in the course, there are no pat routines, procedures, or recipes for solving the problem" (Wiggins, 1993b, p. 205). Wiggins judges the authenticity of tasks by the following factors which are briefly outlined here: Engaging and worthy problems; faithful representation of the contexts in a field of study (no secrecy); nonroutine and multistage tasks--real problems; tasks that require the student to produce a quality product and/or performance; transparent or demystified scoring criteria, interactions between assessor and assessee; response-contingent challenges which the effect of both process and product/performance determines the quality of the result; trained assessor judgment, in reference to clear and appropriate scoring

criteria (possibility of altering a result because of oversight); and the search for patterns of response in diverse settings (pp. 206-207).

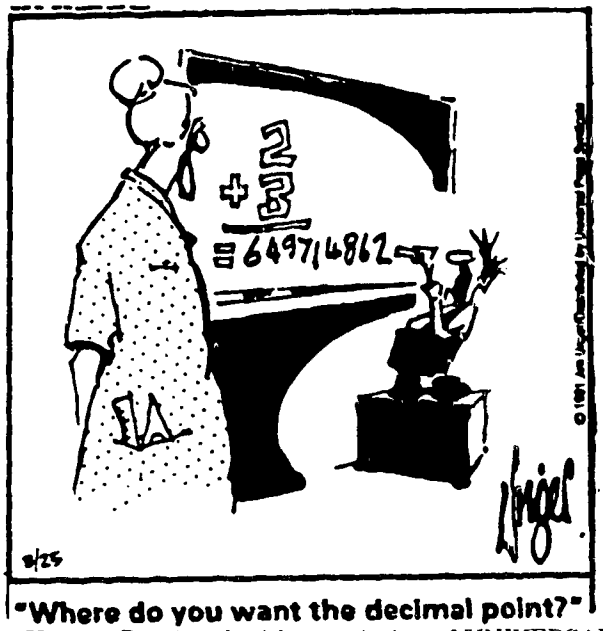
It is impossible to assess for understanding if we have not first taught for, or engaged students in, understanding. Traditionally, and still today, sadly, much emphasis has been on repeating, or miming, newly presented information (Jackson, 1986) in reports or on quizzes and tests. On the other hand, the kind of assessment and instruction that the students valued in this investigation, is closely related to constructivism, as Grennon Brooks and Brooks (1993) describe:

> Constructivist teaching practices, on the other hand, help the learner to internalize and reshape, or transform, new information. Transformation occurs though the creation of new understandings (Jackson 1986; Gardner 1991b) that result from the emergence of new cognitive structures. . . . . Deep understanding occurs when the presence of new information prompts the emergence or enhancement of cognitive structures that enable us to rethink our prior ideas. (p. 15)

Figure 12 illustrates what understanding is not. Again while we may chuckle at this cartoon, it really does target the truth, particularly as I have found, in mathematics. Many teachers of mathematics and their students, do not have a good understanding of mathematics. They have learned their mathematics instrumentally and while the algorithms may "work" for them, they have little understanding as to why they work.

Figure 12

Cartoon: Teaching for Understanding



Constructivist approaches as described by Duffy and Bednar (1991) refer to engaging in projects or topics on which students can work collaboratively or individually in pursuing their interests in rich context. Such projects emphasize reflective thinking and model problem-solving by "experts or in apprenticeship roles." Gardner (1991) believes in the project and apprenticeship approaches which he feels can take on a developmental perspective from novice, advanced to journeyman and master levels. He posits the idea that these can be regarded as authentic ways to approach learning and assessment. He presents the following criteria:

> How well is it conceptualized? How well is it presented? How well has it been executed in terms of technical facility, originality, and accuracy? To what extent, and how accurately is the student able to assess the project on these criteria? It is possible to secure reasonable consensus on such evaluation, and the fact that projects can be so described and

evaluated allows them to be considered seriously by the entire community rather than being dismissed as a frill. (p. 217)

Gardner (1991) cautions, however, that projects are not a panacea:

[P]rojects can be assembled hurriedly on the last day and can draw heavily on the child's own previous work or the work of a friend or classmate. Some materials must be learned by drill; others are more readily presented by classroom lecture or by textbook reading rather than by hands-on, museum style-style activities or by participation in an apprenticeship arrangement. (p. 219)

Constructivism: What The Research Says

What is constructivism? Constructivism can be contrasted to traditional teaching where students passively absorb transmitted information from others (Clements & Battista, 1990). It is a Piagetian-based concept in that it holds the view that knowledge is actively created or invented from previously-held knowledge by the child, and not passively received from the environment. In addition, as children create new knowledge, that newly-created learning is a social process in which children grow into the intellectual life of those around them (Bruner, 1986, in Clements and Battista, 1990; Grennon Brooks & Brooks, 1993). Finally, constructivism holds that no one true reality exists, but rather only individual interpretations of the world, and these interpretations are shaped by experience and social interactions. Therefore, when a teacher demands that students use set mathematical methods, for example, the sense-making activity of students is seriously curtailed. "Students tend to mimic the methods by rote so that they can appear to achieve the teacher's goals" (Clements & Battista, 1990, p. 35). Katz and Chard (1989) note that young children in particular like to please the teacher and that teachers should be aware that children may not have understanding and are pretending to understand. Many children also pretend to understand for fear of reprimand or to avoid shame. They mindlessly go about activities but no real learning is taking place--only learning how to keep the teacher under the illusion that they are learning.

Bruner (1960) in Perkins (1992) believes that instruction should have the objective of leading the child to discover for himself rather than being told what to write and think, and then being tested on what they have been told. This kind of instruction and assessment leads to producing bench-bound learners whose motivation for learning is likely to be extrinsic to the task--pleasing the teacher, getting into college, and artificially maintaining self esteem.

Constructivism in the Classroom: Instruction

"The myth is that learning can be guaranteed if instruction is delivered systematically" (F. Smith, 1986, p. ix).

How does constructivism appear, for example, in a Grade One classroom setting? Put simply, it replaces workbooks and worksheets with activities in context as Kamii and Lewis (1990) describe:

"Paper-and-pencil exercises cause social isolation, mechanical repetition, and dependence on the teacher to know if an answer is correct. We therefore replace the textbook, workbook, and worksheets with two kinds of activities: games and situations in daily living" (p. 37).

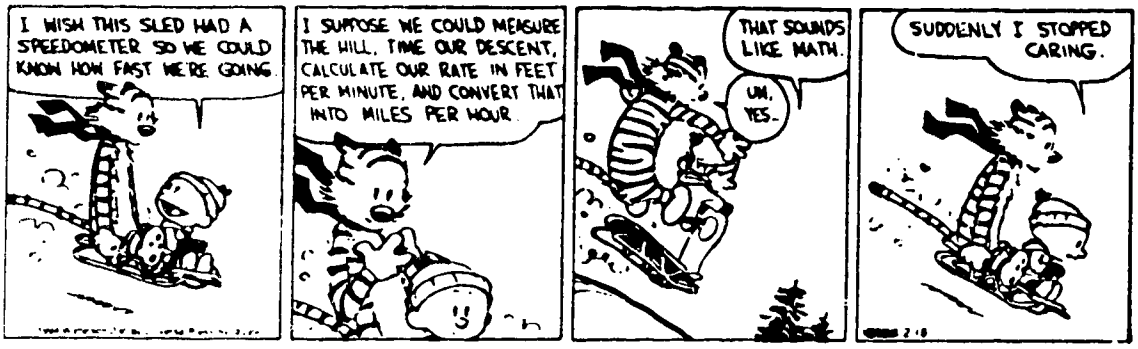
Jonassen (1991) agrees that effective learning is learning in context:

> Many educators and cognitive psychologists are working to develop more constructivistic environments and instructional prescriptions (Duffy & Jonassen, in press). Perhaps the most important of these prescriptions is the provision of instruction in relevant contexts (Jonassen, 1991a). Situated cognition (Brown, Collins, & Duguid, 1988; Resnick, 1987) argues that learning occurs most effectively in context, and that context becomes an important part of the knowledge base associated with that learning. So, rather than decontextualizing learning in isolated school environments, we should create real-world environments that employ the context in which the learning is relevant. . . . Another important strategy is the presentation of multiple perspectives to learners. Cognitive flexible theory is a conceptual model for instruction that facilitates advanced acquisition of knowledge in ill-structured knowledge domains. (p. 11)

Learning in context brings learning to life--it does not kill it as Calvin's learning seems to be in Figure 13. It seems that already Calvin cannot see very much relevance in learning mathematics, a subject that is so relevant in life.

Figure 13

Cartoon: Relevance of Instruction to Life



The Students' View of Constructivism in the Classroom

How does the student view the constructivist approach? Students brought up on the traditional diet of worksheets and trivia tests sometimes have difficulty with this approach. Many teachers have experienced the students' displeasure and often teachers give up and go back to their old ways of teaching, keeping the students happily mired in their comfortable workbooks. Competitive students often find the constructivist approach particularly perplexing and somewhat irksome. Some time ago I had a very competitive Grade 2 student in my class who had already learned how clever he was and who was not at all comfortable working cooperatively, sharing his understandings, or being assessed without letter grades. He wanted his report card to show the usual column of "A"s to show off to his family and to his friends. What a malcontent he was for the first few months in the classroom! The adjustment was quite difficult for him, and although he finally accepted the approach, he really did not feel quite comfortable with it. Older students often react in the same way, particularly when they face more and more "high-stakes" tests. When students are first introduced to the constructivist classroom, there is a certain amount of "buying in." It is a whole paradigm shift for some of them, as there was for my Grade 2 student. Most often we talk about "paradigm shifts" for teachers forgetting that even students have become entrenched in meaningless learning and competition and find it difficult to come to grips with the concept of learning for understanding. For many students who have learned to "work the system," this is a significant paradigm shift for them and not one that they find at all to their liking.

Jonassen (1991) agrees:

> Whatever the challenges of cognitive complexity and task management, a rather different kind of challenge concerns learners' attitudes toward the enterprise. When learners are asked to thrash around for themselves to some extent, there are often characteristic reactions such as, "Why don't you just tell me what you want me to know?" Such learners are not "buying in" to the constructivist agenda of the instruction, a problem that inevitably stands in the way of a fully engaged learning experience. (p. 20)

Gardner (1991) also has found that some students don't want to "waste their time" developing theories and exploring ideas if the teacher already knows that they are "on the wrong track." Teachers sometimes feel pressure from students to give them the "right" answer and to stop wasting their time.

When students learn in a problem-solving hands-on way, they often think and say they aren't doing anything in class. Students and some adults tend to confuse "doing something" with filling in a workbook. The workbook is tangible. Thinking is not. Again this goes back to whether or not the students' buy into this approach, and also recalling F. Smith's (1986) comment that successful learning is inconspicuous. An interesting illustration of devalued learning is provided by Grennon Brooks and Brooks (1993):

> Five 7th grade students were working with the librarian after their social studies class reacted to the U.S. Constitution's "three-fifths rule," which stipulated that five votes by Africa-American males are counted as three votes by white males. One of the five students seemed rather impatient and pre-occupied. The following dialogue ensued:
> Librarian: Ava, you seemed far away. Is anything wrong?
> Ava: Why do we have to spend so much time talking about this?
> Librarian: We're talking about it so that you'll understand it better. I want you to learn about it.
> Ava: We don't have time to learn it. We have to get the assignment done. Mr. Smith is going to collect it. (p. 119)

It seems that Ava does not have time to make sense of her material; she only wants to get it done to hand in for the grade. Because she does not make real sense of it, she never really does get to understand it. Many students who pass tests don't understand how to apply their knowledge, for example, the results that the National Assessment of Education Progress (NAEP) show that, although the majority of students can compute, they lack the knowledge and ability to apply those computing skills to solve problems (Carl, 1991, p. 2).

Grennon Brooks and Brooks (1993) note that constructivist teachers inquire about students' understanding of concepts before sharing their own understandings of those concepts:

> It's hard for many teachers to withhold their theories and ideas. First, teachers do often have a "correct" answer that they want to share with

their students. Second, students themselves are often impatient. Third, some teachers adhere to the old saw about knowledge being power. Teachers struggling for control of their classes may use their knowledge as a behavior management device: when they share their ideas, the students are likely to be quiet and more attentive. And fourth, time is a serious consideration in many classrooms. The curriculum must be covered, and teachers' theories and ideas typically bring closure to discussions and move the class on to the next topic. (pp. 107-108)

Constructivism in the Classroom: Assessment

"The aim is better assessment, not more" (Clarke, 1992, p. 24).

What does the constructivist's view of evaluation look like? The constructivist teacher uses strong context and meaningful tasks; tasks and activities that are real-world and in a relevant context, just as the instruction is; in other words, better assessment where the authentic instruction matches the authentic assessment.

Jonassen (1991a) views meaningful evaluation tasks that are set in a real-world context as "context-driven" evaluation. He believes that it is equally important to evaluate in the same rich context as it is to instruct in a meaningful context. These assessment tasks are known as "authentic" tasks as opposed to the kinds of decontextualized multiple-choice questions, for example, that perhaps are not so authentic. According to Jonassen:

> Authentic tasks are those that have real-world relevance and utility, that integrate those tasks across the curriculum, that provide appropriate levels of complexity, and that allow students to select appropriate levels of difficulty or involvement. We cannot all become masters of every content area. (pp. 29-30)

Appropriateness of Authentic Tasks

Authentic assessment is not a panacea, however. It is as inappropriate to evaluate students who have had a steady diet of worksheets and drill by using rich authentic tasks, as it is to evaluate students who have had rich contextualized instruction and by using decontextualized tasks. Not only is this unfair to the student, but it is invalid assessment as well. It is highly unlikely that the former would occur; most teachers who teach through a workbook test through workbook-type drill sheets and not by using authentic tasks.

These authentic tasks which are promoted in the constructivist approach are not used to sort winners and losers. This is rather refreshing for the students. Varying levels of difficulty are provided so that students can attempt the tasks and demonstrate what they do know. Once that is established, then the student and teacher can take steps to set goals and plan further instruction.

When students are repeatedly given tasks that they cannot do, day in and day out--and this often happens--they soon lose confidence in themselves as Katz and Chard (1989) illustrate:

> A third effect of the dynamic dimension is the long-term cumulative effect of repeated or frequent experiences. An experience may have a benign effect on a child's development if it occurs only once in a while, but a harmful effect if it occurs frequently over a long period of time. Teachers might not worry if a child is occasionally confused by the directions for completing school tasks, but frequent confusion may have strong cumulative effects on the child's self-confidence. (p. 18)

Goal-free Assessment

The constructivist view of evaluation of learning becomes less criterion-referenced and more goal-free because as learners interpret perspectives in different ways, the evaluation makes allowances for this wider variety of responses (Jonassen, 1991b; Scriven, 1983). That way, and happily for the students who experienced abuse of power in testing, "[e]valuation would become less of a reinforcement or control tool and more of a self-analysis tool" (Jonassen, 1991b, p. 12). Self-assessment then plays an important part, and why not? Surely the learner is more aware of how much and how well he or she is learning than anyone else.

Self-assessment

Self-assessment empowers students as learners. Teachers have a chance to give the learning back to the students. Again, it is the pedagogical leave-taking--taking leave, trusting the student--to learn and assess his or her learning to make informed decisions about the next action to take. Young children arrive at school already having had experience in self-assessment--it has been a part of their learning that they have had the first six years of life. How well the student did that the first six years of life! Why do we think that once he or she comes to school that now we have to take self-assessment away and empower ourselves with the issuing and assessment of his or her knowledge and understanding? Perrone (1991) agrees:

> We often hear that students aren't interested in their own growth as learners, that they don't want the responsibility for being involved in assessment practices. Yet when students have sustained opportunities to be active participants, to review for example, their own writing over time, they have become increasingly more articulate about their progress and what they need to work on to improve their performance and enlarge their understandings. (p. 166)

Scoring Guide Criteria

If the tasks are authentic, then the real-world environment that is being modeled in the constructivist environment will recommend the most relevant variables to evaluate the tasks. Johanssen (1991) presents this example:

> [W]e have been developing constructivistic, case-based learning environments in transfusion medicine, to prepare residents and third year medical students in how to assess transfusion risks. The criteria for success in a medical diagnostic environment are clear--a correct diagnosis and prescription of treatment that will save the patient within a time period limited by the severity of the problem and at a cost acceptable to the patient, the hospital, the insurance company, and so on. True enough, real-world may be very objective. But they are real-world, and to the extent that they reflect real-world criteria, they are meaningful. (p. 30)

Reporting the Results

As well as being tested fairly, students should have the right to have the results reported fairly, as well as clearly and honestly. This notion is endorsed by Alberta Education (1990) and British Columbia Ministry of Education (1994). The report should be summative and formative: summative to indicate to the student and parent how well the student is progressing compared to others of his grade or age, and formative to indicate the student's strengths and weaknesses for the purposes of program planning. If letter grades are used, they should clearly describe the outcomes of that grade level. In addition to letter grades, anecdotal reporting should be included as well to further individualize and clarify the description of the student's academic result and growth. Letter grades or their equivalent are best used after Grade 3 because up until that time, students are in very early stages of development and letter grades do not seem appropriate or even useful; anecdotal reporting for young children is more effective and fair. The reporting should as fairly as possible include all areas of development, not just the academic area. The report should clearly describe a plan for the progression of instruction that has been discussed by teacher and student, and or parent. From my experience, a three-way conference with student, parent, and teacher is a positive step to reporting results. That way, the reporting is a shared responsibility and not "done to the student." Three-way reporting does not objectify the student but instead shows respect to the other in a pedagogical relationship and minimizes misunderstanding about the student's growth and results. Davies, Cameron, Politano, and Gregory (1992) state:

"Together is better when the way we report includes parents, students, and teachers as valued contributors. Communication is improved when everyone has the opportunity to take part, ask for clarification, see specific examples, and know that they've been heard" (p. 21).

The students in this investigation wrote about their desire to do relevant tests; tests that were meaningful real-life. Students did not appreciate being required to study hard for understanding and then being given a test of trivia--of meaningless memory work. I believe that if they do not want to be tested on it they probably don't want to be instructed on it in the first place. On the other hand, it really depends on how the instruction is done. Take Paco, for instance, who was not by any means interested in the French Revolution. Paco was angry about the requirement to study the topic for the test. Perhaps if the information were presented differently as Katz and Chard (1989) suggest, Paco may have become interested. Like many teachers I do realize that not every minute of the day can we expect to have students spellbound, but I do think that we can make school a little more engaging and relevant. To do that, it often means taking risks and handing some of the responsibility over to the students. Aoki refers to this as "letting learn" or "pedagogical leave taking". Aoki (1990) writes:

> Often a pedagogical tact in teaching is to say to a student, "I leave it to you," suggesting a letting go of decision-making to the student. Such an understanding reflects teaching understood as delegating or allocating power assumed to reside in the teacher. (p. 39)

"Letting learn" seems to be a matter of power sharing and it is interesting that the student recognizes this and teachers don't--or do they? Students are more likely to learn prescribed material if they have some choice in the context of the learning, for instance, learning to write a friendly letter through the real letter writing activity to a penpal or cousin. Assessing by checking for punctuation, the appropriate greeting, and addressing of the letter are learned through a meaningful activity rather than correcting a contrived error-laden letter in a workbook. Knowledge in the real sense develops this way when there is meaning and purpose underlying the activity. Perkins (1992) gives an example of a situation in which students learned for understanding in context, could remember and apply the knowledge better than students who learned through conventional means. According to Perkins there are various kinds of knowledge that students develop. He labeled them "inert knowledge," "naive knowledge," and "ritual knowledge":

> Conventional instruction--reading textbooks and listening to lectures--tends to produce inert knowledge. For example, cognitive psychologist John Bransford and his colleagues conducted an experiment in which some students read items of information about nutrition, water as a standard of density, solar-powered airplanes, and other matters in the usual textbookish way, with the intent to remember. Other students read the same items of information in the context of thinking about the challenges of a journey through a South American jungle. For instance, the students read about the density of water in the context of how much the water the traveler would have to carry.

> Later, both groups of students were given the task of planning a desert expedition. The students who had studied the information in the conventional way made hardly any use of it. But the students who had studied the same information in the problem-solving context of the jungle journey made rich and extensive use of the information, pondering the kinds of foods that would sustain people the best, worrying about the weight--and so on. (p. 22)

Conclusion

It is clear that good assessment is not possible without good instruction, such as the kind Perkins (1992) recommends. Perrone (1991) agrees:

> I raise these teacher, school, and pedagogical issues . . . to make clear that the larger conception of evaluation cannot go forward without a larger conception of teaching. If teaching is skill sheets, work sheets, textbooks, basal readers and simplified explanations, a larger view of assessment is not likely to take root. Who wants, for example, a portfolio of skill sheets? We have a chance to construct something better. (p. 166)

We do indeed have that chance to construct something better, but it is not easy. I have found that when I have tried to introduce to some teachers various authentic assessment tasks such as performance-based assessment and portfolio assessment, teachers do not know how to apply the information I am giving them. In the first place, how can they implement this kind of assessment when their instruction is through workbooks and worksheets? What needs to change first is the instruction. Only when meaningful instruction is in place only then can they consider or comprehend meaningful assessment.

Real-world, meaningful, engaging instruction deserves the same kind of assessment; assessment that is situated in context and that requires students to make judgments. The scoring criteria should be open-ended so that multiple viewpoints can be accommodated to meet the criteria of acceptable performance levels of the student. In addition, care must be taken to introduce this kind of assessment to students, considering their past experiences of assessment, that often focus on grades and competition. For many students this is a very real paradigm shift.

Linking Good Instruction to Good Assessment

> In a constuctivist setting, assessment of student learning is done naturally within the context of lessons and activities. Teachers analyze student products and exhibitions as benchmarks and garner information for use in developing future activities and informing ongoing practice. (Grennon Brooks & Brooks, 1993, p. 122)

Often, good instruction is not linked to good assessment, nor is assessment seen as a natural or ongoing process, as Grennon Brooks and

Brooks (1993) advocate. Students are shortchanged with decontexualized learning and tests. As one would surmise, then, poor assessment matches poor instruction as this example shows:

> In a next-door twelfth grade class, another teacher dictates a definition of "interpretive literature." All the students are writing it down in their notebook. They are not personally engaged by those terms. There is no time in this class to use the "student discussion method" because the teacher feels the pressure to teach for the final exams. She hopes the students will respond properly to a multiple-choice question about the concept of interpretive literature. Yet, it is not likely that this concept will help them make sense of their experience in reading literature." (Hoffmann, 1964, p. 171)

The example in the literature class indicates that little real learning is taking place. Not much sense is being made of the literature and no doubt after the exam, the learning will vanish. This example is not unique. Katz and Chard (1989) would agree that insignificant understanding is indeed taking place in many schools. Learning for a test is meaningless learning, and yet strangely enough, students are rewarded consistently for this. This seems odd that students are rewarded for not really understanding. So why should they bother trying to understand? It takes too much time and effort anyway. Little wonder students are confused, but most often they do not question the authority of the teacher (Apple, 1979). Teachers often knowingly teach to the poorly constructed tests instead of for understanding because there's no time for in-depth learning. It seems it's a vicious circle. Teachers themselves are rewarded indirectly for having their students score high on these kinds of tests so they give students ample practice on them. There seems to be something inherently wrong when teaching for understanding and learning for understanding are often not the focus nor rewarded.

"We learn every time we make sense of something; we learn in the act of making sense of the world around us. Understanding takes care of learning" (F. Smith, 1986, p. 28).

Performance-based Assessment Outside of the School Context

Although it has not been done so very much in the past, learning for understanding is being valued more and more in the real-world. It seems that schools will have to take the lead from the business world and re-look at the way they carry out instruction and assessment. For example:

> Authentic activities (tasks and problems already relevant or of emerging relevance to students) also relate to a particular body of knowledge, but rather than structuring assessment around specific bits of information, they invite students to exhibit what they have internalized and learned through application. (Grennon Brooks & Brooks, 1993, pp. 96-97)

Business leaders are aware of what good measurement is. In the article, "Measurement Traps" in the Globe and Mail, Clemmer (1994) says that "weighing yourself 10 times a day won't take off pounds." Similarly, educators should realize that testing students 10 times a week won't improve learning. He further reports that "Measurements that don't lead to meaningful action aren't just useless--they're wasteful" (p. B26). Teachers could learn from business leaders such as Clemmer as well. If the measurement or assessment does not anticipate action, then teachers should reconsider doing it.

Interrelation of Assessment and Instructional Tasks

Students must see the link between assessment and to instruction as well. The two should be so well integrated that an instructional activity such as essay writing, with some slight adjustment, can become an assessment activity. The students in both instruction and assessment have choice of topic and a scoring guide. In both cases, time is given for planning, writing drafts, and polishing. When students work with the open-ended scoring guide criteria, and that means criteria that are not stiflingly specific, students will then know the targets and write to them. The criteria ensure that students have the signposts to a good essay, but do not restrict the student's voice. During instruction a teacher in a constructivist classroom coaches students in developing their own criteria for stories and poems they write (Perkins, 1992).

Student Involvement in Peer Assessment

Another way students can become directly involved with the assessment as it relates to instruction, is suggested by Perkins (1992). He suggests that for some quizzes, the teacher has the students grade one another but that the teacher gives strong feedback where necessary about the kinds of answers that make sense. This kind of instruction is giving up the power or ownership that the teacher has. Another example demonstrates this: The teacher involves students in writing story problems in mathematics for one another to heighten the intrinsic interest of the problems. Students automatically write problems in real contexts, contexts that have meaning for them, for example problems that are embedded in the context of their school instead of context made up by an expert for a workbook. Students as young as Grade 1 can write these problems for peers and their peers can assess their story problems as they do them (Silverman, Winograd, & Strohauer, 1992).

Student-Constructed Tests

"Student-constructed tests offer both an effective assessment tool and a powerful review strategy to assist students in organizing their knowledge of the topic" (Clarke, 1992, p. 27).

Clarke (1992) suggests a meaningful activity for students is to have them construct their own tests occasionally. To do this successfully, students must be aware of the course content, and make interesting and personalized

questions for the class. In addition, they also have to prepare an appropriate scoring guide. This works successfully in university as well, although when I first introduced this idea, students were a little apprehensive. They have difficulty with the idea of knowing the test content before the test takes place because some realize that they lose the competitive edge. Of course what happens is that the students learn the material in a meaningful way, and by the time the test day comes, writing the test is just a formality. Students have appreciated this form as a change and found it was far less stressful, but on the other hand many commented that they actually spent a lot more time on the test than they would have expected. The following are two journal entries that students wrote about the experience, and shared with me:

Education Student (Trish) Journal Tuesday, November 10, 1992

> It was an absolutely wonderful experience building questions for our own exam. When we broke up into our groups we began to discuss what we felt stood out in the course. What had taken up a major part of the course? What struck us as most important? What would make a good exam question? Quite often, jokingly, we discussed what type of questions we could make that would stump the rest of the class. Other times we came up with extremely simple questions that could have made the exam a cinch. It is amazing though, how fair the exam questions turned out to be. I think this was a great exercise for beginning teachers. Too often we're fearful that if we allow students to design their own test, or to take ownership of what they do in any way for that matter, they'll run away with it and make a cakewalk out of anything they design. Therefore, we feel we must design everything so the students will be challenged. We fail to realize that students are often harder on each other and themselves than we are on them. If we'd just step back once or twice, I think we'd be utterly amazed by what we see.

Education Student (Lily) Journal November 17, 1992:

> I just had the most wonderful experience!! My evaluation test, can you believe it? A test! Nola told us in advance that it would be a no-stress test but I didn't really believe it would work. We wrote the test and I think everybody tried to make questions that would test our knowledge of the course. Many of them dealt with the same topics but she left them in anyway. I guess that shows we placed a lot of value in those areas. This is just great! I can't believe it. I could have written this test by only talking about portfolios but I chose questions that covered every type of assessment we studied. I'll bet lots of others did too! I can hardly wait to ask them. I didn't have to study because I knew the material and knew I did. What a feeling! And the really good thing is, I know it, not just have it in short-term memory. If this strategy works in the classroom it'll be the greatest thing since sliced bread! It was so friendly!

Students valued this way of testing and I do believe, as it was for Lily, they knew and understood the material so well that the test was really only a

formality. Students appreciated that I "stepped back" and trusted them to make the well balanced, meaningful and context-rich test that they did.

"Time" for Teachers

"Assessment should anticipate action. The most useful criterion when deciding whether to engage in a particular assessment activity is, 'What action will result from this assessment?' If the answer is 'None,' do not carry out the assessment" (Clarke, 1992, p. 29).

Teachers As Reflective Practitioners

Purposeful Activity

Much assessment and instruction that is carried out daily lacks purpose. Large amounts of time are lost because many teachers thoughtlessly engage students in purposeless activities. If more teachers reflected and thought about purpose as Clarke (1992) proposes, then the time in class would be used more productively. Purposeless instructional and assessment activities occur in any classroom at times--teachers get so caught up in the busy hubbub of teaching that the purpose can easily be lost. As a teacher I often found myself and my students very busily engaged in an activity. Sometimes I stopped and asked myself, why are we doing this? As I gained experience and confidence, I decided that if I could not justify doing an activity for educationally sound reasons, we stopped doing the frivolous activity immediately because enough time had been wasted already. By continuing the meaningless activity, we were going through the motions of schooling: I was busily providing irrelevant experiences and students were busily having them.

Who Owns the Store?

Perkins (1992) notes that a testing culture promotes the notion of a teacher-centered classroom while an assessment culture requires a student-centered classroom. In an assessment centered classroom, the teacher shares power as he or she does knowledge. It really means that the teacher is no longer the "sage on stage" but the "guide on the side." Teachers need to step back as the student, Trish, suggested and trust the students to take some of the responsibility for their own assessment and learning. Perkins believes that:

> It is this act of stepping back that enables teachers to practice and infuse the habit of reflection into their own pedagogical approach. In this light, teachers become researchers in the classroom, posing central questions to better inform their sense of student learning, their approach to teaching strategies, and the development of their own reflective habits. (p. 65)

A Parent Speaks Out

Parents value their children taking responsibility for their own learning and ownership of work too. As a parent I was frequently disappointed with the

carbon copy "art" work that my young children brought home. It is hard to cover up disappointment and I commented on what evidence of creative work that the teacher "allowed" my children to do. I see that I was not alone in my disappointment; other parents were, and still are, disappointed with the lack of ownership of learning that their children have. The following is a letter from a parent lamenting this lack of ownership:

"A Question of Ownership"

My son Nicholas came home from his first grade class today carrying a paper puppet, his first art project. Full of excitement and some degree of pride he handed me his puppet waiting for my reaction. As he had attended a wonderful kindergarten program and had come home with many an inventive project, I'm sure he expected my usual enthusiastic response.

My reaction this time, however, was a bit different. In my hands I held someone else's drawing that my son had carefully colored, cut out, glued together and then handed to his teacher to staple on the strip of paper that they were to use to hold the puppet. All I felt was hurt and disappointment, after all, I was losing a bit of a dream I had for my son's education.

I carefully commented on the interesting color combination he had chosen, on the excellent way he had colored and glued the picture together. I asked him if his teacher knew he could use a stapler, he simply replied that he hadn't told her he could. That was all really I could say and I put the puppet away.

But the puppet didn't go away because I was still left with a bad feeling. Later when my five-year-old daughter came downstairs with her puppets that she had carefully created on her own, using Nicholas' puppet as a stimulus, the message was driven home and I knew why I had been so upset.

I loved my daughter's puppets because we could truly celebrate them; they were hers! She really owned them. She owned the idea she used to create them, the skill in cutting them out, and she even owned the problem-solving she did in trying to figure out just how she was going to attach the stick part to her puppets. Because the materials she needed to do her chosen project are always available she didn't have to ask me for anything and risk my interference.

I sat down with my son then, with a large puppet he had created last year. I asked him who thought of the idea, who designed him and put him up on our wall. He had of course. I then explained to him that, that was why I loved this puppet. He owned it. He also owned the learning that went with it. I then asked him who owned the puppet that he had brought home earlier today. We decided that the artist owned the drawing, the teacher owned the attachment and that Nicholas owned the coloring and gluing--the cutting didn't count because he just followed a line that was already drawn. So we could comment on what Nicholas owned, but we both agreed that his earlier puppet gave us more to celebrate.

It is no surprise that the Mother's Day cards that I have chosen to keep are those that my children had so carefully drawn and made themselves--the ones that they owned and created for me. The cards were a part of them, their personalities, their growing skill and creativity in drawing, designing, coloring and cutting, that they gave me for Mother's Day. But I did not keep the commercially-made cards--they belonged to someone else.

Teachers and Change

The constructivist approach seems to be favored by students and researchers, and yet many teachers who hold on to traditional approaches do not see the need for change because their current approaches seem to work well for their students. For example, Grennon Brooks and Brooks (1993) note that students take comprehensive notes and pass important tests; perform well on worksheets; complete assignments neatly and on time; write well-structured and well-researched individual or group reports; and receive good grades for their work. Students have learned to conform in these classrooms, the teacher being the determiner of what is right and wrong, and no questions asked. What happens to students is that as some students contentedly go along with this, the rest disengage. Many teachers are more than comfortable with this too and prefer to keep it that way because they see the sharing of power as "a threatening break from the unwritten but widely understood hierarchical covenant that binds students and teachers" (Grennon Brooks & Brooks, p. 102).

Coping With Change

The students' stories about their experiences of taking a test contained many uncomplimentary remarks about teachers--prompted most likely by the topic that they wrote about. Students do have a case for not being positive about their teachers or tests because much of the time they have been unfairly and inappropriately assessed. This has occurred because many teachers are not well informed about testing and unknowingly are providing poor tests and conditions under which students must perform. Many teachers are uncomfortable about testing and grading and this is unfortunate because they are missing the real celebration of growth and effort that authentic assessment can provide. Teachers need support at this time, not criticism that is not constructive. I believe that most teachers care for their students and work very hard at being the best teachers they can be, but external factors and pressures make it difficult for many to do the pedagogical job they would like to do. Perkins (1992) concurs:

"[M]ost educational settings neither labor very hard to build teachers' understandings of new instructional perspectives nor allow teachers the flexibility or freedom from the coverage fetish to pursue more enlightened instruction" (p. 52).

In building new educational understandings teachers must be given the freedom and flexibility to make errors, because change is transformative and

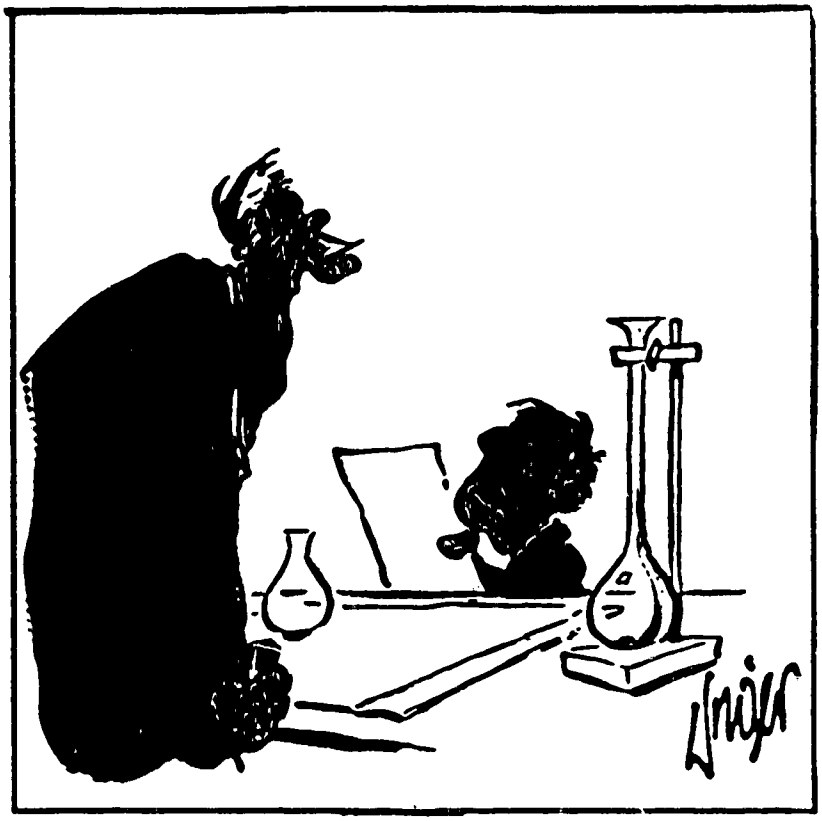
does not occur overnight. There are going to be wins and losses during this time of change. Doll (1989) perceives change through a post-modern view:

"A post-modern view looks upon change in an entirely different light. Change is seen in transformative, not incremental terms; and errors are seen as necessary actions in the process of development" (p. 249).

The cartoon in Figure 14 illustrates the fact that education is a shared responsibility. While a good part of success in learning depends upon the effort, ability, and disposition of the student, we forget often that poor marks may indicate some weak instruction and assessment by the teacher.

Figure 14

Cartoon: Education is a Shared Responsibility



"I think you'll find my test results are a pretty good indication of your abilities as a teacher."

Suggestions for Teachers

From research and my own experience, I have some suggestions for teachers to help them cope with change particularly in their implementation of authentic instruction and assessment.

Less Is More

To help students learn better, teachers must do less. Teachers must trust the students and let them learn by allowing students to take responsibility for their own learning (Doll, 1989).

Uncover the Curriculum

Second, teachers must stop "covering" the curriculum because indeed, that's all they are doing--only covering it instead of uncovering it to reach understanding. Covering a curriculum discourages deep learning (Doll, 1989). Marzanno, Pickering, and McTighe, (1993) agree: "When teachers plan lessons, they often do not consciously consider activities or strategies they might use to help students develop productive habits of mind. They focus instead on content and the need to 'cover the curriculum'" (p. 3). Gardner in Brandt (1993) too, believes that "covering the curriculum" is detrimental to real learning:

> The greatest enemy of understanding is coverage. As long as you are determined to cover everything, you actually ensure that most kids are not going to understand. You've got to take enough time to get kids deeply involved in something so they can think about it in lots of different ways and apply it--not just at school but at home and on the street and so on.
> Now this is the most revolutionary idea in American education--because most people can't abide the notion that we might leave out one decade of American history or one formula in math or one biological system. But that's crazy because we now know that kids don't understand those things anyway. They forget them as soon as the test is over--because it hasn't been built into their brain, engraved in it. (p. 7)

What Gardner is saying is that teachers are so busy covering the curriculum that they don't give themselves time to teach--at least not for understanding. For example once they've "covered insects" even though a child might come excitedly into the classroom with a newly discovered insect a month after the "insect unit" has been covered, its discovery is played down because it has already been "covered." No wonder the child goes away confused. Only a month ago the teacher was excited about insects! Stepien and Gallagher (1993) advocate:

"In the place of covering the curriculum, teachers give learners opportunity to probe deeply into issues searching for connections, to grapple with complexity and to use knowledge to fashion solutions" (p. 26).

Having advocated not to cover the curriculum, I empathize with teachers who must "cover" topics for the Provincial Achievement Tests.

Teachers are in a dilemma--doing something that they feel is inappropriate pedagogy such as rushing students through material before they have had time to internalize concepts. It is little wonder that teachers soon lose their sense of professionalism--or is it being taken away by external testing for the purpose of accountability? Van Manen (1991) is concerned about the relation between pedagogy and politics too:

"Rather than think of themselves as pedagogue-scholars and intellectuals, teachers are encouraged to see themselves mainly as instructional disseminators, knowledge retailers, technicians of externally determined curricula" (p. 213).

Shake off the Albatross

Third, teachers could consider carefully the use of commercial materials, or put them to rest completely. Workbooks and texts tend to "cover" material--it is shallow learning. Shaking off the workbooks like Coleridge's albatross, is difficult. Teachers often feel that they must be used because everyone uses them, or every page must be used whether it has purpose or not. Worksheets and workbooks are tangible evidence that students are appearing to work and learn. Students may look like they're working, but filling in blank after mindless blank does not necessarily indicate that they are learning. Real learning, however, is not so visible. It is inconspicuous until it is applied in the real-world. There is not a lot of authentic instruction or learning in workbooks. The tangible proof or learning through workbooks and tests has run rampant according to Mayer (1961) in Grennon Brooks and Brooks (1993, p. 96):

> [T]he rush of tests and examinations and weekly quizzes, of workbooks and homework, of recitations and catechisms by which children everywhere--but especially in America--are made to prove that they have learned their lessons. If the child cannot give back on demand what he has been taught, it is assumed that he has not learned it. (p. 87)

Rather than testing frequently and superficially, it is better to test less frequently and more authentically to find out what students have just memorized or actually internalized. Grennon Brooks and Brooks (1993) agree:

"[A]uthentic assessment tasks require students to apply prior knowledge to new situations, the teacher is able to distinguish between what students have memorized and what they have internalized" (p. 97).

Other albatrosses might include inflexible time tabling, rules without purpose, and adherence to methods and practices that "have always been done that way." By reflecting on and questioning practice, by searching for educationally sound purposes, only then will the albatrosses that have been weighing teachers down will leave.

Once teachers discard these artificial learning devices, many find that they become more resourceful and creative--and so have their students. Students are now not mindless puppets filling in blanks--they're now students who have to think and make judgments. Some students do not like to have to think to create their own story, and use punctuation instead of using someone

else's story to punctuate. Some teachers say students like worksheets, but we must be careful about that generalization even though it may be true. As Katz and Chard (1989) say, just because students like something it doesn't mean it is good for them. A little junk food is acceptable occasionally, but a steady diet of it does not promote optimum growth and development.

Meaningful Professional Development

"When teachers stop growing, so do their students" (Barth, 1990, p. 50).

Teachers must take time for themselves to reflect and take a chance to do less. They must have time to think how they can allow their students to play a greater part in their own learning and assessment. And they must be supported in order to do that. As Perkins (1992) says:

"Nothing drains energy more than having too many things to do and too little time to do any of them near well" (pp. 4-5).

Teachers are drained because they are too busy, most often doing busywork. Many teachers are the first to admit this, that things are done without real purpose because they've always done it that way, or that they just don't have time think about it. Once teachers do get time to think and talk with other teachers, they will be more likely able to reconsider different and more effective ways of instructing and assessing. Teachers need time to learn how to work with others because they are not going to be able to operate in isolation any longer because there are just too many demands on teachers. Costa (1991) in Perkins (1992) agrees:

"When the conditions in which teachers work signal, promote, and facilitate their intellectual growth, they will gradually align their classrooms and instruction to promote students' intellectual growth as well" (p. 221).

Without doubt teachers need time for themselves. Such professional development is a worthwhile investment because the benefits go directly to the students. In many cases, and increasingly so in light of the recent cutbacks in Alberta, teachers do not and will not have sufficient professional development for any change in thinking to take place. In fact, the professional life of a teacher is not something to be envied. Barth (1990) likens the professional life of a teacher to a 'tennis shoe in a laundry dryer' (p. 1). Fullan (1991) stresses the importance of professional development too. He argues that teachers will never improve learning in the classroom unless they help to improve the conditions that surround the classroom.

It is unfortunate that often the public does not appreciate the importance of giving time for teachers to regroup, reflect on, and rethink their practices. Time for teachers to reflect must come into place before we can see effective change in the classroom; change that includes authentic assessment that is truly linked to authentic instruction. Not only are professional

development days necessary, day to day professional development among the school staff is as, or even more effective; in other words, it is a matter of "collegiality." Perkins (1992) agrees:

> What is key to a more orderly and enlightened profession? Of many factors, Barth underscores collegiality. Collegiality means something different from congeniality . . . . It's not just good manners and telling jokes in the teachers' room. Collegiality means working together in a mutually supportive and thoughtful way at the business of education. (p. 222)

Barth (1990) borrows a four-way characterization of collegiality from Judith Warren Little who says that in a collegial atmosphere teachers not only talk about teaching but they observe each other, teach each other and work on curriculum together. Barth concludes then, that a school that serves as a home for teachers' minds is one that will become one for students' minds as well.

The collegial atmosphere contributes to the empowerment of teachers when they feel comfortable enough to share their stories or narratives, their ups and downs (Connelly & Clandinin, 1990, p. 4). Once an atmosphere of collegiality occurs, a feeling of "connectedness" among teachers follows, and teachers are more likely then to take risks and make errors in a supportive atmosphere.

Even in a collegial atmosphere, to implement this idea of authentic assessment, Perkins (1992) says, teachers must "Think big, start small," and do not work in isolation. Perkins believes that if too much is taken on at once, and in isolation, teachers will become disillusioned and go back to their safe, secure, and orderly ways. In assessment and instruction there are practical ways to work as colleagues. Most important, teachers need to rethink their instruction as they rethink their assessment, or even before they rethink their assessment. Authentic assessment will not occur with inauthentic instruction. Until teachers understand authentic instruction, they will not truly implement it. Their classrooms may look as though authentic instruction is going on, but unless they understand why they are having students do the activities they do, it will be superficial--and so will be the assessment. When engaging in constructivism for the first time, teachers need to articulate the wins and losses. They need to compare notes, share strategies that work, and share assessment ideas and materials. Working collaboratively enriches instruction and assessment, and the teacher's whole professional life.

CHAPTER IX

Conclusion: The End of the "Hodos"--And New Beginnings

I would like to conclude with the rest of my story; where I have come from, and where I think I am going in assessment. The possibilities are many but I feel the students have given me a fair indication of the appropriate direction I should take. This chapter addresses my beliefs in assessment and instruction--those reaffirmed, and those that have been "put to the test." I will share my growth as a learner, how I have changed in my thinking, and finally how I will implement changes in my future instruction and assessment.

My Growth as a Learner--The Eighth Assumption Now Unconcealed

During the latter stages of my research journey I discovered another assumption that was not evident to me when I began the research. I believe this is the time to address it, just as it evolved in the story of my research. My eighth assumption, then, is that **assessment should drive instruction**. If the assessment is a good one, one that involves, for the most part, ill-structured tasks in rich context, then that will force good instruction that is meaningful, relevant, and worthwhile for the student. I took part in preparing good assessment tasks at Alberta Education where we developed performance-based assessment tasks in mathematics. These ill-structured tasks were used to help guide and encourage teachers in the province to teach using such tasks instead of the predictable textbook exercises in mathematics. Because these were prepared for the Provincial Achievement Tests, the teachers would be forced to teach to them, in other words, teach using authentic tasks so that students would be familiar with them on the test. We did the same with the multiple-choice component, where problems were situated in rich real-world context and structured so that students were really perplexed in problem solving instead of just routinely performing step-by-step problems.

My Beliefs: Re-looking At Instruction and Testing

How has my thinking changed through the process of this research? I have discovered that this research has reaffirmed some beliefs and it has caused me to question others. I did not come up with answers, but instead, more questions and possibilities. I will deal with my reaffirmation of beliefs first, and then the beliefs that are now up for further question.

Reaffirmed Beliefs: Kitchen Math and Backyard Science

I firmly believe in what I call my constructivist-based "kitchen math and backyard science" approach to authentic teaching and assessment. It is a real-world, meaningful learning in rich context. For example, students use real objects in context to learn. Cooking in the classroom takes care of a good part of the reading and math program as do the activities and projects done in the outdoors without the aid of artificial props. Real money is used in

measurement, real apples are used in fractions. Students write their own stories to learn spelling and grammar. They construct their own booklets, design their own covers in projects. The students' creations belong to no-one else, in stark contrast to Nicholas, who "owned" very little of his puppet as his mother so astutely pointed out. The students in the study indicated that they learned best through meaningful instruction and assessment, and research supports the students' wishes for authentic instruction and assessment. I notice that Gardner (1991) recommends activities such as these as well because the mathematics and language are situated in rich context and are needed and valued in the real-world (p. 212).

I realized during the study why teachers had such difficulty in understanding the various alternative assessment approaches. No wonder teachers are bewildered by authentic assessment. Those relying on workbooks and packaged instruction (instead of "kitchen math and backyard science") cannot begin to assess authentically because they don't have anything authentic to assess! The plethora of prepared "educational" packages is taking away genuine learning. I noticed that now students' journals have been taken away from them. Phi Delta Kappan has produced "My Stories" which contain outlines for journals and topics for students to write about. Included are "fun" pictures for motivation. The pages are 8 1/2 oy 11, and consist of 25 pages for writing. There are suggested activities and notes that "provide[s] guidance on how to use that page of the writing journal." The children do not have choice on the size of pages for their journal; they do not have choice on how many pages they might want in their journal, or color of pages. Nor do they have opportunity to construct and staple the journal. The topics are selected for them. Any thinking that students would have done has been taken away for someone else to do. The students will own very little of their journal. I realize that at times teachers need the "crutch" of packaged materials, but do they consistently need them? If so, when will we stop interfering and let the student learn? As long as we keep stealing learning opportunities from children and their own stamp, we will have superficial learning. The children own nothing. If children are allowed to make their own booklets, their own projects and decorate them with their own drawings instead of just coloring prepared drawings, they will experience more empowerment, responsibility, and ownership. It will be theirs. If teachers are interested in their students' real learning, they will have to move away from the superficial packages that are fed to students and the test-teach-test mode that many are in now, and embrace a more constuctivist approach before any teaching or learning or assessment is authentic.

Beliefs Open to Question

Measurement Driven Instruction

I do not believe that we can begin to change assessment practices until we have sound teaching practice in place. Enforcing authentic assessment by including it on the mandatory Achievement Tests is really putting the cart before the horse, as teachers have already pointed out. In theory it does sound workable, but in practice in most cases, it has mediocre success. Teachers not only have to buy into constructivism, they have to understand it as well.

Merely carrying out "authentic" assessment is not authentic at all unless it is tied to authentic instruction. Measurement Driven Instruction (MDI) rests on the belief that if the test is good, then teach to it. In other words, good assessment implies that good instruction will have to be in place. This brings me to my question of "teaching to the test." I can see how logical this is but my experiences in administering authentic assessment to students who have not had authentic instruction makes me re-look at this practice. A couple of years ago when I worked on performance-based assessment and carried out some field testing for the International Assessment of Educational Progress (IAEP) I had to administer performance-based assessment (authentic assessment) to several classes of junior high students. To this day I wonder about the validity of the results. The authentic assessment consisted of lab-type activities in math and science. The context was rich, as were the activities rich in meaning and purpose. In one class I set up the activities in the science laboratory in the school. I noticed the students taking great interest not in the activities so much, as the room. I soon learned that this was the first time they had ever been in their science lab. Their science lessons were taught out of the science textbook--chapter by dreary chapter, and their only activity was answering the perfunctory questions at the end of each chapter. The students were unfamiliar with hands-on activities, yet were visibly delighted with the change of pace, and were fully engaged in problem-solving for the two-hour period. Nevertheless, this was not like their regular instruction, and the assessment did not link to the instruction in very many ways at all. The poor results that the students achieved were not a comment on students' problem-solving abilities so much as the decontextualized instruction that they had received. Students spent a good part of the time becoming accustomed to using manipulatives and managing them for the first time. A more appropriate (although not more educationally sound) assessment for these students would have been a paper-and pencil test containing short-answer and objective-type questions.

E. Eisner (personal communication, 1990) attests to Measurement Driven Instruction (MDI), as does Wiggins (Brandt, 1992) in his statement of "teaching to the test" as a way to implement change. They argue that better assessment instruments will drive better instruction, but I believe that this is a somewhat naive assumption. While that may occur in some cases--and it certainly is logical--from my experience in working with hundreds of teachers, it just doesn't necessarily work that way, nor is it that easy to do. We talk about students having to understand to make sense of things and yet we don't give the same opportunity to teachers to understand before they buy into a new teaching approach. For several years when I worked at Alberta Education I worked with teachers and advocated teaching to the authentic test. That was my first mistake. Many teachers did not know what authentic instruction was, let alone authentic assessment. Putting the cart before the horse clearly was not making a lot of sense to teachers. Many teachers valiantly tried to put authentic assessment in place; for example, they constructed authentic activities in assessment, and implemented portfolios, but their instruction did not match the assessment. The portfolio of worksheets somehow missed the mark. And, naturally enough, frustration set in. For one thing, workbooks worked fine before, and one teacher cited the adage, "if it ain't broke, don't fix

it." As it was, students worked page by tedious page through the workbook without much protest, most likely because they did not know any better. This observation is supported by Herman et al. (1992) who believe that "it is not that tests ought to drive the curriculum, or that teachers ought to teach to the test. Rather, good assessment is an integral part of good instruction" (p. 5).

A second mistake in implementing authentic assessment is not taking the students' point of view into consideration. Imposing authentic assessment on students overnight was also inappropriate. Students' enthusiasm in handling these new approaches disturbed the harmony and security that they had with their familiar passive learning. Students need to know why these changes are taking place so that the purpose is clear. It was a similar experience for teachers when they attended a one-day workshop on math manipulatives. The teachers diligently constructed manipulative math activities for students and expected the students to handle these appropriately at the first introduction. The purpose for using the manipulatives was not clear to teacher or students; in other words, the activities were in place before the thinking was, and the result was failure. Consequently, the workbooks were hauled out once more. Perkins (1992) cautions:

"We must not expect new technologies, the grouping of students, and the innovations to do the job by themselves. We must accept the responsibility of mediating students' good use of person-plus resources" (p. 148). Gardner (1991) agrees:

> The presence of a curriculum that is worth assessing is a step in the right direction. Unless teachers accept the curriculum, however, and not only believe in it but embody its precepts in their teaching, the best curriculum and ways of assessing are of little value. To the extent that master teachers believe in what they are teaching and how to assess progress and understanding, they become excellent, indispensable leaders of the educational process. (p. 254) [emphasis added]

Third, I believe that enforcing authentic assessment through external testing is underhanded and questionable, in practice and in moral terms. I, too, assumed that if we enforce authentic assessment on the Achievement Test, teachers would be obliged to change, and be enlightened about the "right" practice. Teachers would be eternally grateful to us for showing them how better to teach. Whether or not they understood what they were doing was not a real consideration. While I believe authentic assessment and instruction is best for students, for some teachers it might not be "right" for them at this time. But then here's my dilemma--accept the inappropriate practice that is now going on in some of the classrooms, or push for MDI? There are countless examples where students are not being instructed well in various subjects. For example, we know how little mathematical understanding by students is occurring in schools at this time. Most instruction is driven by paper-and-pencil activities with few, if any, concrete experiences provided. There needs to be change, but should it be through MDI, such as by the "top down" external tests? Are there better, more humane, and effective ways to do this? One of

the dangers of MDI is that teachers realize what gets approval and so they make their classroom appear that appropriate instruction is occurring. The classrooms may look as if authentic instruction is taking place but unless it is thoughtfully done and understood, it is only an appearance. I have visited several classrooms with centres and math labs attractively set up, but really little was happening in the students' minds--their minds were not engaged in thinking as they mindlessly moved objects about. Children were working passively side by side, in isolation, and not working cooperatively or using language to make thoughtful connections in creating new knowledge.

Gardner (1991) supports my observations about superficial authentic instruction and assessment, and the pressures that teachers experience:

> A school will not work better simply because it styles itself after a children's museum or institutes reciprocal teaching or mandates the keeping of process-folios. Children will not attain understanding just because they watch masters who exemplify understanding in their own practices. (p. 205)

Gardner goes on to make the important point that if reasons underlying such innovations are accepted and teachers and administrators are searching for ways to implement such innovations then there is a possibility that the assessment and instruction will authentically be put in place.

It is possible that eventually understanding of authentic instruction and performance will occur through external testing, but I believe it is the more difficult route to take. In the meantime, teachers have built up a lot of resentment towards authentic assessment because of the top down nature of "teaching to the external test" which is mandated in Alberta. While these are well-constructed tests developed by teachers on the Alberta curriculum, these features provide little comfort for the teacher who has to administer them.

There are merits to our provincial Achievement Testing Program. The tests are well-constructed by teachers and they are valuable in obtaining a snapshot of student achievement. The results do improve programs and teaching practice. However, they leave a sour taste in the mouths of educators because of the top down nature of the program. If we trust teachers like we do students in their constructing of their own tests as Lily and Trish said, the program would be welcomed. Regarding the merits of top down implementation, Fullan (1994) writes:

> Small and large-scale studies of top-down strategies (whether employing voluntary or mandatory methods) have consistently demonstrated that local implementation fails in the vast majority of cases. (p. 186)

I often wondered if I were the only one feeling uncomfortable about the MDI as it was externally administered by Alberta Education? Upon further investigation of this, I read Popham's research in his text Modern Educational Measurement (1981). He believes there is a conflict between teachers and measurement specialists as each operates in separate worlds. He believes that neither teachers nor measurement specialists can afford to be specialists in only one area; each should have knowledge of both to be effective. He found

that, "Teachers would whip up an instructional storm, then testers would swoop in to see how much students had learned" (p. 412); in other words, "Ne'er the twain should meet." Popham adds that instructional specialists were often intimidated by the "measurement crowd." How true! I confidently "swooped in" with my precious boxes of testing equipment and teachers upon sighting me, virtually scattered. Very rarely did a teacher peek in the room to see what was taking place with their students and yet I certainly put out the welcome mat. But I see now that I was viewed as a "specialist" and what I did had little to do with their real-world. Popham thinks that for the well-being of students, this is truly tragic and this situation must be altered. There needs to be collaboration between both camps. We do have some collaboration in the Achievement Testing Program in that teachers and test development specialists do create the tests. However, the number of teachers involved in this process is very small considering the number of teachers in the province. In addition, the teachers involved in these test construction activities are usually not the run-of-the-mill teacher; these are teachers chosen by principals and superintendents who work for Alberta Education because of their excellent teaching expertise and content knowledge. How they view and teach instruction and assessment is not often typical of what is done by the majority of the teachers in the province. Popham believes that the chief impediment to collaboration is ignorance and that when people are ignorant of someone else's operation, they fear it, or at least stay clear of it. I think from my experiences, Popham is right. Once these barriers are broken down, only then will measurement driven instruction be effective.

Multiple-Choice Test Questions

For several years I defended and touted the value of multiple-choice questions to teachers, parents, administrators and students. However, now I have less respect for multiple-choice questions than I did before I started this research. While they seem valid--the right students get them right most of the time, why are they so unpopular? When I think of "the bottom line" that students addressed in their stories, I now understand that poorly-constructed multiple-choice questions are found most unacceptable to them because they are unfair. My own university students still dislike them, even though the students through my course in student evaluation are more informed about them. I have found when I use multiple-choice that while most of the bright students get the questions right, there are some bright students who don't. I always reassured them with, "Oh well, that's why we have a variety of ways to assess--those that don't do well on multiple-choice have a chance on essay questions." But why not give students opportunity to do well on everything? Why does the format have to be the obstacle, or the leveler? What are we testing--the ability to do multiple-choice or the ability to solve problems? Some of my most insightful students were tripped up on multiple-choice questions last semester and I do believe that I constructed the questions reasonably well. That seemed unfair for the more deserving students. What does concern me though is how much these are used at university and from what I have seen, for the most part, they are poorly constructed and improperly analyzed. How can I resolve this dilemma? Perhaps multiple-choice tests are best left to the external tests where the "snapshot" of students' knowledge is captured and

where the results do not count in the students' record. For myself, I will reduce the use of multiple-choice questions in my tests considerably.

How I Will Change My Instruction

I will address what constitutes authentic instruction first in my workshops and university classes and follow naturally with what constitutes good assessment. I will not assume instant understanding of good instruction as I once did. I believe there really is no difference between instructional activities and assessment activities and I will link the two even more than I have in the past.

Practice That I Will Continue

To do that I will continue to use context-rich situated learning activities as I have done in the past. Not only are they more interesting and effective, but students find this a refreshing change from the usual lecture-style instruction. I will extend the idea of the Application Assignment and provide even more choice to students in the ways they would like to be assessed. I will give more responsibility back to the students and trust that they know how they learn and can demonstrate their learning best.

Assessment As a Positive Step

I will continue to view assessment as a positive activity that is integrated into instruction. Zessoules and Gardner (1991) hold that authentic assessment will have to have meaningful and engaging instruction in place first and that it is no longer tacked on the end of a lesson or unit of instruction:

> [A]uthentic assessment involves a complicated re-evaluation of classroom activities and responsibilities, transforming the classroom along many dimensions; changing the kinds of activities students engage in on a daily basis; altering the responsibilities of students and teachers in increasingly sophisticated ways; and transforming the static, mechanical, and disengaging moments when learning stops and testing begins into a continuum of moments that combine assessment, instruction, and learning. By integrating assessment into the day-to-day classroom experience, one changes its role dynamically. No longer a weapon for rooting out and combating students' weaknesses. (p. 63)

Assessment is a celebration and not a means or "weapon for rooting out and combating students' weaknesses" as Zessoules and Gardner (1991) suggest is currently being done. Rather than a punishment or something to be feared as the students in their stories indicated, assessment becomes an additional way for students to learn and set goals for further learning. Ultimately, assessment is a way to discover the strengths, possibilities and potential of the child. It is a way for students to truly move forward and celebrate their being and becoming.

Significant Learning From the Research: Creating A Space for Learning and Assessment

To authentically instruct and assess students, teachers must first create a space for them. I think that that is one of the most significant things I have learned through this research. It was difficult to listen to the students--I kept wanting to speak. I finally let go of the control I had of their voices and I listened to them and reflected. I was apprehensive about doing that. I was worried about the shape that the research would take. I took a risk and let the students share the lead with me, and I learned more than I ever thought possible. Students reaffirmed much of what I already knew about education, but I learned new things too. I remembered reading Gadamer's suggestion to open the conversation and listen for the possibilities. In learning and assessment I have discovered some of those possibilities by listening to the students, just as Ashton-Warner (1956) did. She listened to the students and gave up the teacher power, empowered them to learn in their context--the real-world context that made sense to them and to her. By doing that she could watch them "develop like the flowers that are interested in the rain and the sun; in their own time and way." My last story is an excerpt from her novel, Spinster:

> Today I work on Rangi, a five-year-old Maori. Nothing will make him learn the first words of the imported books. Yet they seem normal enough words. "Come and look." "See the boats." "Little dog." "See my aeroplane." Words that I had taken for granted as having been chosen by adult educationalists for their emotional significance. Indeed, the glamorous visitor from the department claimed that they were the mean and it's all but impossible for a teacher to contest the rightness of anything from the department.
> But Wiki and Rangi and others like them, sit and smile and never recognize them again. All this toil, I think, trying to force them to like something they hate. Why must we? Why don't I teach them something that does interest them? Then they might develop like the flowers that are interested in the rain and the sun; in their own time and way. What does interest them?
> "What's Rangi's background?" I ask the Head.
> "His father is a pugilist who runs a gambling den at the pub."
> "What are you frightened of, Rangi?" I ask as he sits in a knot of others.
> "P'lice."
> "Why?"
> "P'lice they takes me to gaol and cuts me up with a butcher-knife."
> I print these words on separate cards and give them to him. And Rangi, who lives on love and kisses and thrashings and fights and fear of the police and who took four months to learn "come," "look," "and" takes four minutes to learn: butcher-knife; gaol; police; sing; cry ; kiss; Daddy; Mummie; Rangi; haka; fight.
> So I make a reading card for him: out of these words, which he reads at first sight, his first reading, and his face lights up with understanding. And from here he goes on to other reading, even the imported books. His

mind is unlocked, some great fear is discharged, he understands at last and he can read. (Ashton-Warner, 1956, pp. 176-177)

Ashton-Warner in taking Rangi's circumstances improvised and created a space to allow Rangi's difference show. Immediately, instead of being stifled, he bloomed like the flowers. Musicians improvise also--music is more than a matter of skills and techniques. Aoki (1991) says that for music to be lived, it must call for transformation of instrument and music into that which is lived bodily. It is beyond the instrumental reasoning as it is in curriculum where we have performative words such as goals and objectives, processes and products, achievement and assessment (Apple, 1979). Real teaching and learning is real world. Letting learn instead of insisting on the way to teach and assess, instead a way to teach, learn, and assess students. Curriculum improvisation instead of implementation then leaves space to breathe and grow. Letting go means becoming attuned to the students' voices and genuinely addressing their ideas, concerns, and views on learning. By letting go we come nearer to the truth: "The essence of truth reveals itself in freedom" (Heidegger, 1977, p. 130).

Journal Entry, May 21, 1991:

The text of "The Art of the Fugue" appears structured, static, directed, (riddled with dynamics and the like: crescendos, diminuendos, fortes, and allegrettos, and so on)--and so it should be. These "suggestions" or "guidelines," that are written in the text are for poetic license only--not really considered "right" or "wrong," although some might question that I suppose, and hold that some are "more right" or "more wrong" than others, perhaps. Whether to recognize the suggestions or not depends on the ear of the Other, the performer, and the ear of the Other, the auditor. The eye of the performer sees the same text as Others as performers, but "between" the eye and the ear, the magic of interpretation occurs. This is where the freedom of choice occurs, choice of decisions based on past experience, the mood of the moment, and of course in performance, the mood of the Other, the audi-ence.

And so it is with our instruction and assessment. This is our text of music. It too, is riddled with "suggestions" or "guidelines", some of this here, a little less of this there; in other words, poetic license, or freedom to act. The possibilities are infinite. But what of this license? Is this the license (freedom) to teach or improvise with license? The more experienced the musician, the more appropriate license she probably will take. The more license taken, the more likelihood of a sound improvisation; an improvisation of the text that brings the musician's audi-ence to the point of the highest auditory participation. This participation is as alike as it is different, because for each auditor there is a different sound, a sound as heard, and a sound as interpreted--a sound to form a unique connection. Like the young child reading a book, the text is read, listened to, and connected to a new lived-experience from a unique background of lived-experience. So it is with the reader of a music experience living her own improvisation.

Improvisation is living discovery, surprise, creativity, excitement, joy, and freedom. Instruction and assessment beg for the improvisation that the great works of the masters enjoy. Teachers are licensed to improvise instruction and assessment, so . . . .

Let the concert begin!

And we, as the audience, will assess the concert in real-world context, as it is happening. That is real assessment--authentic assessment.

I am an early childhood education teacher who has worked for the Achievement Testing Program. Some of my colleagues wondered how I resolved what they saw as a conflict of philosophies between the two, but I did not see it that way. Assessment was important to me as a pedagogue. It was the "icing on the cake" during, and particularly at the end of the day with my class of grade ones. At the end of the day we would assess it by sharing what we did and learned that day. For the children it was a celebration and a time to set goals for follow-up learning. We celebrated what we did that day and reflected on how we might improve. It was a self-assessment and peer assessment in one. Oftentimes I was assessed too! The child who said "Boring!" during one of the lessons will always stand out in my mind--particularly because he was right! (He was working at a rather meaningless activity, quite inappropriate for his ability and interest--I should have known better.) We came to terms with our strengths and weaknesses, and our attitude was that no-one is perfect but we can always try the best we can. We didn't put a numerical value on the day, or grade it in any way, although I can hear some students say, "Today was a '10'!"; or "Not for me--today was just a '5.'" Good or bad, we learned something and we went on to a whole new day, with new goals and fresh expectations.

The cartoon in Figure 15 illustrates that particularly strengths need to be recognized and how sad it is that usually only the weaknesses are. It is unfortunate that teachers, for many reasons, some of which are beyond their control, are too busy covering the curriculum to relate to students and celebrate their strengths.

Figure 15

Cartoon: The Importance of Celebrating Success

I recognize, however, that assessment is not one that focuses on ; the students spoke very loudly about it, but I do think that it is mostly a matter of attitude. If we take a positive attitude toward assessment, to evaluate what we value, to learn and grow, then assessment or testing will itself flourish--but in a pedagogical sense. Not just as we have done so narrowly in the past rooting out winners and losers, but in a genuine way. Sometimes for bureaucratic or instrumental reasons we do need to tack a number on growth and achievement. If this is done as fairly as possible however, students least of all disagree with that. Of course the key word here is fairly--students said time and time again, that yes, tests were necessary and they did not quarrel with that. What students opposed most of all were unfair tests. By abiding by

principles of fair testing practice, we will be providing appropriate pedagogic practice to students.

I am now at the end of my "hodos," or journey in my pursuit for knowledge. And what did I learn from it? My search to make sense of the meanings and experiences of students taking tests, the looking back, the looking ahead in the journey, and dwelling in the present, has helped me gain insights to "come up with something" as Pirsig (1974) discovers:

> You look at where you're going and where you are and it never makes sense, but then you look at where you've been and a pattern seems to emerge. And if you project forward from that pattern, then sometimes you can come up with something. (p. 149)

My last question: "Did I come up with something in my journey?" My orientation is pedagogic. If this research into the students' experiences of taking tests has in some small way made it a better place in assessment for children, then I have indeed achieved my purpose.

LAST WORDS

What is the teacher?
A guide, not a guard.
What is Learning?
A journey, not a destination.
What is discovery?
Questioning the answers,
not answering the questions.
What is the process?
Discovering ideas,
not covering content.
What is the goal?
Open Minds, not closed issues.
What is the test:
Being and becoming,
not remembering and reviewing
What is the school?
Whatever We choose to make it.

Alan A. Glatthorn

REFERENCES

Adams, W. A. (1980). The experience of teaching and learning. Seattle, WA: Psychological Press.

Aitken, E. N. (1987). A Comparative Study in the Learning of Basic Mathematical Skills in Grade One Using CAI and Traditional Teaching. Unpublished master's thesis, San Diego State University, CA.

Aiken, L. R. (1991). Detecting, understanding, and controlling for cheating on tests. Research In Higher Education, New York: Human Sciences Press, 32(6), 725-736.

Aksamit, D. L., (1990). Naturalistic inquiry applied to the evaluation of a teacher education program. Teaching and Teacher Education, 6(3), 215-226.

Albas, D. C., & Albas, C. M. (1984). Student life and exams: Stresses and coping strategies. Winnipeg, MB: Kendall/Hunt.

Alberta Education. (1990) Program continuity: Elementary education in action. Curriculum Branch. Edmonton, AB: Author.

Alberta Education. (1991). Vision for the nineties. Edmonton, AB: Author.

Alberta Education. (1992). Grade 9 Highlights. Student Evaluation Branch, Edmonton, AB: Author.

Alberta Education. (1993a). Achievement testing program provincial report: June 1992 administration. Student Evaluation Branch, Edmonton, AB: Author.

Alberta Education. (1993b). 1993-94 General information bulletin: Diploma examinations program. Student Evaluation, Edmonton, AB: Author.

Alberta Teachers' Association. (1993). Trying to teach. Interim Report of the Committee on Public Education and Professional Practice as approved by Provincial Executive Council for discussion at the 1993 Annual Representative Assembly. Edmonton, AB: Author.

Alberta Teachers' Association. (1994). Trying to teach: Necessary conditions. Report of the Committee on Public Education and Professional Practice. Edmonton, AB: Author.

Alapack, R. J. (1986). The outlaw relationship as an intertwining of two identity crises: A phenomenological/psychotherapeutic reflection upon female awakening at late adolescence and male rejuvenation at mid-life. Journal of Phenomenological Psychology, 17, 43-63.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards of educational and psychological testing. Washington, DC: American Psychological Association.

Aoki, T. T., Langford, D., Williams, D. M., & Wilson, D. C. (Eds.). (1977b). The British Columbia social studies assessment summary report: A report to the Ministry of Education. Victoria, BC: Ministry of Education.

Aoki, T. T. (1978, April). Toward curriculum inquiry in a new key. Paper presented at the Conference on Phenomenological Description: Potential for Research in Art Education Sponsored by the Division of Graduate Studies in the Fine Arts, Concordia University, QU.

Aoki, T. T. (1990). Inspiriting curriculum and pedagogy: Talks to teachers. University of Alberta, Edmonton, AB: Curriculum Praxis.

Apple, M. (1979). Ideology and curriculum. London: Routledge & Kegan Paul.

Armstrong, D. (1991). A celebration of learning: Assessing, evaluation and communicating student growth. Issues, Events, and Ideas: Early Childhood Council of the Alberta Teachers' Association, (62), 7-8.

Arter, J. A., & Spandel, V. (1991). Using portfolios of student work in instruction and assessment. Northwest Regional Educational Laboratory, pp. 1-21.

Ashton-Warner, S. (1958). Spinster. New York: Simon & Schuster.

Bachor, D. G., & Crealock, C. (1986). Instructional strategies for students with special needs. Scarborough, ON: Prentice-Hall.

Badger, E. (1992). More than testing. Arithmetic Teacher, 39(9), 7-11.

Barritt, L., Beekman, T., Bleeker, H., & Mulderij, K. (1983). The world through children's eyes: Hide and seek and peekaboo. Phenomenology + Pedagogy, 1(2), 140-161.

Barritt, L., Beekman, T., Bleeker, H., & Mulderij, K. (1984). Analyzing phenomenological descriptions. Phenomenology + Pedagogy, 2(1), Edmonton: University of Alberta.

Barth, R. S. (1990). Improving schools from within. San Francisco: Jossey-Bass.

Becker, C. S. (1986). Interviewing in human science research. Methods, 1, 101-124.

Bell, S. (1991). Appropriate assessment in the early childhood classroom. Issues, Events, and Ideas: Early Childhood Council of the Alberta Teachers' Association, (62), 8-9.

Bergum, V. (1989). Woman to mother: A transformation. Massachusetts: Bergin & Garvey.

Berlak, H., Newmann, F. M., Adams, E., Archbald, D. A., Burgess, T., Raven, J., & Romberg, T. A. (1992). Toward a new science of educational testing and assessment. New York: SUNY.

Berman, L. M., Hultgren, F. H., Lee, D., Rivkin, M. S., & Roderick, K. A. in Conversation with Ted Aoki. (1991). Toward curriculum for being: Voices of education. Albany, NY: SUNY Press.

Blishen, E. (Ed.). (1969). The school that I'd like. Harmondsworth, Middlesex, England: Penguin.

Bollnow, O. F. (1960). Lived-space. Philosophy Today, 15(4), 31-39.

Bollnow, O. F. (1989). The pedagogical atmosphere. Phenomenology + Pedagogy. Edmonton: University of Alberta, 7, 5-63.

Borasi, R. (1994). Capitalizing on errors as "Springboards For Inquiry": A teaching experiment. Journal for Research in Mathematics Education, 25(2), 166-208.

Brandt, R. (1989). On misuse of testing: A conversation with George Madaus. Educational Leadership, 46(7), 26-29.

Brandt, R. (1992). On performance assessment: A conversation with Grant Wiggins. Educational Leadership, 49(8), 35.

Brandt, R. (1993). On teaching for understanding: A conversation with Howard Gardner. Educational Leadership, 50(7), 4-7.

Bredekamp, S., & Shepard, L. (1989). How best to protect children from inappropriate school expectations, practices, and policies. Young Children, 44(3), 14-24.

Brimfield, R., Roderick, J., & Yamamoto, K. (1983). Persons as researchers: Observation of the participation. Curriculum Inquiry, 13(4), 5-12.

British Columbia Ministry of Education. (1994, July). News release: New reporting standards for students finalized. (NR2994, p. 30). Victoria, BC: Author.

Broadfoot, P., & Fenner, R. (1985). Pupil profiles: The promise and the peril. The New Era, 66(2), 30-39.

Broadfoot, P. (1986). Profiling and the affective curriculum. Journal of Curriculum Studies, 19(1), 25-34.

Bruinsma, R. (1992, March/April). Teaching, testing and evaluation. The Alberta Teachers' Association Magazine, 72(3), 8-10.

Burgess, R. G. (Ed.). (1984). Field methods in the study of education. Philadelphia: Falmer Press.

Butt, R. L., Raymond, D., & Yamagishi, L. (1988). Autobiographic praxis: Studying the formation of teachers' knowledge. Journal of Curriculum Theorizing, 7(4), 87-164.

Cannell, J. J. (1987). Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average. Daniels, W. V.: Friends for Education.

Carl, I. M. (1991). Better mathematics for K8. Streamlined seminar. National Association of Elementary School Principals, 9(4), 2.

"Children's Voices on Provincial Testing." (1993). Issues, Events, and Ideas, Early Childhood Council of the Alberta Teachers' Association, (70), 10.

Clarke, D. J. (1992). Activating assessment alternatives in mathematics. Arithmetic Teacher, 39(6) 24-29.

Clements, D. H., & Battista, M. T. (1990). Constructivist learning and teaching. Arithmetic Teacher, 38(1), 34-35.

Clemmer, J. (1994, April 19). Brain storming: Measurement traps. The Globe and Mail, p. B26.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. Educational Researcher, 19(5) 2-14.

Cornett, J. W. (1990). Utilizing action research in graduate curriculum courses. Theory Into Practice, 29(3), 185-195.

Craig, C., (1993). Children's questions about provincial testing. Issues, Events, and Ideas, Early Childhood Council of the Alberta Teachers' Association, (70), 9.

Craig, T. (1983). Perspectives: Self-discovery through writing personal journals. Language Arts, 60(3), 373-379.

Davies, A., Cameron, C., Politano, C., & Gregory, K. (1992). Together is better: Collaborative assessment, evaluation, and reporting. Winnipeg, MB: Peguis.

Davies, A., Paul, B., Schroeder, D., & Scott, D. (1993). The writer's story. Early Childhood Education, 26(2) 1-42.

Denton, D. E. (1979). Concepts and strategies of phenomenological research. Unpublished manuscript, University of Kentucky.

Depressed over grades. (1993, April 1). Lethbridge Herald, 90(110), p. B7.

Doll, W. E. (1989). Foundations for a post-modern curriculum. Journal of Curriculum Studies, 21(3), 243-253.

Duffy, T. M., & Bednar, A. K. (1991). Attempting to come to grips with alternative perspectives. Educational Technology, 31(9), 12-15.

Edwards, J., & Craig, T. (1990). A teacher experiments with drama as a teaching tool: A collaborative research project. The Alberta Journal of Educational Research, 36(4), 337-351.

Ellis, J. (1993). Stories of women who don't write. McGill Journal of Education, 28(3), 455-468.

Eisner, E. W. (1985). The art of educational evaluation. Philadelphia, PA: Falmer Press.

Etobicoke Board of Education. (1987). Making the grade. Scarborough, ON: Prentice-Hall.

Fairhurst, D. (1993). Achievement tests and dissonance. Issues, Events, and Ideas, Early Childhood Council of the Alberta Teachers' Association, (70), 8.

Feuer, M. J., Fulton, K, & Morison, P. (1993). Better tests and testing practices: Options for policy makers. Phi Delta Kappan, 74(7),. 530-533.

Fiske, E. B. (1988, February, 17). Standardized test scores: Voodoo statistics? New York Times, pp. 7, 9.

Fiske, E. B. (1988, April). America's test mania, Alexandria, VA: ASCD.

Flexner, S. B., & Hauck, L. C. (Eds.). (1987). The Random House dictionary of the English language (2nd ed.). Toronto: Random House.

Floyd, P. (1979). Pink Floyd the wall. Don Mills, ON: Pink Floyd Music Publisher (BMI) CBS Columbia Records Canada. Columbia No. AL 36184

Foucault, M. (1979). Discipline and punish. (A. Sheridan, Trans.), New York: Vintage Books.

Fullan, M. G. (1991). The new meaning of educational change. New York: Teachers College Press.

Fullan, M. G. (1994). Coordinating top-down and bottom-up strategies for educational reform. The Governance of Curriculum: The 1994 ASCD Yearbook, Alexandria, VA: ASCD.

Fulwiler, T. (1982). The personal connection: Journal writing across the curriculum. A. Young & T. Fulwiler (Eds.), Reality Across the Curriculum, Urbana, IL: National Council of Teachers of English.

Fulwiler, T. (Ed.). (1987). The journal book. Portsmouth, NH: Boynton/Cook.

Gadamer, H. E. (1989). Truth and method. (2nd rev. ed.). New York: Crossroad.

Gardner, H. (1991). The unschooled mind: How children think and how schools should teach. New York: Basic Books.

Gitlan, A. D. (1990). Educative research, voice, and school change. Harvard Educational Review, 60(4), 445-466.

Gordon, T. (1974). Teacher effectiveness training. New York: Peter H. Wyden.

Grennon Brooks, J., & Brooks, M. (1993). In search of understanding: The case for constructivist classrooms. Alexandria, VA: ASCD.

Guba, E. G. (1981). Criteria for assessing trustworthiness of naturalistic inquiries. ERIC/ECTJ Annual Review Paper, Educational Research Information Center, 29(2), 75-91, Syracuse, NY: Eric Clearing House on Information Resources.

Hargreaves, A. (1989). Assessment and surveillance. Curriculum and Assessment Reform, Toronto: OISE Press.

Hawking, S. W. (1988). A brief history of time. NY: Bantam Books.

Hebert, E. A. (1992). Portfolios invite reflection--from students and staff. Educational Leadership, 49(8), 58-61.

Heidegger, M. (1977). Basic writings. San Francisco: Harper.

Hoffmann, B. (1964). The tyranny of testing. New York: Collier Books.

Husserl, E. (1991). Cartesian meditations. London: Kluwer Academic.

Inmac. (1993, Mid-winter). MacSelect. pp. 1-66.

Janesick, V. J. (1981). Using a journal to develop reflection and evaluation options in the classroom. Paper presented for the conference on "Reading in the Classroom: New Ways of Looking," Albany, NY.

Jonassen, D. H. (1991a). Evaluating constructivistic learning. Educational Technology, 31(9), 28-33.

Jonassen, D. H. (1991b). Objectivism versus constructivism: Do we need a new philosophical paradigm? Educational Technology and Development, 39(3), 5-14.

Kaiser, R. B. (1981, March). The way of the journal. Psychology Today, pp. 64-76.

Kamii, C., & Lewis, B. A. (1990). Constructivism and first-grade arithmetic. Arithmetic Teacher, 38(1), 34-35.

Kansanen, P. (1991). Pedagogical thinking: The basic problem of teacher education. European Journal of Education, 26(3), 251-260.

Katz, L. G. & Chard, S. C. (1989). Engaging children's minds. Norwood, NJ: Ablex.

Kermode, F. (1979). The genesis of secrecy on the interpretation of narrative. London: Harvard University Press.

King, M., & Ranallo, J. (Eds.). (1993). Teaching and assessment strategies for the transition age. Vancouver: EduServ.

Kite, J. (1991), January). Dear diary. TWA Ambassador, pp. 60-63.

Kvale, S. (1983). The qualitative research interview: A phenomenological and a hermeneutical mode of understanding. Journal of Phenomenological Psychology, 14(2), 171-196.

Langeveld, M. J. (1983). The secret place in the life of the child. Phenomenology + Pedagogy, 1(2), 181-191.

Langley, R. (1983). Keeping an intensive journal: A guide to tapping your hidden inner potential. People Who Motivate Others, 2(3), 1-4.

Le Box, A. (1991). Standardized testing for standardized grade ones? T. T. Aoki (Ed.), (Voices of Teaching, Monograph #2), pp. 26-31, Vancouver, BC: British Columbia Teachers' Federation Program for Quality Teaching.

Lincoln, Y. S. (1988). The role of ideology in naturalistic research. Paper prepared for Presentation at the Invited Symposium, "Ideology in Qualitative Research Methodologies," Division D and SIG/Qualitative Research, American Educational Research Association, New Orleans, LA, April 5-9, 1988.

Linden, J. D., & Linden, K. W. (1968). Tests on trial. Boston: Houghton Mifflin.

Livingstone, C., Castle, S., & Nations, J. (1989). Testing and curriculum reform: One school's experience. Educational Leadership, 46(7), 23-25.

Maeroff, G. I. (1991). Assessing alternative assessment. Phi Delta Kappan, 73(4), 273-281.

Majesky, D. (1993). Grading should go. Educational Leadership, 50(7), 88, 90.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). Assessing student outcomes: Performance assessment using the dimensions of learning model. Alexandria, VA: ASCD.

McGuire, W. L. Approaching "D-Day": Experiencing pedagogical suffocation. In T. T. Aoki (Ed.), Voices of Teaching, Monograph #2 (pp. 32-34). Vancouver, BC: British Columbia Teachers' Federation Program for Quality Teaching.

McLaughlin, M. W. (1991). Test-based accountability as a reform strategy. Phi Delta Kappan, 73(3), 248-251.

Meisels, S. J. (1989). High-stakes testing in kindergarten. Educational Leadership, 46(7), 16-22.

Merleau-Ponty, M. (1962). Phenomenology of perception. New York: The Humanities Press.

Meyer, C. A. (1992). What's the difference between authentic and performance assessment? Educational Leadership, 49(8), 39-40.

Morris, W., (Ed.). (1980). Houghton Mifflin Canadian dictionary of the English language. Senior Edition. Markham, ON: Author.

Mouly, G. J. (1978). Educational research. Boston, MA: Allyn & Bacon.

Musca, T. (Producer), & Menendez, R. (Director). (1988). Stand and deliver [Film]. Los Angeles, CA: Warner Bros.

NAEYC Position Statement on Standardized Testing of Young Children 3 Through 8 Years of Age. (1988). Young Children, 43(3), 42-47.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

Noddings, N. (1991). Caring and continuity in education. Scandinavian Journal of Educational Research, 35(1), 3-12.

O' Neil, J. (1992). Putting performance assessment to the test. Educational Leadership, 49(8), 14-19.

Osborne, J. W. (1990). Some basic existential-phenomenological research methodology for counsellors. Canadian Journal of Counselling, 24(2), 79-91.

Oxford English Dictionary. (1933). London: Oxford University Press. Author.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. Educational Researcher, 20(5), 12-20.

Perkins, D. (1992). Smart schools. New York: The Free Press.

Perrone, V. (Ed.). (1991). Expanding student assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Perrone, V. (1991). On standardized testing. Childhood Education, 67(3), 132-142.

"Phase out standardized tests, AFT says." (1990, July 12). Education USA Newsline. Online Source, pp. 10-11.

Pike, G. (1991). Reflections of a failing grade. In M. van Manen (Ed.), Texts of pedagogy, University of Alberta: Human Science in Education Project.

Pinar, W. F. & Reynolds, W. M. (Eds.). (1992). Understanding curriculum as phenomenological and deconstructed text. New York: Teachers College Press.

Pirsig, R. M. (1974). Zen and the art of motorcycle maintenance. New York: Bantam.

Polkinghorne, D. (1983). Methodology for the human sciences. Albany, N.Y: State University of New York Press.

Popham, W. J. (1981). Modern educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Progoff, I. (1975). At a journal workshop. New York: Dialogue House Library.

Rabinow, P. (Ed.). (1984). The Foucault reader. New York: Pantheon Books.

Reigeluth, C. M. (1991). Reflections on the implications of constructivism for educational technology. Educational Technology, 31(9), 34-37.

Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? The Alberta Journal of Educational Research, 37(2), 179-192.

Sartre, J. P. (1957). Being and nothingness. London: Methuen.

Sharma, J. K. (1985). Time and T. S. Eliot. New York: Apt Books.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46(7), 4-9.

Shepard, L. A. (1991). Will national tests improve student learning? Phi Delta Kappan, 73(3), 232-238.

Sherman, A. (1990). Stop the Grade 3 Provincial tests! Issues, Events, and Ideas, Early Childhood Council of the Alberta Teachers' Association, (57), 1-2.

Sherman, A. (1991). Message from the president. Issues, Events, and Ideas, Early Childhood Council of the Alberta Teachers' Association, (61), 2-3.

Silverman, F. L., Winograd, K., & Strohauer, D. (1992). Student-generated story problems. Arithmetic Teacher, 39(8), 6-12.

Smith, D. G. (1988). Children and the gods of war. Phenomenology + Pedagogy, 6(1), 25-29.

Smith, D. G. (1991). Hermeneutic Inquiry: The hermeneutic imagination and the pedagogic text. In Forms of Curriculum Inquiry, Edmund C. Short (Ed.), (1991), pp 187-209.

Smith, F. (1986). Insult to intelligence. New York: Arbor House.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20(5), 8-11.

Smith, M. S., Fuhrman, S. H., & O'Day, J. (1994). National curriculum standards: Are they desirable and feasible? The Governance of Curriculum, Alexandria, VA: ASCD, 12-29.

Spiegelberg, H. (1975). Doing phenomenology. The Hague, Netherlands: Martinus Nijhoff.

Stepien, W., & Gallagher, S. (1993). Problem-based learning: As authentic as it gets. Educational Leadership, 50(7), 25-28.

Tan, A. (1989). The joy luck club. Toronto: Random House.

Valle, R. S., & King, M. (Eds.). (1978). Existential-Phenomenological alternatives for psychology. New York: Oxford University Press.

van den Berg, J. H. (1970). Time. Things: Four Metabletic Reflections, Pittsburgh, PA: Duquesne University Press.

van Manen, M. (1984). Practicing phenomenological writing. Phenomenology + Pedagogy, 2(1), 36-69.

van Manen, M. (1986). The tone of teaching. ON: Scholastic-TAB.

van Manen, M. (1990). Researching lived experience: Human science for an action sensitive pedagogy. ON: The Althouse Press.

van Manen, M. (1991). The tact of teaching: The meaning of pedagogical thoughtfulness. Albany, NY: The Althouse Press.

van Manen, M. (1992, March). On pedagogy as virtue and the question of the education of teacher educators. Text of a Lecture/Discussion on the occasion of the Faculty of Education 50th anniversary Lecture Series, University of Alberta, Edmonton, AB.

van Manen, M. (1993). Pedagogical politics? Political pedagogy? Keynote invited address at the International Human Science Research Conference, University of Groningen, August 1993, pp. 1-25.

Vygotsky, L. S. (1978). Mind in society. London: Harvard University Press.

Warmington, E. H., & Rouse, P. G. (1956). Great dialogues of Plato. Toronto: The New English Library.

Wertz, F. J. (1984). Procedures in phenomenological research and the question of validity. In C. M. Aanstoos (Ed.). Exploring the lived world, West Georgia College Studies in the Social Sciences, 23, 29-48.

Wiggins, G. P. (1993a). Assessing student performance: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass.

Wiggins, G. P. (1993b). Assessment: Authenticity, context, and validity. Phi Delta Kappan, 75(3), 200-214.

Willis, S. (1990, September). Transforming the test. ASCD Update. Alexandria, VA: ASCD, 32(7), 3-6.

Willis, S. (1993). Are letter grades obsolete? ASCD Update, Alexandria VA: ASCD, 35(7), 1, 4 , & 8.

Witte, D. L., & Blakey, J. M. (1991). The child's view of evaluation: Voice as pattern through time. The Journal of Learning About Learning, 1(1), 72-79.

Yinger, R. (1985). Journal writing as a learning tool. Volta Review, 8(1), 21-33.

Zachariah, M. & O'Neill, L. (1986). The re-introduction of public examinations in Alberta: School context, reasons, implications. The Canadian Public Education System, Calgary, AB: Detselig Enterprises.

Zessoules, R., & Gardner, H. (1991). Authentic assessment: Beyond the buzzword and into the classroom. In V. Perrone, (Ed.), Expanding Student Assessment, VA: ASCD.

APPENDIX A

Notes

"Authentic Assessment"

Some teachers take exception to the term "authentic assessment." They have asked, "If this is 'authentic assessment, is all other 'inauthentic' then?" To that I would say yes and no. Certainly a lot of assessment that is occurring is authentic, but so has a lot been inauthentic--the kinds of mindless activity that has been passed off as a test is deplorable. The students have described it well. The trivia, the meaningless questions, and the irrelevant. I believe most of us are familiar with this--most of us, like the students, have experienced these kinds of tests that don't authentically test knowledge and understanding at all. These tests are written by burnt-out teachers, by teachers out of their league in the subject matter, by teachers who are overloaded with other work, by teachers who have retired before their time and just putting in time, by teachers who are 'winging it," by teachers who are not dedicated or committed, and by teachers who just don't know any better--they were brought up on a steady diet of inauthentic tests themselves. most likely, and so the cycle continues.

Test, Assessment, and Evaluation

What are the differences among test, assessment, and evaluation? In everyday conversation we consistently find these three terms used interchangeably. For the purpose of this investigation, they will be used the following way.: The term "test" will be referred to as the measuring tool or instrument, much like a thermometer or scale; "assessment" will be understood as the process of assessing students' knowledge; and finally, the judgment made from the collected assessments will be the "evaluation." Therefore, we could use Badger's (1992) definition: "[W]e use tests in the process of assessing students' knowledge to make judgments, or evaluations" (p. 7). Chittenden (Perrone, 1991) agrees that assessment is a process and is distinguished from testing:

> Assessment, as opposed to simple one-dimensional measurement, is frequently described as multitrait-multimethod; that is, it focuses upon a number of variables judged to be important, and utilizes a number of techniques to assay them. . . . Its techniques may also be multisource. . . and/or multijudge (p. 27). (p. 24)

Chittenden, on the other hand, refers to tests as: ". . . the full range of devices developed commercially or by teachers, for checking up on student learning. In this he includes informal reading inventories, end-of-unit tasks, teachers' quizzes, etc." (p. 25).

Assessment Is further illustrated by Chittenden cited in Perrone (1991):

> In its derivation, the word assess means "to sit beside," to "assist the judge." It refers to a process of collecting and organizing information or

> data in ways that make it possible for people--teachers, parents, students--to "judge"' or evaluate. (p. 25)

It therefore seems appropriate . . . "to limit the term assessment to the process of gathering the data and fashioning them into an interpretable form; judgments can then be made. . . . Assessment, then, as we define it, precedes the final decision-making stage in evaluation" (Anderson et al., 1975, p. 27) (Perrone 1991, p. 26).

In summary, we can view "test" as simply the measuring instrument such as a quiz that is given to students; "assessment" as the process of collecting and sorting of the data resulting from the use of the measuring instruments; and finally, the "evaluation" as the last step, or the judgment of the outcome of the analysis and interpretation of the data gathered in the assessment.

APPENDIX B

Sample of Data Recording and Organizing of Students and Topics.
(Student in italics are elementary school students)

UNPREPARED FOR TEST

didn't/forgot to study	Oscar; Elmo; Lyn; Sandy
forgot about test	Jean; Josephine
forgot to take home books	Moe
thought s/he knew content already	Jean

REASONS TO STUDY

study to ace the test/honor roll	Nancy; Laura
avoid powerlessness	Saphire
study to be exempted from finals--need 85%	Marie
study so parents won't get mad at them	Sandy
rewards from parents: smiles; material things, money	Elmo; *Michael; Blaine: Karen; Jerry;* Jessie
expectations of achievement and results not correlating	Kathleen; Lee la; Ann; *Joe;* Jessie; Brad
expectations of test content and test not correlating	Hercules Wes; Suzy Q
didn't prepare but did OK/well prepared but did poorly	Scott; Doug; Dana; Josephine; Bill
blanked out	Linda; Tonto; Jean; Michelle; Martha; Ann; Murphy: *Nicole;* Lyn; Sara; Harry; Rob; Shianne; Rookie; Paul; Veto; Columbo
worried about parents' reaction	Linda; TJ; Shainne; James; Tonto; Terri; Jerry; Ellen; Matt
putting on a front	Tonto; Pat

EMOTIONAL EFFECTS

tension	Ben
angry	Allen
scared	Joe; James
nervous	Jayne; Elizabeth; Valerie; Jenny Lee; Moe; *Nicole: Ellen; Karen;* Catalina; Rookie; Ray; Dran; Paul; Brad; Ice; Veto; Bobby; Matt; Columbo
feel stupid	Saphire
memorization exercise	Kathleen; James; Darlene; Michael; Murphy; Bobby; Sara; Catalina; Ben

PHYSICAL EFFECTS

headache	Jayne; Lisa
feel sick to stomach	Lee; Jayne; Petrovitch; Shainne; Kennedy
weak legs	Jayne; Josephine; James;Robin
sweaty palms	Kathleen; Lee la; Jean; Elizabeth; Michelle; Michaila; *Matt; Wendy; Rob;* Kennedy; Bobby; Ice; Sterine; Paul; Bryce; Robin; Mike; Matt
pounding heart	Kathleen; Michele
hot	Jayne; *Nicol;* Jessie; Rebecca Jane; Major; Ice
cold	*Nicol;* Jessie; Ray; Matt; cold
butterflies	Mark; Jean; TJ; Valerie; Martha; Jim; *Haley; Marie; Karen; Keith; Karen Kay; Matt; Lloyd; Maria; Charlene;' Malissa*
teeth grinding	Lee la
bitten nails	Lee La
dizzy	Saphire
aware of surrounding people	Saphire
frustrated and hot	Brad; Sparky
sick and panicky	Saphire; Robin; Brad; Moses
shaking for lack of control	Saphire
pounding head	Martha; *Nicole; Joe;* Matt
sick--test looks big	Clay; Dibs; Catalina; Kennedy; Matt
sick--tempted to cheat/worried that teacher thinks s/he's cheating	Clay; Sandy; Roberto

TEMPORAL

clock/time is power	Saphire; Lee
time goes fast	Ben

CONCERNS

worried about mark	Leann; Matt
fear of failure	Leann; Andrew; Kennedy
good luck charms/prayers	Valerie; Sandy
exams aren't fair	Saphire

SPATIAL

cold and dark	Verna
hot space	Richard
people staring	Rob
irregular space	Ray
stuffy room	
noise in test space	Hercules; Martha; Michaila; Saphire; Serria; Dawn; Richard
unfamiliar space and relieved test is over	Scott; Kathleen; Lee la; Lee; Jayne

APPENDIX C

Samples of Letters

The
University of
Lethbridge

4401 University Drive
Lethbridge, Alberta, Canada
T1K 3M4
403-329-2251
FAX: (403) 329-2252

FACULTY OF EDUCATION

1993 11 03

Mr. J. McLellan
Superintendent, County of Warner
210 3rd Ave.
Warner, AB
T0K 2L0

Dear Mr. McLellan:

I am embarking on research for my dissertation entitled The Students' Perspective on the Meaning of Taking a Test. At this point in my research I need to interview students as part of my data gathering. May I have permission to approach principals to ask them if they would be interested in participating in this research project? Informally I have spoken to Kim Tsuji and she has indicated that she is interested in having her students participate.

In return for the school's participation I would be very willing to do a half-hour or one-hour workshop on student evaluation for a professional development day or staff meeting.

I am enclosing my application for the Ethics Review which was accepted recently by the review committee at the University of Alberta.

I am currently teaching Evaluation of Student Learning at the University of Lethbridge, so you may contact me there at 329 2429; or at my home at 380 3250 if you are interested in my project.

Yours sincerely,

Nola Aitken

Enclosure
cc: Kim Tsuji Principal Raymond School

The University of Lethbridge

4401 University Drive
Lethbridge, Alberta, Canada
T1K 3M4
403-329-2251
FAX: (403) 329-2252

FACULTY OF EDUCATION

February 15, 1994

Dear Parents,

I am currently researching the nature of testing from the points of view of students. To do this research I need to collect my data by tape-recording interviews of elementary-school children about their experiences of taking a test. If you are willing to allow your child participate in this research, and if your child is also willing, please indicate this by signing the Permission Slip below.

If your child decides not to participate during the research activity, he or she is free to withdraw at any time. There is a possibility also that I may have to have a follow-up conversation with your child at a later date. Again, this is an optional activity for your child.

Yours sincerely,

Nola Aitken

Nola Aitken
Assistant Professor, University of Lethbridge

Permission Slip

My child MAY participate in the study as outlined above.

Parent's Signature----------

My child MAY NOT participate in the study as outline above.

Parent's Signature----------

The
University of
Lethbridge

4401 University Drive
Lethbridge, Alberta, Canada
T1K 3M4
403-329-2251
FAX: (403) 329-2252

FACULTY OF EDUCATION

1994 02 22

Mr. Cal O'Brien
Assistant Principal
Catholic Central High School
405 - 18 St. S.
Lethbridge, Alberta
T1H 3E5

Dear Cal:

Thank you very much for providing me class time with your grade 10 and 11 students so that I could gather data on their experiences of taking a test. The students cooperated extremely well; they took the exercise seriously and in so doing they wrote very thoughtfully about their experiences. I am very delighted with this rich data that will be so very important to my dissertation.

Yours sincerely,

Nola

Nola Aitken

cc: Ed Rocheleau, Principal

Chapter Four

LITERATURE REVIEW

The provision of water supply and sanitation facilities in rural areas of the developing world has been seen as a way of sparkling socio-economic development and improving quality of life. Yet most studies assessing the impacts of water supply and sanitation programmes limit their assessments to well-defined groups of diseases (for example, diarrhoeal and cholera diseases) [Cvjetanovic, 1986], rather than considering the impact of such programmes on a range of indicators essential to societal development. (The problems hindering such an undertaking have already been discussed in Chapter One.)

Although the former approach is the simplest, it should be realized that it only presents partial evidence, while the latter, presenting formidable obstacles in terms of quantification, more truly reflects the overall impact. The objective of this chapter is to review the impact of water supply and sanitation programmes, in order to highlight both the direct and indirect impacts which that such programmes have on the development of rural areas of the developing world. Another focus of the chapter is an examination of evaluations in the rural water sector. The emphasis is on the methods utilized in the studies, the results emanating from using such methods, and their inherent problems. The approach adopted for the study is discussed at the end of the chapter.

Impacts of Rural Water Supply Programmes in Societal Development

Safe water for drinking and domestic use and adequate sanitary services are considered basic human needs in the developing world (ILO, 1977). The underlying assumption is that the provision of the above facilities would greatly improve health, and provide benefits not directly related to health (Saunders and Warford, 1976; Okun, 1988). These benefits and the role they play in the development process in the rural areas of the developing world are addressed below.

Disease Prevention and Improved Nutrition

An initial input in the improvement of water supply and sanitation produces "direct" health benefits by preventing the spread of water borne diseases (Saunders and Warford, 1976). The transmission of dracunculiasis (guinea worm), for example, a delibilitating disease that affects 10 to 15 million people in West and Central Africa and Western India, depends upon direct contact of infected individuals with water used for drinking, generally in shallow ponds or wells where suitable cyclop species are present (Okun, 1988). In a 1982 Workshop, the United States National Research Council concluded that the most effective means of preventing dracunculiasis has been to provide safe water supplies (National Research Council, 1983).

This assertion has been confirmed by the introduction of hand pump-equipped boreholes in Kwara State, Nigeria. In Kwara State, the introduction of improved water supply systems resulted in the complete elimination of guinea worm (Rooy and

Donaldson, 1990). The elimination of the guinea worm improved the health status of the people, and enabled them to engage in economic and social activities, thereby increasing productivity and incomes, hence improving their quality of life.

In regards to diarrhoeal diseases, which are mostly spread through unhygienic food and inappropriate hygienic practices, water supply and sanitation programmes do have substantial impacts on their incidence and morbidity (WHO, 1985). In 1984, Feachem reported that reductions in diarrhoeal incidence attributable to hygiene education components of water and sanitation programmes alone, were between 14% and 48% in Bangladesh, the United States and Guatemala.

Similarly, in their critical assessment of studies dealing with the effects of water supply or excreta disposal on diarrhoea morbidity, Esrey et al. (1985), noted that most studies showed beneficial impact from water supply and sanitation interventions. Based on an analysis of 67 studies, they concluded that 53 of them showed a median reduction of 22%, for water supply or excreta interventions. Their study confirms the views expressed in earlier comprehensive reviews by Saunders and Warford (1976) and McJunkin (1982). Reductions in diarrhoea infection, especially in young children, not only improve their health status and reduce household medical expenses, but also allow the people (especially women) more time for productive work and other pursuits, thus increasing their general well-being. As well, preventing or reducing diarrhoea infection through appropriate hygienic practices, ensures normal growth in children, which in the long-term results in a healthy society.

The prevention of diarrhoeal diseases also improves nutrition, because enteric

infections decrease food intake and increase metabolic losses which in turn produces malabsorption of nutrients (Okun, 1988). The relationship between diarrhoea diseases, other infections and malnutrition has been documented by Schrimshaw et al. (1968), and has been confirmed by other researchers. In a two year study in St. Lucia, the population of each of three villages located in valleys was provided with different water supply and latrine facilities. As a result of the use of larger quantities of water and appropriate human excreta disposal by households, the incidence of diarrhoea and skin diseases decreased and the nutritional status of children improved (Henry, 1981).

Esrey et al. (1985), have also demonstrated the relationship between improved water supply or excreta disposal and nutritional status of children with variations related to different factors. Based on a comprehensive nutrition study conducted by the Institute of Nutrition of Central America and Panama in Guatemala, Torun (1983) concluded that improvements in water supply and sanitation aided and enhanced other measures tending to ameliorate the population's nutritional status.

Safe drinking water and improvements in sanitation are also known to be effective in controlling cholera and ascariasis, and other diseases such as typhoid, shigellosis and other helminthic diseases (Azurin and Alvero, 1974; Esrey et al., 1985; Okun, 1988). Any one of these diseases is likely to be a significant health problem in developing countries suffering from high diarrhoeal disease rates. In a study in the Philippines over a period of four and a half years, Azurin and Alvero (1974), observed that where both water supplies and toilets were provided, the incidence of

cholera was reduced by as much as 76%. In another study in St. Lucia, Henry (1983) reported significant reductions in ascariasis and trichuriasis, along with diarrhoea as a result of improved water supply and sanitation.

Although, it has been argued that water related health improvements are greatest among children who are not members of the labour force (Sork[illegible], 1988), a healthier population can reduce some categories of expenditure currently made for health and medical services. Specifically, fewer funds might be necessary for the eradication programmes (for example, guinea worm and onchocerciasis), hospital and health centre facilities and equipments, physicians and staff, drugs and medicine and transportation for health personnel.

Time Released for Women

Apart from improving the health status of people in rural communities, and its overall impact on community development, the provision of an adequate quantity of safe water, preferably in the home or at least at a reasonable distance from the home results in time savings and releases the burden of water collectors. For many rural people in the developing world, getting water is time consuming and heavy work, taking up to 15% of women's time (Briscoe and Ferranti, 1988). The time saved by not collecting water, frees the water collectors for many rewarding activities, such as child-caring, tending home gardens and animals, and proper food preparation.

Cairncross and Cliff (1987) report that, in a village on the Muenda Plateau

in Mozambique, the average time that women spent collecting water was reduced from 120 to 25 minutes per day, as a result of the introduction of improved water supply systems in the communities. Time savings have been reported in Belu District and in a rural periphery of Chaoz City in the People's Republic of China (Narayan-Parker, 1990; Jong, 1991). In all cases, the time saved enabled women to participate in village committee meetings and primary health care programmes, and in the case of China participated in township-run enterprises. These indicate that the time saved by not collecting water is essential in the development process of rural communities in the developing world. Even if there was no evidence that time saved from no longer carrying water is directly used for productive activities, "saving time is development, for time saved from humdrum tasks is time to invest in human capital" (Birdsall and Greevey, 1978:36).

Community Strengthening

The involvement of the programme beneficiaries in the planning, construction, operation and maintenance as well as the financing of their water supply and sanitation facilities, also has implications for the development of communities. Effective participation of the people in such activities, leads to community cohesion and strengthens the community and community institutions. These characteristics are important in the development process of rural communities in the developing world, since they can be used for other community projects. The lessons learned and experiences gained through their involvement can lead to the development of local

initiatives for other projects. With such experiences, the community can identify and resolve problems, organize themselves for action, and raise funds locally. Thus, the community is empowered through the process.

Experience in Nusa-Tenggara-Timur Province in Indonesia has demonstrated that the participatory experiences of the villagers in rural water projects led to community cohesion, and encouraged them to build three rain water tanks, as well as construct one new household toilet every month (Narayan-Parker, 1990). The water project, in effect, empowered and motivated the people to undertake a project of a completely different nature.

Improved Social Standing of Women

Associated with the participation of the beneficiaries is the enhancement of the status of women as a result of the introduction of water supply programmes. The recognition of women's tasks and organization and training for new tasks (for instance as community water organizers [CWOs] or hand pump caretakers), are not only essential to achieving maximum benefits from improved facilities, but also provide other professional jobs for women. It also improves their social standing, both in the households and within their communities, and this marks an important step in their empowerment.

Improved social standing of women as a result of their participation in water supply and sanitation programmes has long been recognized. However, it has rarely been thoroughly addressed in impact studies, due to problems related to methods of

collecting such data (World Bank, 1987; Stephenson and Peterson, 1991). As such this issue is addressed in this study.

Improved Economic Activity

Improved access to water can generate economic activity through increased livestock, crop production (especially vegetables) and small-scale industries in rural communities. These are likely to provide positive effects on nutrition as well as on income generation, both of which are essential to improved quality of life and especially, the welfare of women, in themselves essential to the development of rural communities. Improved nutrition and income earning result if households use some of the animals and crops in the diets and sell some of the products in the market. Such benefits have been realized from water supply projects in Indonesia, Guinea-Bissau and Togo (Kompaore,1989; INSTRAW, 1990; Narayan-Parker, 1990). In all three cases, there were increases in vegetable production, with resultant improvements in nutrition and increase in incomes, especially for women.

Impact on the Environment

Studies undertaken in different parts of the world have thus revealed the positive impact of water supply and sanitation programmes in the development process of the developing world. However, some negative impact may result, especially as they relate to the environment, thereby hampering the overall development of communities. One such impact relates particularly to the fixed

nature of hand pumps and other water supply systems, and is a serious problem in arid and semi-arid regions of the developing world, where wells attract excessive numbers of humans and livestock.

The fixed nature of hand pumps suggest that increased activity can lead to land degradation, which has implications for both production and groundwater resources, and hence, affects the community in its development efforts. In the Sudan, Carrier (1988) found that the increase in and concentration of human and animal population around hand pumps led to such areas becoming "moderately to severely" desertified. Much of this problem stemmed from concentration of cattle at pump sites, as a result of nomadic herding. Under such circumstances, a programme that was initiated to aid in the development of rural communities had tended to hinder the development process of these areas.

Summary

Investments in water and sanitation services are vital to the development of rural communities in the developing world, since they offer high social and economic returns. The health benefits provided by better water and sanitation services are huge. The World Bank (1993), for instance, estimates that, if all people had adequate water and sanitation facilities, about two million fewer children would die from diarrhoea each year. Improved access to water and sanitation facilities also yields direct economic returns, through time savings and the cultivating of vegetables. These returns promote the health of the general population, encourage the

development of rural industries, increase incomes and enhance the status of women which inevitably promote community development.

Evaluations in the Rural Water Sector

Evaluating the impact of projects/programmes is not a new activity. It emerged as a specialized function after World War II, and therefore has a brief history (Suchman, 1967). Its emergence as a developing speciality in the social sciences was in response to public accountability for the continued support of social programmes (Franklin and Thrasher, 1976).

Project/programme evaluation shares with most specialities an over-abundance of definitions and a paucity of consensus. There are as many definitions as there are evaluators because definitions of programme evaluations tend to have a strong flavour of the disciplinary background of the definers.

Among the earlier definitions of evaluations was the one provided by Riecken (1952: 4), who defined evaluation as: "the measurement of desirable and undesirable consequences of an action that has been taken in order to forward some goal that we value." In this sense, an evaluation presupposes the existence of some programme or activity to be evaluated. Riecken further delimits the evaluation process when he concludes that any intentional social action can be the object of an evaluation study. According to this approach, the activity being evaluated will usually be one of deliberate social change. In other words, evaluation is the process whereby man attempts to check upon his own ability to influence other men or his environment.

This definition is supported by Borgatta (1966: 182), who finds that "research problems in evaluative research ... recur in the many circumstances where programs operate manifestly to improve existing conditions, or where efforts are being made to prevent or stop deterioration of existing conditions."

The emphasis upon social change and outputs as the subject of evaluation is underscored by writers such as Klineberg (1955), Suchman (1967), Fink and Kosecoff (1978) and Freeman et al. (1980), who take a somewhat broader perspective to include the analysis of efforts (programme inputs) as well as outcome. Fink and Kosecoff (1978: 1), for example, included in the definition of evaluation: "a set of procedures to appraise a program's merit and to provide information about its goals, activities, outcomes, impacts and costs."

Fink and Kosecoff advocate a systems approach to programme evaluation and, if taken literally, would leave very little outside the legitimate concern of the evaluator. In a similar context, Freeman et al. (1980: 25) defined evaluation as: "any information obtained by any means on either the conduct or the outcome of interventions, treatments or of social change projects."

Broad these definitions may be, but they do not describe usefully the current boundaries of the evaluation field and exclude the basic approaches to the gathering of valid and reliable evidence that have been developed in the social sciences. In view of these shortcomings, Rossi and Freeman (1985: 19) provide a "simple" definition of evaluation as: "the systematic application of social research procedures in assessing the conceptualization and design, implementation and utility of social

intervention programs."

Implicit in this definition is the use of social science research methods to judge and improve the planning and monitoring of the effectiveness and efficiency of human service programmes. In addition, evaluations are systematic to the extent that they employ social science approaches to the gathering of valid and reliable evidence.

Attempts have been made in the past years to employ social science research methods to evaluate the impacts of water supply programmes in the developing world. A review of some of these studies is the focus of the next section. The methods employed and the results that emerged from them are examined and the shortcomings of the studies related to this research are discussed.

Overview of Evaluations in the Rural Water Sector

Several evaluative studies have been conducted in the rural water and sanitation sector in the developing world. These studies have employed a variety of methods and measures in their assessments. One such study is the comprehensive interdisciplinary evaluation conducted by Feachem et al. (1978) in Lesotho. In that study, the evaluators investigated the social and economic benefits that could be traced to a water supply programme that was implemented in rural communities in Lesotho. The evaluation design was based on comparisons between villages with and without improved water supplies. The indicators used for the assessment were impacts of the programme on health, sesotho beer (*joala*) brewing, livestock husbandry and communal gardens. The time frame for this interdisciplinary, cross-

sectional study was twenty-five months.

With the use of observational methods, Feachem and his colleagues concluded that there were no measurable reductions in water-related diseases in the villages with improved water supplies. As well, the programme had no impact on brewing, livestock and communal gardens in the beneficiary communities (Feachem et al., 1978). A shortcoming of this study was the reliance on hospital records in their assessment of the health impacts, in the absence of baseline data. Rather than relying on hospital records and observations, interviews with the project beneficiaries could have provided useful information on the measured indicators.

Utilizing a rather different approach, McGowan and Burns (1988) assessed the health, economic, social and environmental impacts of a CARE/Sudan Interim Water Supply and Management Project through interviews with project staff, Village Water Committees, random groups of water users at project water sites and non-project sites, as well as visits to water source sites of both project and non-project areas. Like the Lesotho study, this evaluation was conducted without baseline data. A more directly measurable economic benefit, according to McGowan and Burns, was the use of distribution-point wasted water for growing vegetable gardens and tree nurseries. Many villagers at project sites expressed considerable enthusiasm about both dietary benefits and income generation as a result of the gardens raised with wasted water. Socially, the attention given to women through extension lessons raised the awareness and educational levels of women in the project areas, thereby improving their social standing in the communities. However, according to the

evaluators, the lack of baseline data made it difficult to assess the health and environmental impacts of the project.

Although the evaluators utilized a variety of methods, the study had some flaws. They relied heavily on project staff rather than on project beneficiaries in their assessment of the impacts. As well, the time frame for the study puts into question the validity and reliability of their results. It took the evaluators fifteen days for the entire study, with only six days spent in all ten communities selected for the study, thus, contributing to the lack of information on health and the environment.

A similar time period was spent in evaluating another CARE Water Supply and Sanitation Project in the Dominican Republic. Silva et al. (1989), evaluated the economic, health, convenience and service level benefits to communities as a result of the project. In addition, they analyzed the Health Education component of the project and its impact on knowledge, attitudes, and practices of project beneficiaries. The evaluation team developed three instruments to assess the project's performance in the field. The instruments developed were: a) an engineering survey to collect data on water quality and quantity and to assess the knowledge of pump technicians and plumbers on operation and maintenance, b) a questionnaire for water committees to gather information on management practices of the water committees and, c) a households survey which sought information on water use and sanitation practices and the role of women in the project. Based on these methods, they concluded that the relationship between water and health was understood and that some health practices had improved. The failure of the evaluation team to specify

their measures/indicators made it difficult for them to assess the economic impacts and service level benefits of the project. Moreover, the short time frame prevented intensive observations in the households, in the assessment of the health education component of the programme.

In her study in Kati County, Mali, Belli (1988) assessed the impacts of the provision of three boreholes (one specifically installed for vegetable gardening) on vegetable production in Sougoula village. Based on participant observation, she observed how women utilized an Indian model pump for community vegetable gardening activities. She concluded that one hundred and twelve women were involved in market gardening activities, due mainly to the provision of the hand pumps, and technical advice from a woman agricultural expert.

Frankel (1974) in his impact study of water habits, sanitation, and health conditions, compared two Northeast Thai villages (Ban Fang and Ban Phongsawang) with similar socioeconomic characteristics and both having the advantage of a potable water project. His primary data collection method was through interviews with villagers. He concluded, based on responses from his interviews, that water habits were very similar in both villages and sanitation practices were identical. However, more frequent bathing, more water used for gardening and an observed improvement in gardening were apparent in Ban Phongsawang. His approach was systematic, yet, the study lacked a control sample. Moreover, the use of observations could have revealed water habits and sanitation practices better than direct interviews.

In assessing the benefits of a rural water supply project in Malawi, Glennie

(1983), utilized over ten years of his experience in the water sector as an approach for his evaluation. The assessment was therefore based on his own perspective. The most significant benefit he noticed was the reductions in time spent drawing water. The water supply also enabled people to return to their former deserted villages and thereby increased the area of land under cultivation. The problem with this study was the limited use of scientific research methods, thus rendering the analysis very subjective and the results difficult to verify.

In their evaluation of the Integrated Water and Sanitation Development Project in Nigeria, Rooy and Donaldson (1990), observed that there has been migration into some participatory villages with hand pump-equipped boreholes in Kwara, Imo and Gongola States. They also observed that in some areas of Kwara State where there has been complete eradication of guinea worm, quality of life has improved and this was reflected in increased school attendance and higher productivity of farmers. The findings of the study depended solely on observations, thus making it difficult to substantiate whether migration was due to water availability or to some other confounding variable. In addition, improved quality of life of the beneficiaries was assessed based on interviews with project staff, rather than soliciting the information from the beneficiaries.

Bah et al's., (1991) study in rural Sierra Leone utilized both qualitative and quantitative measures to assess the socioeconomic impacts of improved wells in four villages. Their approach involved the use of standard survey methods including structured questionnaires, semistructured interviews with key informants, participant

observation and quantitative monitoring of water extraction. They observed that 74% of the drinking water for the households studied came from traditional sources, rather than from the improved wells. The limited use of the wells was attributed to the taste of the water from the concrete lining.

According to the evaluators, the time spent in collecting water in the four villages under study was significantly longer from the improved wells than from the traditional sources, even though the wells were nearer than the traditional sources. This was largely attributed to the low flow rate of water from the wells resulting in long queues, hence the longer time. On the economic impacts of the project, the study showed that, there was no clear evidence that wells had affected lifestyles and commerce in the villages. The problem associated with the study is that, the indicators used to assess the economic impacts of the project were unknown, thus making verification of the results difficult. As well, the interview with key informants, rather than a random selection of individuals in the communities may tend to produce biased responses.

Although the USAID suggested an evaluation model which focused on system operation, performance and impact in 1980 (Warner, 1981), impact studies conducted by the agency or on its behalf had methodological problems. In the Philippines, an impact study was conducted for USAID after five years of the completion of a provincial water project. The purpose of the study was to assess the health and economic impacts of the project (Magnami et al., 1983). Through interviews with programme beneficiaries, the evaluators concluded there was no evidence the project

had a substantial impact on health. In addition, the evaluators could not assess the magnitude of economic gains resulting from the project due to the problem of quantification. Although one objective of the project was to establish an evaluation methodology for USAID, the disappointing results, due to the method employed, rendered its applicability in the future difficult. Rather than relying on quantitative information, qualitative information through observations and interviews with the users could have revealed some of the economic gains resulting to the programme.

Warner et al., (1986), in evaluating the USAID/Malawi Self-help Rural Water Supply Programme, focused on the actual performance of project institutions and the measurable health, economic and social impacts arising from the programme. Primary data collection for this study involved interviews with officials of USAID (both in Malawi and in Washington, D.C.), Ministries of Health and of Water Supply and visits to project sites. Based on these methods, the evaluators concluded that the project has developed leadership and organizational skills within the rural population, and in some areas, the project resulted in substantial time savings for women.

In another USAID Rural Water Supply and Sanitation Project in Togo, Roark et al (1988) considered not only the inputs and outputs but also placed emphasis on the project's impact on health, economy, environment and women. Similarly to the other USAID sponsored study in Malawi, data collection methods were based on a review of project documents, meetings with project staff and paying visits to sampled project villages. On the impacts of the project on the beneficiaries, the evaluators conceded that, direct evidence of impacts was often difficult to measure, even if there

had been a pre-project baseline study. Nonetheless, the evaluation team was confident that, because of the overall success of the project in meeting most of its objectives, the project's goal of improving the living conditions of the Togolese rural people had been achieved and that health had generally improved.

A shortcoming of the Malawi and Togo evaluations was the reliance on project staff and government officials. This approach made it difficult for the evaluation team to assess the economic and social impacts of the projects on the beneficiaries. As well, both studies lacked comprehensive research designs, control samples and observational methods.

In 1991, VanSant et al. assessed the health, environmental, community and institutional impacts of a Rural Water Borne Disease Control Project in four regions in Swaziland. Their design involved a random selection of seven sites in four regions, interviewing central, regional and field staff from home country implementing agencies, USAID and non-governmental organizations involved in the project. Based on this, the evaluators concluded that, the single most visible impact of the project was the provision of plentiful potable water for domestic use to an estimated 54,000 rural Swazis in 52 communities through some 529 water taps. According to them the available data were inadequate to documen. decreases in diarrhoea and schistosomiasis. Like other USAID studies, users views were not considered.

The focus of the Zomba East Piped Water Project evaluative study, also in Malawi, was on the socioeconomic impacts of the project on the beneficiaries, including the effectiveness of local organizations for management and maintenance

(Kandoole and Msukwa, 1981). Using Zomba South as the control area, the detailed village survey included a census, observations (for water collection and use) and household interviews about socioeconomic impacts. However, due to lack of reliable data, the study left out the impact of improved water supplies on the health of the rural population, which is one of the main justifications for introducing the rural water supply scheme.

With the use of observations and interviews, Narayan-Parker (1990), assessed the impacts of two year self-help rural water projects in four communities in Kupan and Belu Districts in Indonesia. Among the direct benefits of the project was a marked decline in diarrhoea and skin diseases as a result of the improved water supply in Belu District. There was also a marked increase in the number of women selling vegetables after two years of the project. The project also induced feelings of pride, self-confidence and competence in women, enabling them to emerge as leaders and undertake more challenges. A limitation of the study was the lack of a control sample to compare health status, in the absence of baseline data.

In analyzing the socio-economic impacts of the Upper Region Water Supply Project on the lives of women in Bolgatanga District, Harkness (1983) dwelt on time savings for women, recognizing the problems of resources, expertise and time. Utilizing questionnaires and participant observation techniques, her study looked at the lives of village women to ascertain how the provision of drilled wells has affected their lives. She compared this with earlier anthropological studies done in the area. She concluded, among other things, that the economic impacts of the water supply

project appeared to have been insignificant with respect to the income earning activities of women. A design flaw in the Harkness study was the lack of a control sample to compare income earning activities of women.

The studies discussed above have focused on a variety of indicators in evaluating the impacts of rural water supply programmes. However, other impact studies have concentrated on only single indicators, particularly health and on single diseases. The focus of the impact study in Mirzapur in Bangladesh, for example, was on health only (Aziz et al., 1990). In this study, a case control study design was adopted to assess the health impacts of a water supply, sanitation and hygiene education package implemented in rural communities in Mirzapur. The study was a longitudinal one, with the intervention and non-intervention areas followed up over a period of four years. Standard questionnaire surveys, combined with occasional observational studies to confirm the accuracy of responses, were used as data collection instruments.

The results of the Mirzapur study showed that 90% of households in the intervention area used handpump water for practically all domestic purposes. In addition, the project had a significant impact on childhood diarrhoea disease in the intervention area, where the incidence of diarrhoea fell to three-quarters of that in the control area. Furthermore, the project resulted in reductions in ascaris infection by more than one-third. A shortcoming of the Mirzapur study, apart from its focus on selected health indicators, was its time frame (five years). A study of this scale is costly and hardly provides a ready method for the operational evaluation of water

supply programmes.

A similar time frame was utilized by Azurin and Alvero (1974) to test the effect of either improved water supply or improved waste disposal (or both) against cholera infection in the Philippines. The method employed for the study was a house to house canvass by an epidemiological aid who visited the communities (both control and intervention) daily, taking rectal swab from persons found to have diarrhoea. The specimens were later brought to a project laboratory on the same day and examined for cholera vibrio. With this method, their results showed that improvements in either water supply or toilet facilities or both were effective in significantly reducing the incidence of cholera in the corresponding study communities as compared to the control area. After four and a half years of study, they concluded that the provision of sanitary facilities for human waste disposal can reduce the incidence of cholera by as much as 68%, while the provision of a safe water supply can reduce it by 73%. Where both toilets and water supply are provided, the incidence can be reduced by as much as 76%. This study has a similar methodological problem as the Mirzapur study.

In Upper Burma, the National Department of Medical Research carried out an evaluative study to determine the health impacts of a water supply and environmental sanitation programme. The study included both a cross sectional comparison between villages, with and without water supply and sanitation facilities, and a longitudinal comparison of villages before and after the introduction of the improved system (Rosenhall, 1990). The results of the study, based on observational

methods and household questionnaires, indicated large proportions of the population in the four of the six study villages used the improved water supply source. However, water from the improved source was free from contamination, but usually got contaminated during storage and handling at home. Thus, the impact of the improved system on the morbidity rates of diarrhoea was rather marginal, and there was no evidence of marked changes in villages with access to and use of the improved system.

Lindskog et al. (1987), using three villages (two with an intervention and one acting as a control), evaluated a piped water supply in rural Malawi. Observations and interviews were the instruments used in collecting the data. Both the comparison and intervention groups were examined for slightly more than one year before and one year after the intervention. The results of their study indicated that the storage and handling arrangements of water in dwellings were prone to contamination. There was also a 50% reduction of diarrhoea with increased water quantity and sanitation.

In another study in Nigeria, this time in Imo State, a quasi experimental study design was adopted. In this study data was collected from intervention and control areas, to evaluate a water supply and sanitation project (The Imo State Evaluation Team, 1989). The evaluators utilized both longitudinal and cross-sectional surveys to assess the health impacts of the rural water project. Approximately 850 households in the intervention area, and 420 in the control area were selected for the sample. The team concluded that, due to the widespread use of water from the

boreholes, the project showed an impact on dracunculiasis in the intervention area. In regards to diarrhoea, the impact was not clearly shown. The impact on diarrhoea was found to be limited to certain subgroups of the population within the intervention area. In their opinion, young children in particular were at greater risk of diarrhoea in the wet season if the estimated household water collection time was more than two hours per day. The short life cycle of the programme (two years) led to the inconclusiveness of the results.

Gaps in the Literature

The anticipated impacts of improved water supply and sanitation facilities, as well as the need for accountability for the huge investments in the sector, have initiated a number of evaluative studies in rural communities in the developing world. The preceding studies indicate that the methods and indicators/measures employed do vary, ranging from the evaluation of water and sanitation impacts on specifically defined diseases to multiple measures, involving health, economic, social and environmental. Results of these studies, therefore, vary and the validity of the conclusions are likely to be affected by the choice of the evaluation method. The question that needs to be asked is, with the numerous impacts studies in the rural water sector, what is the justification for another study? A summary of the studies reviewed (Table 4.1) brings to the fore certain issues that need to be addressed in the rural water evaluation literature.

From the literature review, it appears the core of most impact studies has

been to concentrate on health impacts and the reliance on project staff for information on the programmes' impacts. The literature has failed to focus on the rural peoples' knowledge, awareness and impacts of issues related to programme objectives. These issues are addressed in this study in an effort to fill this gap in the literature. In regards to the environment, the literature has rarely dealt with the practices related to the environment and which have been addressed by rural water supply and sanitation programmes. This study addresses these gaps in the existing literature. Previous studies have also ignored the views of decision makers (government officials): their awareness, involvement and impacts of the programme on the beneficiaries. Where their views were solicited, they have been limited to specific ministries at the national level. The viewpoints of local and regional government officials are seldom solicited. An examination of these issues is important before any sound plan to improve programmes can be instituted.

Table 4.1
Summary of Some Evaluations in The Rural Water Sector

Reference	Type of Study	Indicators/ Measures	Data Collection Instruments	Time Frame	Evaluators
Feachem et al., 1978	Case control	Health, gardens, beer and livestock	Observations	25 months	Consultants
McGowan & Burns, 1988	Case study	Health, economic, social and environmental	Interviews, observations	15 days	Consultants
Silva et al., 1989	Case study	Health, economic, convenience, service level	Interviews, household and engineering surveys	23 days	Consultants
Belli, 1988	Case study	Vegetable production	Participant observation	n/a	Consultant
Frankel, 1974	Case study	Water habits, vegetable gardening, sanitation	Interviews, visits to water supply Sites	n/a	Local specialist
Glennie, 1983	Case study	Migration, time savings	Personal experience	n/a	Consultant
Bah et al., 1991	Case-control	Socio-economic	Personal and key informant interviews, observations	n/a	Ph.D. study
Magnami et al., 1983	Case study	Economic, health	Interviews	n/a	Consultants
Warner et al., 1986	Case study	Health, economic, institutional and social	Interview with project Staff, visit to project sites	n/a	Consultants
Roark et al., 1988	Case study	Health, economy, environment and women	Interview with project Staff, visit to project sites	25 days	Consultants and Local specialists

VanSant, et al., 1991	Case study	Health, environment, institutional	Interview with project staff and site visits	21 days	Consultants
Narayan-Parker, 1990	Case study	Health, socio-economic	Community observations, interviews	n/a	Consultant
Harkness, 1983	Case study	Socio-economic, women	Interviews, observations	n/a	M. A. study
Aziz et al., 1990	Case-control, longitudinal	Health	Questionnaires and observations	4 years	Consultants and local specialists
Rooy & Donaldson, 1990	Case study	Migration, health	Observations	n/a	Project staff
Rosenhall, 1990	Case-control, longitudinal	Health	Observations, household interviews	n/a	Project staff
Blum et al., 1983	Case-control, longitudinal	Health	Observations, household interviews	n/a	Consultants and local specialists
Kandoole & Msukwa, 1981	Case-control	Socio-economic	Observations, household interviews	n/a	Local specialists
Lindskog et al., 1987	Case-control, longitudinal	Health, sanitation	observation, interviews	n/a	Consultants
The Imo State Team, 1989	Case-control, longitudinal	Health, nutrition	Observations, interviews	3 Years	Consultants

It is hoped that information from users and decision makers will provide further insights for both aid agencies and governments to tailor future efforts to those deficient areas of the programme. If differences exist, education and management efforts could be better designed and targeted to express local needs, concerns and knowledge levels of the beneficiaries who rely on programme facilities for a wide range of their needs.

Finally, and more importantly, by examining the views of programme beneficiaries on impacts related to programme objectives, as well as those of decision makers, the study will contribute to, and perhaps broaden, the methods employed in conducting evaluation studies of the rural water sector in the developing world.

Chapter Five

METHODS OF DATA COLLECTION AND TECHNIQUES OF ANALYSIS

This chapter sets out to provide the research design adopted for this present study. It also describes the various instruments used in collecting the data and the levels involved, as well as the sampling procedure. A documentation of the techniques used in analyzing the data is also presented. The final section of the chapter deals with data validity and reliability as well as the problems encountered in the field.

Research Design

The study designs of the studies reviewed in Chapter Four varied considerably. Some of these studies were case studies. Others were longitudinal, with a before and after comparison of conditions in intervention areas. The Upper East Region Water Supply Programme, which is the subject of the present study, was implemented without a baseline study conducted in the area. Under such circumstances, the choice of an appropriate study design becomes critical. This is because the validity and the verification of the conclusions from the impact study depends on the choice of the evaluation method (Schlesselman, 1982).

A method which has been proposed for evaluating the impacts of improvements in water supply, hygiene and sanitation in the absence of baseline data is the case-control method (Ibrahim, 1979; Esrey et al., 1985; Briscoe et al., 1986). Although this method has been utilized in impact studies in the rural water sector, its use has been largely limited to health impact studies. In a case-control study,

individuals with a particular condition are selected for comparison with a series of individuals in whom the condition is absent. It does not require the same population to be studied both before and after an intervention. It is relatively quick to mount, conduct and inexpensive and requires comparatively fewer subjects. Moreover, it can give reliable results (Schlesselman, 1982; United Nations, 1987; Lindskog et al., 1987).

Due to logistic reasons, the absence of baseline data and the relative merits of the approach, the case-control method was utilized in this present study. The object was to conduct a survey in areas that received the programme and areas that did not. This was to ensure meaningful interpretation of the data, because without a control sample, there was no way of distinguishing the impacts resulting from the water supply programme and the impacts that would have occurred in any case due to other factors of social, economic, and environmental change.

Methods of Data Collection

To achieve the research objectives, and in line with the approach selected for the study, data was collected from both primary and secondary sources. Primary data was collected by means of formal and informal interviews, participant and spot observations as well as through personal communication. A review of regional, central government and CIDA documents was conducted as secondary sources.

Interviews

Interviews were conducted at different levels. The first level involved interviews in the communities, in both programme and non-programme areas. The broad scope of information required necessitated the use of a household, interviewer-completed semistructured interview schedule. Unstructured group interviews with women constituted the second level of the interviews and were used to supplement information revealed by some of the interview schedules. The third and final level was to interview government officials involved with the sector.

Direct Observation

Direct observation by means of the participant observer approach was used to supplement data and also as a primary method of data collection. Two observation sheets were prepared. Village-wide observation was used in order to provide information on the physical environment and water sources (Appendix 1). The second observation dealt with the household and took into consideration household water use, personal hygiene, sanitation practices, waste disposal and food preparation and storage (Appendix 2). These structured observations were complemented by the use of spot checks.

Five female assistants were assigned the task of conducting the observations. They presented themselves at about 6 a.m. and continued observations till 6 p.m. or until members of the household left the house, then on their return, they (the assistants) returned to conduct the observations. The observations proved to be

useful in terms of identifying the environmental consequences of people's activities. Three days were devoted to each household for the observation. This was to ensure that what was observed was a consistent practice in the household.

Secondary Sources

Documentary information from regional and central government, as well as CIDA offices was the major secondary source. These places provided background information of the study area, village population size, economic activities, the health status in the region and programme activities. In addition, the Regional Ministry of Health Statistics Unit provided data on reported cases of water and sanitation-related diseases in the region.

Interview Schedule Design and Administration

Two different sets of interview schedules were designed for the programme and non-programme areas. All questions were open-ended. Open-ended questions were thought to be useful in order to fulfil the research objectives. It does not force the respondent to adapt to pre-conceived answers, allowing the respondent to answer questions more freely (Nachmias and Nachmias, 1987; Babbie, 1992). In this way their answers reflect their opinions, views and experiences. Thoughts can also be expressed more spontaneously, which can provide deeper insights into the respondents' attitudes, views and opinions. Despite the utility of open-ended questions, they are more difficult and time consuming to analyze.

The two sets of questions were written in English, and translated to Buli, Frafra and Dagbani, when administered in Builsa, Bongo and Bawku (the programme areas), and West Mamprusi (the non-programme area) respectively. The answers were recorded in English. The interview schedules were administered by the author with the assistance of nine interviewers, two each in the individual programme districts and three in the non-programme area.

The broad scope of the study required personnel who were familiar with particular community and fluent in both the local language and English. Interviewers were chosen to eliminate possible biases in the interview responses that are common in cross-cultural research. To prevent linguistic and conceptual errors, it was ensured that interviewers were from the community. This proved useful because it secured the support and approval of village leaders in conducting the study.

The selection process of the interviewers was done in consultation with local Community Development Officers, who suggested the use of both male and female interviewers. Their suggestions were reiterated by Mary Margaret Issaka, a Community Liaison Worker for WARDROP Engineering Ltd, who has been working in these communities, and as such has extensive knowledge of field survey issues in the villages. In all, five females and four males were selected.

Interview Schedule for the Programme Area

The interview schedule for the programme area were designed to provide information in the following areas related to the series of objectives of the

programme and also to programme activities:

a) economic activities pursued by respondents;

b) effects of borehole/handpump on livestock and vegetable production, and other economic activities, including *pito* brewing and shea butter oil processing;

c) knowledge of Water for Health Education;

d) knowledge of the mode of transmission, prevention and treatment of water- and sanitation-related diseases;

e) knowledge of the work and effectiveness of Community Water Organizers and Water User Committees.

f) the effects of the provision of water in the communities on women's workload and

g) the involvement of women in decision making at home as a result of women's participation in the programme (Appendix 3).

The interview schedules were directed to heads of households and spouses of heads of households, and in some cases to female heads of households, who had lived in the community for a period of not less than five years. The interviews were conducted either in the morning, afternoon or evening, depending on the time suitable for the interviewee, and were conducted in their courtyards. Any unusual information from the interview was cross-checked with an observation. For example, if a person indicates during the interview that he/she is a *pito* brewer, an attempt is made to identify the materials and equipments used for brewing in the household.

The questions were pre-tested with sixteen respondents (eight men and eight women) randomly selected from Sambruno, a rural community near Bolgatanga. As

a result of the pre-testing, two questions related to household income and number of livestock holdings were removed from the survey, as the respondents did not feel comfortable in responding to them.

Non-Programme Area Interview Schedule

The schedule administered in the non-programme area was designed in part to collect the same information, as that sought from the programme area (Appendix 4). However, it was not possible to interview both groups using one questionnaire, because the non-programme area did not benefit from the water and sanitation programme. Therefore, a separate interview schedule was justified.

Information collected dealt with economic activities pursued by respondents, knowledge on the mode of transmission, prevention and treatment of water- and sanitation-related diseases, and knowledge on water, health and sanitation education. The questions were pre-tested among ten residents randomly chosen from Kakasunanka, a community near Tamale, the capital of the Northern Region of Ghana. All the questions were answered without difficulty, as such, no changes were made.

Government Officials

The selection of the government officials was based on two basic criteria:

1. the linkage of that institution with the rural water sector in Ghana and,
2. the preparedness of the head or a designated official to engage in the study.

In all, nine government officials were interviewed. The interviews proceeded on the basis of a series of questions related to the research. In each case they were given considerable liberty in expressing their own views. The interview schedule was designed to provide information on: a) their awareness of the programme, b) impacts on vegetable production, c) environmental sanitation, d) health, e) transformation of the beneficiaries in terms of project planning and implementation and, f) the standard of living of the people in the region as a whole (Appendix 5). They were also provided with the opportunity to express their views on the economic and sociocultural impacts of the programme on the people.

The units of investigation were the political heads of the sector ministries: Finance and Economic Planning, Works and Housing, Local Government and Health. In view of the tight schedule of the political heads for Finance and Economic Planning and Local Government, their respective Principal Officers in charge of donor projects, were designated for the interviews. In the case of the Ministry of Health, the National Director in charge of Public Health indicated that the Regional Director of Health Services was the best resource person for such a study, since the head office in Accra was not really aware of what was happening on the "ground".

Interviews at the regional level involved the former Deputy Regional Secretary (now Deputy Regional Minister), the Regional Director of the National Council on Women and Development and, the Regional Representative of the 31st December Women's Movement in charge of projects. Former District Secretaries of two

programme districts constituted the final level of the interviews with the government officials.

The interviews with the government officials were conducted in their respective offices and were recorded on audio tapes. After the interviews the tapes were played to them to ensure verification of the information. Each interview lasted between twenty-five to thirty minutes.

Unstructured Group Interviews

This aspect of the study involved group discussions to obtain in-depth information on the economic, social and cultural impacts of the programme on the community. The specific areas covered were:

i) the impact of the provision of water in changing economic activities in their community;

ii) impact on vegetable production;

iii) impact of water provision on their standard of living;

iv) impact on migration

v) impact on the community's attitude to project planning and implementation;

vi) impact on the traditional role of women;

vii) impact on children's health;

viii) impact on the overall cleanliness of the community;

ix) effect on water and sanitation-related traditional practices and;

x) community cohesion.

An average of seven women in each community were interviewed separately in four randomly selected communities in each of the programme districts. The discussions were held in the late evenings when most people had returned from their farms. In all cases, the discussions took place under shady trees to prevent both interviewers and interviewees from the scorching sun (Plate 5.1). The discussions usually lasted between two to three hours.

Plate 5.1 Interview with a Group of Women

Sampling Procedure

Having defined our study design and specified our data collection methods, the next step was to specify our method for selecting the study subjects. Ideally, all handpump communities should have been included, but time and cost constraints clearly made this unfeasible. In view of this, a form of multi-phase sampling was adopted for the study (Sheskin, 1985). This enabled one overcome the problem of interviewing at geographically dispersed settlements in the region.

In the first phase of the procedure, a representative sample of three districts were randomly selected from a total of six in the programme area. The selected districts were Builsa, Bongo and Bawku West (Fig. 5.1). In view of the spatial location of the selected districts, it was envisaged that the spatial qualities of the districts not included in the study (Bolgatanga, Bawku East and Kassena-Nankani) were catered for by the selected three, since they exhibit similar physical and socio-economic characteristics (Dickson and Benneh, 1980).

A fourth district, the West Mamprusi District, representing the non-programme area, was selected from the Northern Region. This district shares a border with the Upper East Region and lies to the south (Fig. 5.1). It was selected because it has similar social and physical characteristics as the Upper East but was not included in the programme.