University of Alberta

The Representation of Evidence in University Undergraduate and High School Chemistry Textbooks

by

Elizabeth Vergis

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Education

in

Adult Education

Department of Educational Policy Studies

Edmonton, Alberta
Spring 2007

Your file Votre référence
ISBN: 978-0-494-29910-4
Our file Notre référence
ISBN: 978-0-494-29910-4

Dedication

To my Mother and Father

with heartfelt gratitude for their unconditional love and support

Abstract

This study addressed a perennial problem in the High School Chemistry curriculum, which is the failure to treat in a systematic and comprehensive way the nature of scientific evidence. I have completed a thorough examination of the three most widely used High School Chemistry textbooks in Canada and a widely used first-year Chemistry textbook for their inclusion of concepts of evidence. I have found that the treatment of concepts of evidence varies widely across programs, within programs across topics, and within and across programs by the concepts of evidence themselves. The issue that most needs addressing is that curriculum policy makers should lay down a curriculum for procedural knowledge using the guidelines employed for substantive knowledge. Then teachers will be bound to teach procedural knowledge systematically and thoroughly. Teachers themselves may need training to "see" and "teach" concepts of evidence which are the building blocks of procedural knowledge.

Acknowledgements

I wish to thank:

Mathew, my husband, for his steadfast love which gave me the freedom to pursue my dreams,

Priyanka, our daughter, whose interest in my work was an inspiration,

Dr. Stephen Norris, my supervisor, whose interest, support and guidance were invaluable,

Dr. André Grace, who encouraged me to perform at my best,

Dr. Frank Jenkins, who posed thought-provoking questions,

Dr. Margaret-Ann Armour and Eddie Ferris, who were on my CRYSTAL committee,

Gail Briton and Carolyn Carlson, the CRYSTAL secretaries, who were always ready to help and

Joan White, our Graduate Secretary, who patiently answered all my queries.

My heartfelt thanks to one and all.

TABLE OF CONTENTS

List of Tables

Page

List of Figures

Page

CHAPTER ONE

The major question that motivated this research study was "How is evidence represented in the university undergraduate and high school chemistry textbooks in the following eight areas: i) Observation and measurement, ii) Choice of measurement instrument, iii) Calibration and error of measurement, iv) Reliability and validity of measurement, v) Statistical treatment of measurement, vi) Design of investigations and "fair tests", vii) Data presentation, and viii) Patterns and relationships in data?" These eight areas capture the categories identified by Gott et al. (2003) in their classification of "concepts of evidence" (abbreviated as CoEs). These CoEs are described at length in my thesis. The major research question was not dealt with per se, but was answered by addressing the following subsidiary questions: 1) How frequently do these CoEs occur, by grade level, by topic and by publisher? 2) From the different types of CoEs employed what is the range in the density of CoEs/Situation? 3) Are there any gaps in the depth, accuracy and appropriateness of the occurrences of the CoEs, when compared to Gott's classification system? 4) Does Gott's classification system of CoEs require any additions or deletions? 5) Was there was a development in the sophistication of the CoEs as one moved from Grade 11 to Grade 12 to university? 6) (a) Can CoEs be taught or learnt? (b) How would you improve/modify the teaching of these CoEs to make it more effective? 7) Can the CoEs be arranged hierarchically? and 8) What curriculum objectives would you recommend regarding learning about evidence?

Science has been heralded as a core subject in the curriculum at both the Elementary and the Junior High levels, in all schools, right across Canada. In higher adult education as well, science features prominently as an essential subject, especially for students who wish to pursue careers in medical, engineering and related fields. If this status is to be maintained, then science must have universal value to both the common person on the street, working in a non-science environment, and to the minority who will pursue it as a lifelong career. Despite the fact that science is a core subject for all pupils, the proportion of those who will use science for career purposes is relatively minuscule. For the majority, therefore, science is just part of their general education – one aspect of their preparation for life. Consequently, there is a growing call for science education to provide a more effective preparation for citizenship (Jenkins, 1997; Millar & Osborne, 1998).

The aim of this growing call is to improve ***scientific literacy***; that is, to increase the numbers of "scientifically literate" adults in society and hence improve the ***public understanding of science***. Although there is no single definition of scientific literacy, one possible way of characterizing a scientifically literate person may be as "one who has both the interest in scientific issues and the capability to inquire, as well as some in-depth knowledge of a particular area of science" (Hinman, 1998). The components of literacy suggested as vital by Holliday et al. (1994) and Sutman (1996), include, the willingness and the ability to continue learning science; the development of scientific processes; a contemporary view of science; and the ability to communicate *scientific* ideas to others effectively. The theme of scientific literacy most akin to this study is "Science as a way of thinking." Under this theme there are eight sub-themes (Chiappetta et al., 1991), of which the one most relevant for this work is "Discusses evidence and proof," since this study focuses on the representation of evidence in high school and undergraduate chemistry textbooks.

According to the American Association for the Advancement of Science (AAAS) in its report, *Science for All Americans*, the scientifically literate person is:

> … one who is aware that science, mathematics, and technology are interdependent human enterprises with strengths and limitations; understands key concepts and principles of science; is familiar with the natural world and recognizes both its diversity and unity; and uses scientific knowledge and scientific ways of thinking for individual and social purposes. (AAAS, 1989, p. 4)

The *National Science Education Standards* set their reform goal as the attainment of scientific literacy by all. They defined scientific literacy as being able to:

- Experience the satisfaction of understanding the natural world.
- Use scientific thinking in making personal decisions.
- Participate intelligently in societal decisions on science and technology.
- Attain the skills and knowledge that are required for being productive in our current and future economies. (NRC, 1996, p. 13)

Although both the AAAS report (1989) and the NRC standards (1996) treat the problem of scientific literacy as their central focus, neither really defines the problem it is trying to solve, nor does it define the scientific literacy it is trying to attain. Definitions are not given in simple terms that suggest an understanding of how to implement the solutions, or how to recognize the solutions once they are achieved. The crux of the matter is that a clear vision is needed of what scientific literacy means and what it will look like when it is attained. Otherwise, as Wright & Wright (1989) caution; "If the vision is not clear, the implementation will fail."

Without scientific literacy, it is difficult to make informed decisions about the interrelated educational, scientific and social issues that one confronts every day in the newspapers (Glynn & Muth, 1994). Scientific literacy involves more than just science knowledge. To be scientifically literate, students must have the reading ability to evaluate the print-based information presented to them, as well as the writing ability to communicate their thoughts to others (Holliday, Yore & Alvermann, 1994).

The Council of Ministers of Education Canada (CMEC) (1997) has, in addition, identified an attitudinal component as foundational to scientific literacy. This attitudinal component will encourage students "to develop attitudes that support the responsible acquisition and application of scientific and technological knowledge to the mutual benefit of self, society, and the environment." Attitudes, in general, are modeled for students by example and reinforced by demonstrations of approval. Attitudes are not acquired in the same way as skills and knowledge, because "attitude development is a lifelong process that involves the home, school, the community, and society at large (CMEC, 1997)." The development of a positive attitude has an important role in students' growth because it interacts with their intellectual development "creating a readiness for responsible application of what they learn." The attitudes component emphasized six ways in which science education can contribute to attitudinal growth, which are: i) appreciation of science, ii) interest is science, iii) scientific inquiry, iv) collaboration, v) stewardship, and vi) safety.

For a student going through the present public school system, becoming scientifically literate is not as simple an undertaking as it appears, since in practice, the same science curriculum has to accomplish two tasks: of providing the first stages of a training in science for a minority of students, and of giving access to basic scientific literacy for the majority. These two functions often appear to be in tension, because they may lead to differences in curriculum content and emphasis. Both these tasks, however, share two common goals. First, both have the aim of helping students come to an understanding of some parts of the vast corpus of substantive scientific knowledge. Second, both take aim at the fact that scientists are also citizens. Scientific expertise is often limited to a narrow area of specialization. Both as an individual and as a citizen, a scientist may have to take decisions which involve ideas outside his or her narrow area of expertise. In such areas, the scientist is, in many respects, an educated "lay person." In addition, scientists, qua scientist, require the skill to communicate their specialist knowledge to non-expert audiences, being sensitive to the possibility that misunderstandings can occur about the nature of scientific knowledge. Scientists must be made aware that they have a responsibility to chart the direction taken by, and the emphasis placed on, their area of work. Besides scientists must be cognizant of how dependent the scientific enterprise is on wider societal support, and how "education, both internally and internationally, is a political football" (Husen et al., 1992). It is important, for these reasons, that scientists develop a sophisticated understanding of the institutional nature of science and of the controlling power it wields, both socially and politically.

Scientific literacy has become such a fundamental goal of science education that it guides and directs the development of teaching tools, such as the science textbooks. For the majority of students, the science textbook is their first contact with science. The primary instructional tool in many classrooms is the science textbook. "School science curricula can be placed on a continuum from textbook-driven to teacher-driven." (Glynn & Muth, 1994). With a textbook-driven curriculum, the textbook is the engine that drives the curriculum. It guides the teacher in the selection of topics, the organization of lessons, the assignment of activities, and the construction of tests. In a teacher-driven curriculum, however, the textbook may still play an important role as a reference, but it will not function as the curriculum. The teacher in the teacher-driven curriculum is empowered and has much more control over the instructional methods, and the use of other print-based materials, such as trade books, magazines, and biographies of scientists. The teacher-driven curriculum expects the teacher to know a great deal about science, methods of instruction and about the basic skills of reading and writing science. Overall, in both the textbook-driven and the teacher-driven curricula, the science textbook is a critical factor in the development of scientific literacy: so much so, that being able to read and understand the science textbook is a critical part of being scientifically literate. It is important, therefore, that textbooks emphasize the themes of scientific literacy that are believed to be important, because without a scientifically literate population, the outlook for a better world is not promising (Rutherford & Ahlgren, 1990).

Most of the textbooks being introduced into science education currently state scientific literacy as a major goal. For example, *Nelson Chemistry* "strives to develop a system of communication that portrays science as an endeavor in which concepts are constructed in the mind in order to explain and predict observations made by the senses" (Jenkins, van Kessel & Tompkins, 1993, p. 21). In this textbook, students are encouraged to use a familiar concept or experience to predict the results of an experiment. After creating an experimental design, they collect and analyze the evidence. By comparing the predicted result and the experimental result, they learn to evaluate the concept used to make the prediction. In this way, experience in the science classroom reflects the experience of chemists as they construct and test scientific knowledge. According to *Nelson,* a student who understands how scientific knowledge is constructed and how science relates to technological and social issues, will be "better equipped to make a valuable contribution to life in the 21st century" (p.21). This shows clearly that the authors value scientific literacy and consider it as a major goal in writing this textbook.

In the Preface of the first year undergraduate text used at the University of Alberta, *General Chemistry* by Petrucci et al. (Petrucci, Harwood & Herring, 1997), the authors acknowledge that most general chemistry students have career interests not in chemistry, but in biology, medicine, engineering, agricultural science, etc. They reiterate that general chemistry will be the only college chemistry course for some students, and thus "their only opportunity to learn some practical applications of chemistry" (p. xvii).

Despite the growing call for science education to provide a more effective preparation for citizenship, there has been little attempt to develop a curriculum that is commensurate with the goals of citizenship. Science courses scant in the treatment of the nature, practices and processes of science result in most students leaving school with naïve or severely limited conceptions of science (Driver et al., 1996). The science education prevalent in the last century, leaves far too many students with a confused sense of the significance of what they learned, and a negative attitude towards the subject itself. As early as the elementary and middle grades, students begin losing interest in science

(Bordt et al., 2001). By high school, students of all achievement levels find science hard, dull and meaningless (Bordt et al., 2001, p. 9). For example, there is a growing trend in Canada for students to drop science in high school (Bordt et al., 2001). Negative attitudes toward mathematics and science appear to emerge in mid elementary school and increase throughout high school (Bordt et al., 2001, p. 12). Why is this trend prevalent in Canada? As a consequence of this drop-out rate, scientific literacy suffers, and when it does, our students may in future compare poorly with those of other industrialized nations, in spite of the fact that they compare very well at present (Statistics Canada, 2003).

To remedy this weakness, one needs to reconsider the aims and purposes of science education. It is my contention, that an understanding of scientific evidence and its relationship to scientific facts, concepts and theories is very valuable in enabling and empowering citizens to use science in their everyday lives. Evidence is the centre point of any controversy in science, and it is a concept central to the empirical sciences. By definition, evidence is the ground for belief in any scientific theories, laws or models. In science, evidence is made up of observations and facts, tending to prove or disprove something. Contrary to popular belief, evidence is not simply data. Data must be linked to a conclusion by theory and argument before it can become evidence. Dependence on evidence and reasons is a frame of mind that can be developed in science, and then can be applied much more widely across other disciplines.

Evidence is the central point used to resolve any controversy in science. The extent to which science interacts with people's lives is determined by how familiar they are with the language of evidence. Understanding scientific evidence requires a holistic appreciation of science, and teaching designed to increase understanding of evidence needs to engage pupils fully with scientific reasoning and inquiry. If evidence is taught, and is seen to be useful and functional in everyday life, I conjecture that the image and popularity of science as a school subject is likely to improve.

Learning science requires the acquisition of substantive knowledge and of skills. The substantive knowledge includes understanding the facts, concepts, and theories of science. Skills include knowing how to use an analytical balance, how to draw a graph, how to set up a distillation apparatus, and how to focus a microscope. It is not always recognized that these skills have a distinct knowledge base that is connected directly with the understanding of scientific evidence. The skills must be exercised within an understanding of such ideas as variables and their manipulation, accuracy, fair testing, and the validity and reliability of data. It is these ideas that Gott has collectively termed "concepts of evidence" (2004). Concepts of evidence constitute a knowledge base akin to the substantive knowledge that traditionally has been seen as the heart of science.

If we accept that concepts of evidence form an important part of scientific knowledge, and a part that is crucial to the citizen-scientist, then decisions must be made about the best way of teaching them. There is growing suspicion that concepts of evidence are neither systematically nor comprehensively presented in the secondary school science curriculum. There is also a widespread concern that the traditional substantive knowledge is so vast that there is insufficient time to cover it as well as more general issues such as understanding of evidence. As a consequence, suggestions have been proffered that the substantive content of science courses be reduced to make room for such knowledge as the understanding of evidence, because it is this knowledge that students will use in coping with controversial socio-scientific issues that arise in their later lives as adults.

As an overview, the remaining content of my thesis is divided into four chapters. Chapter Two is a literature review that traces the work done in science education on evidence, leading from its genesis to the work done by Richard Gott in Durham, England. Chapter Three delves, in detail, into the design of this piece of research. Chapter Four describes and analyzes the results obtained as a result of the study. The last chapter, Chapter Five, provides a discussion and final conclusion of the research project.

CHAPTER TWO

REVIEW OF THE LITERATURE

This second chapter commenced with a historical review of science education since it was introduced into the school curriculum, to what it has evolved to at the present time. The chapter then surveyed the concept of scientific literacy – for a definition of what it is, and how it has been understood over the ages. The state of school science and the problems it faced were then explored, along with a look at the role of science textbooks in schools. The chapter then proceeded to examine the four major themes for analyzing textbooks of which the third theme - science as a way of thinking – has got a category under it entitled "discussion of evidence and proof." It was this evidence that was examined subsequently in this chapter. There was a summary of the work done on evidence so far; how data becomes evidence; the three types of evidence; concepts of evidence (CoEs) as part of procedural knowledge; as well as students' perceptions about evidence and their facility in using it.

Historical Review of Science Education

Science became incorporated into the school curriculum in the 19th century in Europe as well as the United States, mainly because of the persuasion of scientists. The notable educators and scientists who spoke publicly in favour of teaching science were Thomas Huxley, Herbert Spencer, Charles Lyell, Michael Faraday, John Tyndall, and Charles Eliot (DeBoer, 1991). They did not have an easy task, because the humanities were then firmly established as the subjects that were thought to lead to the most noble and worthy educational outcomes. Science was thought to be so materialistic and devoid of higher virtue, that scientists had to be careful when arguing about its utility. In spite of this, scientists illustrated the practical importance of science in a world that was becoming increasingly governed by science and technology. Scientists also maintained that science provided intellectual training at the highest level, because it advocated the use of inductive logic by the process of observing the natural world and drawing conclusions from it. They proposed that students would learn this way of thinking by carrying out independent inquiries and investigations in the laboratory. Working so independently in laboratories, would protect scientists from the excesses of arbitrary authority and equip them to participate more fully and effectively in an open and democratic society.

John Dewey defended science as a legitimate intellectual study based on the power it provided individuals to act independently. He said: "Whatever natural science may be for the specialist, for educational purposes it is knowledge of the conditions of human action" (Dewey, 1916, p.228). It was chiefly because of writers such as Dewey that during the early years of the 20th century science education was justified more and more on account of its relevance to contemporary life, and the contribution it made to the development of a shared, common understanding of the world. In 1918, the National Education Association set up a Commission on the reorganization of Secondary Education (CRSE). According to this Commission's report entitled *Cardinal Principles of Secondary Education* (NEA, 1918), "the proper role of education was to develop the individual for effectiveness in a social world" (DeBoer, 2000).

By 1932, however, there was concern that curriculum developers had gone so far into the process of making subject matter "relevant" that they had forgotten the fundamental reason for studying science, which was "to provide a broad understanding of *the natural* *world* and the way it affected people's personal and social lives" (DeBoer, 2000, p. 584). The National Society for the Study of Education (1932), in its Thirty-first Yearbook called *A Program for Teaching Science*, reexamined the goals that had been identified fourteen years previously in the Cardinal Principles with the intention of making them clearer and more substantive. The challenge they faced was "to find the right balance between a broad intellectual understanding of the natural world and the scientific way of thinking on the one hand, and the utility of science for effective living on the other" (DeBoer, 2000, p. 584). The Yearbook Committee believed that science should be studied, not only for its usefulness to individuals in supporting their participation in a democratic society, but also as a powerful cultural force and a search for truth and beauty in the world.

In 1947, the National Society for the Study of Education (NSSE) published its Forty-sixth Yearbook, *Science Education in American Schools*. The theme of social relevance was evident again. The Yearbook Committee expressed faith in the link between science and human progress: this idea had been a prevalent part of the 19th and 20th century thinking about science and technology. In the years following World War II, however, this optimism was tempered by an awakening that scientific developments also had the potential to destroy society. The Committee quoted an anti-science source as stating:

> The present prospect is that, no matter what its way of working may be, science may itself be the end of the human enterprise. The belief in social progress, of which science has nearly always been viewed as the efficient cause, has come to a low state in our time. Few people … any longer believe that mankind is moving forward in the direction of a desirable goal. Even the hope that this

might be so is rapidly waning. . . . Indeed, security, peace of mind, loyalty, friendship, kindliness, and the general attitudes associated with the brotherhood of man appear to be becoming less as science moves forward. (NSSE, 1947, p. 16)

Another factor that changed the relationship between science and society was the growing perception that scientific and technological developments were important investments for national security. In the years following World War II, there was an increasing concern in the U.S about their economic and military status internationally. People wanted science education to play a prominent role so that the U.S. would be assured of remaining a dominant force in the world. Two years after World War II, the President's National Research Board was established to study the country's research and development activities and science training programs. The Board declared:

> Account must be taken of the degree of comprehension of science by the general population. For in a democracy, it is upon the popular attitude toward science that the attractiveness of the profession, the resulting selectivity for those finally entering the profession, and the degree of support obtainable for their work will depend. (President's Scientific Research Board, 1947, p. 3)

During the late 1950s, the science education community became more interested in the strategic role of scientific knowledge in society. This was especially significant because of the then recent launching of the earth orbiting satellite, Sputnik, by the Soviet Union in 1957. In 1960, the NSSE in its Fifty-ninth Yearbook entitled, *Rethinking Science Education*, focused on science education for yet another issue (NSSE, 1960). It was proposed that science educators should work to produce citizens who understood science and were sympathetic to the work of scientists. Not everyone, however, was comfortable with science education being justified on the basis of national security concerns. One member of the National Science Foundation appealed to educators to remain focused on the general, liberal education theme saying:

> Not because there are satellites following their elliptical orbits about the earth, not because other nations have given emphasis to training in technology and science, and not because of any alteration of our scale of values, should it suddenly be declared that science must occupy the commanding position at all levels in our educational system. . . . [W]e live in an environment molded by the application of science, and we believe some of the processes used in arriving at conclusions in science have a relevance to our thinking and, indeed, to our behaviour in other phases of life. Hence, education in science should be a part of the intellectual heritage of all. (NSSE, 1960, p. 24)

As seen from the historic review, science in the 19^{th} century developed from being less noble and worthy than the humanities to the status of a legitimate, intellectual study with a great impact on people's lives in a technologically developing society. After World War II people became aware that scientific developments had the potential to destroy society and therefore, it became important for national security and had a strategic significance. Following the launch of the Sputnik in 1957, however, it became apparent that just as there was a dire need for specialized scientists, there was a cry for science to become a part of the cultural heritage of all. This cry could only be addressed by promoting scientific literacy for one and all.

Scientific Literacy

Many science educators during the late 1950s, however, believed that the goals of science education should be qualitatively different. They reckoned that science teaching should still be for personal development and to help individuals adjust to life in modern society, although the world was changing. "Explosive developments in technology and concerns about national security that arose following World War II were compelling enough to command a new approach to science education" (DeBoer, 2000, p. 586). Within this new environment, the goals of science teaching for general education purposes, were termed *scientific literacy*. In June of 1958, The Rockefeller Brothers Fund (1958) issued a report on the state of education in the U.S. The report focused on how the country should respond to the "startling" rate of scientific and technological change taking place especially in areas such as nuclear energy, space exploration, cell biology, and brain physiology. In the report on education, the challenging question was how the educational system could be used to prepare people more effectively to live and work in such a rapidly changing world.

The answer to meet these challenges, was "to turn to organized intellectual effort as never before in history" (Rockefeller Brothers Fund, 1958, p. 347), with a special focus on the most talented and highly educated members of society. All this was deemed necessary for us to keep pace with the "breathtaking movement into a new technological era" (p. 367). In addition to needing an adequate supply of technically trained scientists, mathematicians, and engineers, the society also needed a highly educated citizenry that understood the scientific enterprise. The report said: "Among the tasks that have increased most frighteningly in complexity is the task of the ordinary citizen who wishes to discharge his civic responsibilities intelligently" (p. 351). The solution to this dilemma was scientific literacy. According to the Board:

> [J]ust as we must insist that every scientist be broadly educated, so we must see to it that every educated person be literate in science. . . . We cannot afford to have our most highly educated people living in intellectual isolation from one another, without even an elementary understanding of each other's intellectual concern. (Rockefeller Brothers Fund, 1958, p. 369)

Scientific literacy, therefore, was to provide a broad understanding of science, as well as of the rapidly advancing scientific enterprise, whether one intended to become a scientist or not.

In 1958, Hurd (1958) first used the term 'scientific literacy' to refer to the new goals of science education in the post-Sputnik era. Scientific literacy, at its simplest level, is a shorthand for "what the general public ought to know about science" (Durant, 1993, p.129). Bybee states:

> The phrase 'scientific literacy for all learners' expresses the major goal of science education – to attain society's aspirations and advance individual development within the context of science and technology. (1997, p. 69)

As early as the 19th century there had been the idea (Layton, 1975) that the general public should have some knowledge of science. Bybee alludes to this notion when he notes:

> The idea of scientific literacy has been a key factor in the formation of school science programs throughout our history. . . . The term has been increasingly used as a short-hand version of the fundamental goal of science education. (1997, p. 46)

If that is the case, then "to speak of scientific literacy is simply to speak of science education itself" (DeBoer, 2000, p.582). In this paper, DeBoer argues against defining scientific literacy narrowly and precisely, asserting that the value of the term lies primarily in its potential to advance arguments that have been a part of the science curriculum debate for many decades. Bybee (1997) also elaborates on the value of the term "scientific literacy" as a slogan, because it has become "a rallying cry for contemporary reform; it serves to unite science educators behind a single statement representing the purposes of science education" (p. 71).

In the International Encyclopedia of Education, Jenkins (1994) suggests that "[scientific literacy] usually implies an appreciation of the nature, aims and general limitations of science, coupled with some understanding of the more important scientific ideas" (p. 5346). Thomas and Durant undertook a more detailed analysis, and identified eight characteristics of scientific literacy from the literature which are listed in Figure 2-1 (1989, p. 1-14). This list includes several items (like numbers 2, 3, 6 etc...) that receive little attention in most school science curricula, alongside some (like number 5) that are invariably included. On the whole, the items listed in Figure 2-1, have striking similarities with the curriculum objectives proposed by advocates of a Science, Technology and Society (STS) approach (Solomon, 1993). Aikenhead notes (1994) that there are many varieties of STS courses, and that in emphasizing the importance of some understanding of important science ideas, 'science literacy' courses are likely to belong to the category that Aikenhead labels "Science through STS content".

Figure 2-1. **Characteristics of scientific literacy**

Characteristic	Description
1	An appreciation of the nature, aims and limitations of science; a grasp of 'the scientific approach' - rational argument, the ability to generalize, systematize and extrapolate; the roles of theory and observation.
2	An appreciation of the nature, aims and limitations of technology, and of how these differ from those of science.
3	A knowledge of the way in which science and technology actually work, including the funding of research, the conventions of scientific practice, and the relationships between research and development.
4	An appreciation of the inter-relationships between science, technology, and society, including the role of scientists and technicians in society and the structure of relevant decision-making processes.
5	A general grounding in the language and some of the key constructs of science.
6	A basic grasp of how to interpret numerical data, especially relating to probability and statistics.
7	The ability to assimilate and use technical information and the products of technology: 'user-competence' in relation to technologically advance products.
8	Some idea of where or from whom to seek information and advice about matters relating to science and technology.

Norris and Phillips also look at the range of conceptions of scientific literacy in the science education literature, and draw up a list summarized as Figure 2-2 (2003, p. 225). Although this compendium has considerable overlap with Thomas and Durant's list of characteristics, there are also some differences of emphasis, and a stronger sense of what scientific literacy equips the learner to do (eg. participate in the discussion of issues;

continue learning science after school; experience curiosity and wonder about the natural world). Norris and Phillips then make a powerful case that scientific literacy must be grounded in the fundamental sense of literacy which is the ability to analyse and interpret text. They argue that science could not exist as an oral tradition: texts are essential, not optional. They are a characteristic feature of science - just as empirical data collection is vital for research in science. An understanding of science, therefore, requires the ability to read texts. Literacy, then, is at the core of scientific literacy.

According to Norris and Phillips (2003) there is a clear distinction between the *fundamental* and *derived* senses of scientific literacy. They referred to reading and writing when the content is science as the fundamental sense of scientific literacy, while being knowledgeable, learned and educated in science was termed the derived sense. Though there may be many ways to be scientifically literate (DeBoer, 2000), as discussed above, Norris and Phillips (2003) maintained that at a minimum this path to literacy "must intersect scientific literacy in both senses" (p. 230). They argued that "reading and writing are constitutive parts of science" (p. 226). They defended their argument by stating:

> Constitutive relationships define necessities because the constituents are essential elements of the whole. Remove a constituent, and the whole goes with it. Throw away the cover, and you still have a book: throw away the contents and keep the cover, and you no longer have a book. (2003, p. 226)

If evidence is to be accurately identified, understood, and interpreted, it is imperative that the reader has a high degree of literacy in the fundamental sense of the word.

Figure 2-2. **Conceptions of 'scientific literacy' in the science education literature**

Conception	*Description*
a)	Understanding of basic scientific ideas
b)	Understanding science and its applications
c)	Knowledge of what counts as science; the ability to distinguish science from non-science
d)	Ability and wish to be an independent, lifelong science learner
e)	Ability to use science knowledge in problem solving
f)	Knowledge needed for intelligent participation in science-based social issues
g)	Understanding of the nature of science
h)	Appreciation of, and comfort with, science including a sense of wonder and curiosity
i)	Knowledge of the risks and benefits of science
j)	Ability to think critically about science and to deal with scientific expertise

If evidence is to be accurately identified, understood, and interpreted, it is imperative that the reader has a high degree of literacy in the fundamental sense of the word.

In their definition of "scientific literacy," in the *National Science Education Standard,* the National Research Council, (NRC), lay an emphasis on reading text with understanding:

Scientific literacy means that a person can ask, find, or determine answers to questions derived from curiosity about everyday experiences. It means that such a person has the ability to describe, explain, and predict natural phenomena. Scientific literacy entails being able to read with understanding articles about science in the popular press and to engage in social conversation about the validity of the conclusions. Scientific literacy implies that a person can identify scientific issues underlying national and local decisions and express opinions that are scientifically and technologically informed. A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately. (NRC, 1996, p. 22)

The subsequent paragraphs in the *Standards* (NRC, 1996) disclose that scientific literacy is a continuum with a range of levels, that an individual's scientific literacy might vary across the sciences, and that the development of scientific literacy is a lifelong task. The idea of scientific literacy having various levels is explored by Bybee (1997), who endorsed a 'framework for scientific literacy' with four stages: beginning from *nominal* literacy, through *functional* literacy, and *conceptual* and *procedural* literacy, to *multidimensional* literacy.

Although scientific literacy has been laid down as a continuum, there are critics who have questioned whether scientific literacy, seen in these terms is a feasible goal for all students (Shamos, 1995). In his provocative book, Morris Shamos, a physicist and science educator of very broad experience, argued that universal scientific literacy is a futile goal, and urged a critical review of the purpose of general education in science. He was of the opinion that a meaningful scientific literacy cannot be achieved in the first place, and that the attempt to do so was a gross misuse of human resources. He was skeptical about forecasts of critical shortfalls in scientific manpower, and of crash programs to get more young people into the science pipeline. Rather than giving children the heavy diet of scientific terms and facts that they now get in the present science education, he advocated an appreciation of science as an ongoing cultural enterprise; an awareness of technology's impact on one's personal health, safety, and surroundings; and the need to use experts wisely in resolving science/society issues. He strongly believed that it was impossible to get every American to learn enough science to make independent judgments about major scientific issues, and therefore, it was easier to resort to the advice of science experts in resolving socio-scientific issues.

The goals of scientific literacy are so broad that the way in which these aspirations are interpreted at a more detailed and specific level may show a high degree of variance. The characteristic feature of this endeavor for scientific literacy is that the focus here is on "the citizen," and the ways in which he or she encounters science in the course of their daily life, rather than on the more specialized contexts of particular scientific or technical work. Since our society is deeply influenced and moulded by the ideas and values of science, ordinary citizens need to live and act with reasonable comfort and confidence, so that they do not feel excluded from the whole discourse of science, and thereby do not feel marginalized.

In a technologically-advanced society that is rapidly undergoing change, effort must be taken not only to train the most talented and highly-educated members of society, but also to develop an understanding of scientific enterprises among a highly-educated citizenry who pursue non-science related careers. This can only be done by promoting scientific literacy for all learners: and this should be the major goal of science education. Fundamental literacy should lay the foundation for scientific literacy. Although the goals of scientific literacy are broad, the focus should be on the citizen so that she/he is not marginalized, due to an inability to cope with science and its applications encountered in the course of everyday life. To enable citizens to live with comfort and confidence in a world dominated by science, it is important to introduce scientific literacy through the science courses adopted in the school system especially at the high school and undergraduate level.

Scientific Literacy in the Twenty First Century Science Course

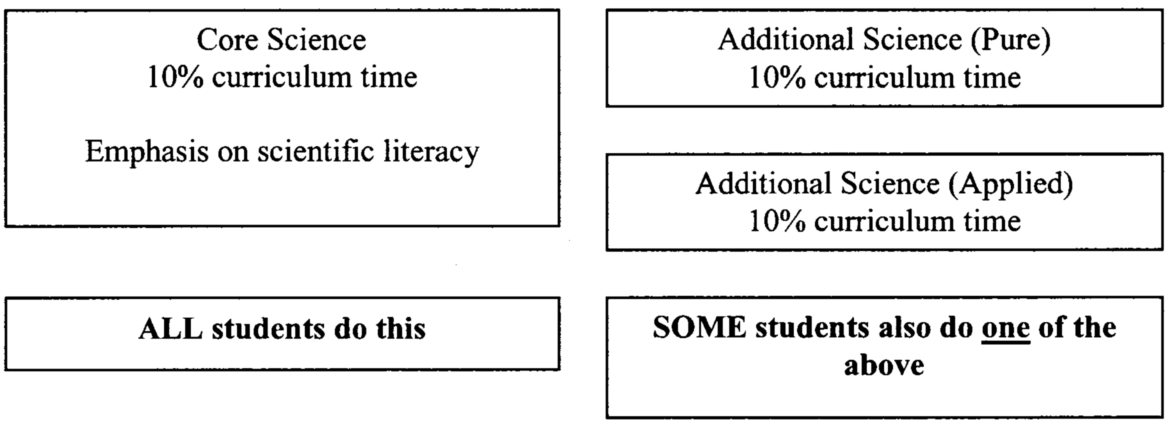
Despite the fact that scientific literacy for all students is widely heralded as an important goal for science education, the science curriculum has another important agenda that it must satisfy. It has to provide the foundational training in science for those students who might pursue a career in science or require more advanced knowledge of science. How the tension between these two different purposes of school science can be resolved is highlighted in the *Beyond 2000* report (Millar and Osborne, 1998) and elsewhere (Norris, 1983). Many curricula have tried to incorporate both purposes of school science, but these attempts have been characterized by the pre-professional training emphasis invariably getting to play a dominant and distorting role in the whole curriculum. The Twenty First Century Science curriculum model (Millar, 2006) is distinct in that it separates these two purposes and tries to address them individually. It achieves this by dividing the curriculum time allocation for science (20% is the norm for students aged 15-16 in England) into two equal parts (Millar, forthcoming, 2006). All students are required to take a Core Science course, which is designed specifically to develop their scientific literacy. Along with this, they have the choice of taking an Additional Science course, which may be offered with either a "pure" or "applied" emphasis (Figure 2-3). Both the Core and Additional Science courses are of the same duration as those offered to students in other subjects leading to the General Certificate of Secondary Education (GCSE) qualification. The Core-plus-Additional courses provide a sound foundation for those progressing to more advanced courses in the sciences at the 16-and-over age group. In this way, the demand of some stakeholders (such as the scientific community, employers and the government), to give priority to school science in providing a firm foundation for those interested in more advanced study of the sciences, was adequately met. The Core science course, on the other hand, was developed for the majority of students, with scientific literacy as its primary aim. This scheme of possible course offerings was intended to adequately resolve the tension between the two different purposes of school science. How well a science curriculum serves the students who follow it, is best measured by evaluating the state of school science, and that is what is examined next.

The State of School Science

The state of science education in schools has been the topic of much discussion recently (Millar, 1996; Millar and Osborne, 1998; Osborne, Driver and Simon, 1998), because of a recognition that not all is well. There are two issues that are of particular concern to science educators:

- Pupils' *attitudes* towards science as a school subject.
- Pupils' *understanding* of the ideas of science.

Figure 2-3. **The *Twenty First Century Science* curriculum model**



A strong indicator of the attitudes of pupils is their choice of subjects. The proportion of A-level students taking science, particularly physical science remains a cause for concern (Osborne, Driver and Simon, 1998). "For whatever reasons, large numbers of pupils, particularly girls, are rejecting science as soon as they get the chance" (Gott and Johnson, 1999). The Third International Mathematics and Science Study (TIMSS), carried out during 1994-1995, provided information on the performance of Canadian elementary and secondary students in various grades on international mathematics and science tests. This study showed that students' interest and performance in mathematics and science decline through elementary and secondary school, and participation shows a decline in high school (Bordt et al., 2001). “Interest in mathematics and science declines between Grade 4 and Grade 8 and continues to drop during high school” (Bordt et al., 2001, p. 9). In 1995, only 42% of students were taking both these courses in their last year of high school (Bordt et al., 2001, p. 12). Most students find them ‘difficult’ or ‘boring’. “Even when they have done well in mathematics and science in the past, and believe that the subjects are important to them if they want to succeed in life, many students are unwilling to pursue them”(Bordt et al., 2001, p. 9).

Some say that the contexts we use for teaching science are inappropriate and irrelevant. However, there is little evidence to suggest that ‘more relevant’ contexts lead to a better understanding of ideas, although this may lead to a higher motivation. There seems to be an underlying problem here that has not been addressed adequately: the teaching of science is flawed. There appears to be a flaw at the very heart of what we are doing as science educators, and our pupils are not being given the chance to understand the science that we teach.

Several research studies have shown that there is a poor match between the scientific content taught in high school and university science courses and the type of success required in science-based occupations (Duggan and Gott, 2002). Duggan and Gott discovered that in all the industrial, science-rich workplaces they investigated, most of the scientific conceptual understanding used by employees was learned on the job and not in high school or university courses. In hospitals across Europe and North America, evidence-based practice now dominates educational and professional development programs in the nursing field (Bonell, 1999; Closs and Cheater, 1999). In a recent study, Aikenhead (2004) investigated what science-related knowledge was actually used by nurses in their day-to-day clinical reasoning as they attended their patients, thereby extending Duggan and Gott's (2002) research program into a science-rich occupation that had not been previously investigated. He looked at the "knowledge-in-use" employed by six acute-care nurses in a hospital surgical unit. The three questions Aikenhead wished to address through his study (2004) were:

- What knowledge-in-use comprises canonical science content found in science curricula ("scientific knowledge)?
- What knowledge-in-use is associated with the technical field of nursing ("professional knowledge of nursing")? and
- Does a nurse's knowledge-in-use include a core set of concepts of evidence similar to that observed earlier in industry?

From the results of his study Aikenhead was able to establish that as all six nurses attended to the data gathered from their patients, they utilized a core set of concepts of evidence. There were some CoEs related to reliability that the nurses used which were missing from Gott and colleagues' (2003) compendium such as: normalcy range, uniqueness of the patient measured, and variability within the patient's physical attributes. Conversely, the nurses seldom used key CoEs such as repeated readings and measurement error. These would seem to be CoEs irrelevant to nursing. The above-mentioned CoEs all deal with the physical attributes of patients. However, missing from Gott, Duggan, Roberts and Hussain's (2003) compendium, but apparent in the surgical unit, is a set of emotion-related CoEs, associated with psychology, sociology, and anthropology (e.g. cultural sensitivity). These emotion-related observations were often cited by nurses as evidence from which to make clinical decisions as to how to improve the condition and comfort of a patient. The nurses' CoEs served two functions: (a) taking the practice to the next level and, (b) initiating a procedure for intervention. Before engaging in either of these two procedures, nurses had to evaluate their observations (especially their credibility) and decide whether they were sufficient or insufficient. Both contexts existed for the main purpose of healing patients. A detailed explanation of specific, emotion-related CoEs, however, requires further investigation.

Remediating this weakness requires a reconsideration of the aims and purposes of science education. It is my contention that an understanding of scientific evidence and its relationship to scientific facts, concepts and theories is more valuable in enabling and empowering citizens to use science in their lives than simply knowledge of the facts, concepts and theories themselves. Since evidence is the centre point of any controversy in science, the point at which science interacts with people's lives, it relies on the participants being familiar with the language of evidence. Understanding scientific evidence requires a holistic appreciation of science, and teaching designed to increase the understanding of evidence needs to engage pupils fully with scientific reasoning. As an additional benefit, I conjecture that if evidence is seen to be useful and functional in everyday life, the image and popularity of science as a school subject is likely to improve. It is with this hope in mind, that I am examining for the first time how evidence is represented in chemistry textbooks used at the university undergraduate level in Alberta, and at the high school level right across Canada.

Science Textbooks

In the English-speaking world, the first widely-used textbooks in elementary physics were written by William Whewell (Stinner, 1992). He argued that students should learn concepts outside the grip of mathematical formulation. Whewell was of the opinion that if students did not struggle through the appropriate arguments based on intuition, space and geometry first, they would only "learn to reason by means of symbols …; and by means of the general rules of combining and operating upon such symbols; without thinking of anything but these rules" (Whewell, 1850). The teaching of science in general, however, has been a textbook-centered affair since Whewell's textbooks appeared in the 1820s. There is evidence that "many students studying science see little connection between their ideas about the world and what they learn in science textbooks" (Aufshnaiter, 1989).

According to Thomas Kuhn, textbooks are the "pedagogic vehicles for the perpetuation of normal science" (Kuhn, 1962). He believed that scientists "never learn concepts, laws and theories in the abstract by themselves" (p. 187), but through the study of the application of a theory to some concrete range of natural phenomena. He maintains that students of physics learn physics by studying specific applications and concrete examples, which he calls "exemplars". These exemplars may be the problem-solutions that students encounter from the beginning of their scientific education in various forms, either in laboratories, on examinations, or at the end-of-chapter questions in science textbooks. Kuhn agrees that textbook-centered science teaching has been very successful in producing adept scientists for research and technology. However, he has misgivings about its effectiveness in producing the kind of high-grade thinking required to examine the foundations of a science from time to time:

> ... But for normal scientific work, for puzzle-solving within the tradition that the textbook defines, the scientist is almost perfectly equipped ... Even though normal crises are probably reflected in less rigid educational practice, scientific training is not well designed to produce the man who will easily discover a fresh approach (Kuhn, 1962, p. 166).

Kuhn further asserts that the textbook does not question the presuppositions of a science and, in fact, it "systematically disguises" the history of its discipline (Kuhn, 1962, p. 136).

The recommendations that Whewell and Kuhn make, present two opposing views. Whewell believed in frequent contact with the evidential plane in the teaching of physics. To achieve this, he advocated the "explication of concepts by appealing to the student's experience and intuition prior to the final mathematical formulation (Whewell, 1850). Kuhn, on the other hand, believed that students learn science from exemplars, i.e. students "by doing problems would learn consequential things about nature" (Kuhn, 1962). He seems to believe that because textbooks have been so successful in producing competent "puzzle-solvers" in normal science, i.e. the working scientist, textbook-centred teaching must also be successful in teaching science in general, namely the science student.

Stinner (1992) thought that Kuhn is right in that ideally, by learning the exemplars of a science, students should be able to make contact with an evidential base in a way that "makes sense" to them. Most students, however, failed to make this contact, and the normal pattern of science learning is memorization and algorithm recitation (Stinner, 1992). One of the reasons science teaching failed to help students make contact with appropriate evidence may be the science teachers' inadequate background knowledge. Many teachers do not have the self-confidence to make frequent connections with the evidential-experiential plane. Another important reason must be the fact that not enough attention is paid to how students learn science concepts. In spite of their diametrically opposed views on how science ought to be taught, it is evident from what is discussed above, that both Whewell and Kuhn are deeply concerned about how students learn concepts in science, and both value scientific literacy as one of the prime goals of science education.

Science textbooks, which play a major role in science instruction throughout Canada, could not be understood without its readers being literate and knowing how to handle science text in its various forms. Norris and Phillips brought this to the fore when they stated succinctly and forcibly:

> Science is in part constituted by texts and by our means of dealing with them. Without the expressive power and relative fixity of text; and without the comprehension, interpretive, analytical, and critical capacities we have developed for dealing with texts; then western science as we know it could never have come into being. (2003, p. 233)

Science textbooks, therefore, have a tremendous influence on how students perceive the scientific enterprise (DiGisi & Willett, 1995; Scruggs, 1988; Yager, 1983). These teaching aids are used frequently and widely in science classrooms across the nation. In many science classrooms, the majority of the instructional support, beyond the teacher, is provided by the prescribed textbooks and consequently science textbooks contain much of the scientific information students receive (Mayer, 1983). Some researchers believe that in the majority of classrooms, "the textbook serves as the ultimate source of knowledge" (Chiappetta, Sethna & Fillman, 1993), and in many cases "actually becomes the curriculum" (Stake & Easley, 1978).

Although there is little question regarding the importance of textbooks in science teaching (Harns & Yager, 1981), the problems associated with science textbooks are numerous. Since publishers have to satisfy the guidelines of the science curriculum committees of many states and provinces, textbooks end up covering far too much subject matter and may do so in a superficial manner (Chiappetta, Sethna & Fillman, 1993). Ruis (1988) pointed out, that chemistry textbooks often do an inadequate job of explaining important concepts and principles, because of their superficial treatment of the subject matter. These textbooks often introduce ideas, but do not develop them adequately around the models that scientists used to form these concepts originally. Besides, science textbooks often fail to present important topics (such as evolution) in a thorough manner because they are afraid of offending special interest groups.

Chemistry textbooks, as well as all science textbooks, should make science interesting, relevant and understandable to students without succumbing to the temptation of diluting the subject matter to the point where it lacks any meaning (Bucat & Cole, 1988). These teaching aids should not become mere vocabulary books "with more new terminology than is found in foreign-language textbooks" (Yager, 1983). This abundance of new terminology – which means students have to memorize a lot of new words that mean very little to them - is a factor that turns many students off science. Instead of helping the learner to understand ideas through new ways of explaining and presenting written information, textbook publishers have gone to extreme lengths to include many colored pictures in order to allure teachers into adopting their textbooks (Ruis, 1988). Publishers are, in addition, providing voluminous resource materials such as laboratory exercises, worksheets, and review sheets to encourage adoption of their textbooks. In spite of all these attempts to get chemistry textbooks adopted by school districts, little has been done to interest more students into taking chemistry in high school. As a result the full-time undergraduate enrollment in Chemistry is dropping (Statistics Canada, 2000). Some of the major barriers to getting youth more interested in science are: lack of appeal of science, the difficulty of science subjects, and the fact that youth are not interested in science (EKOS Research Associates Inc., 2004).

The Science Council of Canada (Orpwood & Souque, 1984) carried out a large-scale study to examine the contents and aims of science textbooks used in Canada. The themes they used to examine the textbooks were those found in the Ministry of Education guidelines, which are related to science content, the acquisition of scientific skills and the interrelationship between science and society. The Council examined 64 textbooks in the elementary, middle and senior high schools. One of their findings was that the principal way in which the aims of a science program in Canada are implemented in practice is through the use of a textbook. The survey indicated that the teachers are generally satisfied with their textbooks, finding them well illustrated, easy to read, and generally well suited to the intellectual maturity of their students. Teachers felt the textbooks had little to offer the slowest student, but did meet the requirements of the bright students. The fact that many of the textbooks did not contain Canadian examples or descriptions of Canadian science applications was a cause for concern among the teachers. The teachers were of the opinion that the science textbook objectives corresponded adequately to their own priorities.

Garcia (1985) analyzed earth science textbooks for their representation of various aspects of scientific literacy. She selected scientific literacy as the major theme of her content analysis "because of its broad.conceptual framework for the outcomes in science education". She looked at the work of many science education researchers and organizations in order to formulate broad and discrete categories of scientific literacy. Among the works on scientific literacy which she analyzed were: Pella, O'Hearn and Gale (1966); Showalter (1974); Harms and Yager (1981); NSTA (1982); Roberts (1983); Fensham (1983); Orpwood and Alam (1984); and Collette and Chiappetta (1986). From the numerous works she examined, Garcia (1985) identified many descriptors, each of which was placed on a card. The cards were given to two science educators to categorize using a modified Q-sort procedure described by Rakow (1985). This procedure led to the identification of *four* distinct categories of scientific literacy. These were: (a) The basic knowledge of science, (b) the investigative nature of science, (c) the thinking processes of science, and (d) the interaction of science, technology and society.

Chiappetta, Fillman and Sethna (1991) found that Garcia's (1985) descriptors, which were used to analyze only earth science textbooks, could be elaborated so that the written text that appears in a wide variety of science textbooks, from various branches of science, could be properly categorized. Chiappetta et al (1991) found descriptors which had a high rate of recognition for the four major themes. This called for many iterations of analyzing a large variety of science textbooks, and resulted in the construction of a 25-page training manual (Chiappetta, Fillman, & Sethna, 1991). The four major themes of scientific literacy and their descriptors, as they appear in the training manual are as follows:

1. *The knowledge of science.* This category is involved if the intent of the text is to *present*, *discuss* or *ask* the student to recall information, facts, concepts, principles, laws, theories, etc. It reflects the transmission of scientific knowledge where the students receives information. This category typifies most textbooks and presents information to be learned by the reader. Textbook material in this category: (a) presents facts, concepts, principles and laws; (b) presents hypotheses, theories, and models; and (c) asks students to recall knowledge or information.

2. *The investigative nature of science.* This category is applicable when the intent of the text is to *stimulate thinking* and *doing* by asking the students to "find out." It includes the active aspect of inquiry and learning, which involves the student in the methods and processes of science such as observing, measuring, classifying, inferring, recording data, making calculations, experimenting, etc...This type of instruction is a combination of paper and pencil as well as hands-on activities. Textbook material in this category: (a) requires students to answer a question through the use of materials; (b) requires students to answer a question through the use of charts, tables, etc...; (c) requires students to make a calculation; (d) requires students to reason out an answer; and (e) engages students in a thought experiment or activity.

3. *Science as a way of thinking.* This category applies if the intent of the text is to show

how science in general or a certain scientist in particular, went about "finding out." This aspect of the nature of science represents thinking, reasoning, and reflection, where the student is told about how the scientific enterprise operates. Textbook material in this category: (a) describes how a scientist experimented; (b) shows the historical development of an idea; (c) emphasizes the empirical nature and objectivity of science; (d) illustrates the use of assumptions; (e) shows how science proceeds by inductive and deductive reasoning; (f) gives cause and effect relationships; (g) discusses evidence and proof; and (h) presents the scientific method and problem solving.

4. *Interaction of science, technology and society.* This category is used if the intent of the text is to illustrate the *effects* or *impacts* of science on society. This aspect of scientific literacy deals with the application of science and how technology helps or hinders humankind. It includes, in addition, social issues and careers. The student usually receives this information and generally does not have to find out. In this category textbook material: (a) describes the usefulness of science and technology to society; (b) points out the negative effects of science and technology on society; (c) discusses social issues related to science and technology; and (d) mentions careers and jobs in scientific and technological fields.

The work of Chiappetta et al. (1991) helped to identify four major themes of scientific literacy which could be applied to science textbooks in all branches of science. Subsequently, I would like to concentrate on the third theme of scientific literacy which is "Science as a way of thinking." This theme focuses on the thinking, reasoning and reflection involved as scientists try to "find out" and probe the unknown for answers.

Textbook material addressing this theme can be classified under eight broad headings as shown above. Out of these eight headings, the category of prime importance to this study is the one that "discusses evidence and proof."

Evidence

Although this appears to be a simple question, it is one that is not asked in many research contexts, partly because evidence is difficult to define. In a dictionary of philosophy, Mautner (1999, p. 184) says that evidence is "that which provides a ground for a belief or a theory." In another philosophical dictionary, Audi (1999, p. 293), says evidence is "information bearing on the truth or falsity of a proposition." A proposition is a statement, usually in the form of a declarative sentence which contains a "that" phrase either explicitly or implicitly (Miller & Fredericks, 2003). According to Miller and Fredericks (2003):

> A proposition then asserts something is the case and we are trying to show that the "information" we possess can establish the truth or falsity of it; or, in another sense, establish its "probability" "likelihood", "warrantability", or something similar. (p. 6)

Evidence, a concept central to the empirical sciences, is used when referring to data or observations that are put forth to support or refute a scientific hypothesis. Whether to believe in a scientific hypothesis or theory is dependent upon the quantity and character of the evidence in its favour.

This section on evidence begins with an overview of the work done on the understanding of evidence, followed by an account of how data become evidence. It then examines the different types of evidence illustrating them with the work of Heinrich

Hertz and J. J. Thomson on cathode rays. The role of practical work in the understanding of evidence in science is then explored. Substantive knowledge based on facts and procedural knowledge leading from skills are defined and explained as being the two essential arms of problem solving in science education. This leads to concepts of evidence, their classification, definition and categorization. The chapter ends with a section on students' conceptions about evidence and their ease in using it.

The Work Done on Understanding Evidence

Kuhn, Amsel and O'Loughlin (1988) focused on the development in young people of the capacity to engage in *scientific reasoning*. They examined the ability of students at different ages to evaluate given (or self-proposed) theories using given evidence, and the influence that the students' own theories had on this process. In this study, a "theory" meant a claim that a given dependent variable does, or does not covary with a given independent variable. Covariation in this context means that a change in one variable results in an increase (or decrease) in the other. Non-covariation, on the other hand, means that a change in one variable results in no change in the other. Kuhn et al. (1988) introduced students to a problem, elicited their "theories" about it, and then asked them to say whether the given pieces of data supported or conflicted with a particular "theory," and to explain their reasoning. They concluded that the process of the coordination of theory and evidence is subject to developmental change and that many young students do not consider the possibility that their theory might be false or that alternative theories might exist. The ability to separate clearly data and explanation (or "theory") takes time to develop.

A year later, according to Kuhn (1989), as the scientist explores the environment, constructs models to understand it, and revises these models as new evidence is generated, so too lay people strive to make sense of their surroundings by processing data and constructing mental models based on these data. The view that she endorsed in this article is that at the heart of scientific thinking is the coordination of theories and evidence. A central premise underlying science is that scientific theories stand in relation to actual or potential bodies of evidence against which they can be evaluated. The scientist is one who:

a) is able to consciously articulate a theory that he or she accepts,
b) knows what evidence does and could support it, and what evidence does or would contradict it, and
c) is able to justify why the coordination of available theories and evidence has led him or her to accept that theory and reject others purporting to account for the same phenomena.

These skills in coordinating theories and evidence are the most central, essential and general skills that define scientific thinking. As Kuhn (1989) reviewed the research in this area, she was able to demonstrate that the processes of scientific thinking differ significantly in children, lay adults and scientists. She proposed a framework for conceptualizing the development of the scientific thinking process, centering on progressive differentiation and coordination of theory and evidence. This development is metacognitive and requires thinking about theories, rather than merely with them, and thinking about evidence, rather than merely being influenced by it, and it reflects the attainment of control over the interaction of theories and evidence in one's own thinking.

Koslowski (1996) questioned whether the form of reasoning termed by Kuhn et al, as "scientific" really matched the practice of scientists. She argued that the interpretations that people place on data depend on whether they can imagine an underlying mechanism that might account for any patterns observed in the data. If a person can imagine a plausible reason why two variables might covary, then a small amount of data showing no covariation is unlikely to make them reject the hypothesis that the variables covary. On the other hand, if they cannot imagine any plausible mechanism linking two variables, a small amount of data showing covariation is unlikely to persuade them that they are related. In other words, people treat their ideas about relationships between variables as working hypothesis, and modify them when confronted with anomalies in a manner that takes account of *theory and data together.*

One critique of Koslowski's work, however, is that her subjects were not presented with primary data, but with a statement of the "results" of an experiment that was described to them. This method omits one very major step in the process of scientific reasoning – getting from the measured primary data to the statement of "results." By omitting this step, Koslowski was only studying "logical reasoning" in a scientific or quasi-scientific setting, rather than "scientific reasoning." Thus although Koslowski's work challenged and extended the findings of Kuhn et al. in important ways, it did not probe one constitutive feature of scientific reasoning, namely *reasoning from data.*

Science educators have argued for decades that it is insufficient to know about scientific theories, and not know how knowledge claims are justified, what counts as evidence, or how theory and evidence interact. Two decades ago, Millar and Driver (1987) argued that how students actually use science process skills was dependent on their understanding of science content. Similarly, Samarapungavan (1992) has shown that changes in children's reasoning on theory-choice tasks can be accounted for by changes in their understanding of the underlying scientific concepts. In a later work, Driver et al. (1996) grouped students' evidential warrants for belief into three categories: "appeal to direct perceptual evidence, ability to make inferences based on evidence, or technological efficacy as a sufficient warrant for believing a claim." From the student responses Driver et al. studied (1996), they found that as students matured, there was a slight tendency toward less (55-45%) reliance on direct, perceptual evidence, as well as an increased use of sophisticated reasoning about evidence and authority. They also noted that older students were more likely to use evidence and explanation consistently and to evaluate explanations in the light of evidence. Their data are supportive of the idea that students' warrants for belief in scientific ideas and their abilities to coordinate theory and evidence varies with context. They found that students talked differently about evidence and explanation when they were given a story about rusting than they did when given a story about balloons filled with air. Recent work by Ryder et al. (1999) emphasized the same theme in that they found that individual college undergraduate students drew upon a range of views about the nature of science depending on the scientific context being discussed.

The scientific views and religious beliefs of college students were explored by Dagher and BouJaoude (1997) using the sensitive topic of biological evolution. The objective of this study was to investigate how some university biology majors in Beirut, Lebanon, accommodated the theory of biological evolution with their existing religious beliefs. This led to the perusal of two main research questions: (a) How do students accommodate their religious beliefs with their understanding of the theory of biological evolution? and (b) What arguments do they present to justify the positions they espouse? The participants in Dagher and BouJaoude's study consisted of sixty-two undergraduate biology majors enrolled in a required senior seminar, in a private university where English is used as the language of instruction. Each student enrolled responded to open-ended questions that addressed the two major research questions stated above, and answered three essay questions as well, based on the theory of evolution. Then, based on their answers, fifteen students were chosen for an in-depth exploration of their written responses. The students' responses fell under four main categories: for evolution, against evolution, compromise, and neutral. This research was important for two reasons: (a) "to explore how the creation/evolution debate, whether addressed or avoided in the formal teaching of evolution, takes place in the students' thinking; and (b) to go beyond outlining the various positions of the debate to discuss their implications to curriculum and instruction" (Dagher and BouJaoude, 1997). Since this study reveals how students' beliefs systems interact with science learning, it has the potential to enable educators to be more mindful of, and skillful at easing, the difficulties their students experience when they are exposed to a scientific theory that has the potential of conflicting with their other beliefs. It is of vital importance that epistemological questions about the nature of scientific facts, laws, hypotheses, theories and evidence are addressed in all science topics, especially is sensitive topics such as evolution. Dagher and BouJaoude contended that a lot of the confusion that comes about in the evolution/creation debate arises from confounding the everyday with the scientific use of these terms. Therefore, a deliberate discussion of these terms, and their characteristics in specific contexts should clarify the specialized meanings associated with them. These type of planned discussions would aid students to "achieve a better conceptual understanding of the theories involved," and would also better enable them to "be conversant in the metalanguage of science" (Norris & Phillips, 1994).

In Dagher and BouJaoude's (1997) work, described earlier, some students demanded 100% proof that the evolution theory was true, or discounted the theory because it was a theory and not a law, or because it relies on probabilities and not on certainties. In doing so, they were in fact expressing their criteria for a proof: for a proof to be adequate, it has to be directly observed and replicated. One student, who had a good understanding of the evolution theory, believed that the evidence presented by the scientists did not justify his accepting the theory and abandoning the religious accounts of creation. This case illustrated that good understanding of the evidence did not necessarily lead to the acceptance of the theory. It also highlighted how religious beliefs can affect the way in which people view and value evidence. They also found that neutrality of instruction did not help students resolve their conflicts between their own religious beliefs and the evolution theory.

Brickhouse et al. (2002) further explored how student views about the nature of science vary with science content. They particularly considered the following areas: distinction between explanation and evidence, warrants for belief or for skepticism, and nature of observation. This study drew on data collected in an university astronomy course. The data collected included three interviews and written work from twenty students, as well as written work from the 340 students in the class. The conclusions drawn from this study suggested that students' talk about the nature of science differed depending upon the particular scientific topic under discussion. The relationship between theory and evidence, warrants for belief, and nature of evidence are described in different ways in the different topics discussed in the course. These differences have several implications for science education researchers and teachers. Researchers should be cautious in making claims about students having any particular view of "the nature of science" which is independent of the subject under discussion. In addition, it follows that attempts to teach the nature of science should be thoroughly embedded in particular science content.

The work by Brickhouse et al (2002), which included Dagher from the previous study, showed that students' abilities to coordinate theory and evidence varied with context. For example, students talked differently about evidence and explanation when discussing gravity than when they talked about examples from astronomy. The use to which evidence is put also varies according to content. For example, in astronomy, scientists observe and use the evidence they obtain to learn new things; whereas in evolution, evidence is collected to test the theory. The authors noted a development in the students' ability to demand evidence; and in differentiating between valid and invalid evidence.

More recently cognitive scientists, Masnick and Morris (2002) asked subjects (spanning a wide age range from U.S. third grade students to college undergraduates) to interpret data sets that varied in sample size, in the direction of difference between measurement pairs, and in variability about the mean. Their work showed that students at all ages had greater confidence in conclusions drawn from larger data sets. Similar methods have also been employed by science education researchers, Allie et al. (1998), who reported an investigation of the understanding of measurement and numerical data of first-year undergraduate physics students in South Africa. One of their probes asked subjects to draw conclusions and justify them from two sets of five measurements with the same mean but different variability around the mean. A second probe had two sets of five measurements with different means but similar variability around the mean. Allie et al. (1998) concluded that their subjects' responses are in agreement with the progression model of Lubben and Millar (1996). In their model, Lubben and Millar proposed a progression of ideas concerning the reliability of experimental data based on a study with UK secondary school students. They were able to identify a pattern of understanding with age and experience. For example, they were able to identify a progression in understanding of the *process of measuring* going from the denial of the need to repeat measurements, via a search for recurring results, and a deliberate variation of control variables to collect a guaranteed variety of results, to the determination of the likely range of results, initially possibly done as science class routine. Measurement is a concept of evidence that has a wide range of applications across the different branches of science. Parallel progressions were identified for the *evaluation of measurements* and about *what to do with 'anomalous results.'*

Kanari and Millar (2004) explored the understandings of data and measurements that school students, in the 10-14 years age range, draw upon, and the ways that they reason from data when carrying out a practical science inquiry task. The two practical tasks employed in the study each involved investigations of the relationship between two independent variables (IVs) and a dependent variable (DV). In both tasks, one IV covaried with the DV, whereas the other did not. This study was partly designed to test the hypothesis: *"Students will experience greater difficulty, and will be less successful in* *reaching the correct conclusion, in investigations of a variable that does not affect an outcome than of a variable that does have an effect"* (Kanari and Millar, 2004). The students' actions were video-recorded for analysis. Each student was subsequently interviewed and asked to discuss and interpret data collected by two other students, undertaking a similar task, shown on a video-recording. An analysis of the students' performance on the practical tasks and their interview responses showed few differences across task contexts or with age in the students'reasoning. The most remarkable finding of this study was that students of all ages had a much lower success rate in investigations where the dependent variable did not covary with the independent variable, than in those where it did covary. Their results showed that it was only a minority of students who could be said to have developed their awareness of the variability of measurements into a notion, however embryonic, of "measurement uncertainty" or "error." The findings of this study have implications for both research and teaching. This could be seen as a good point of entry for teaching concepts of evidence such as uncertainty entailed in measurements, the need to repeat measurements and the sources of error, both random and systematic. The basic competence in carrying out investigations of the relationship between variables that clearly covary can be acquired by many students by the age of ten, if the curriculum places the emphasis on this. The understandings needed to extend this capability to more difficult situations, however, do not develop spontaneously and need to be taught explicitly using carefully chosen examples. Kanari and Millar recommended that school science investigations should include both covariation and non-covariation cases. Non-covariation cases are of special value for research into scientific reasoning because they draw attention to features of students' thinking that remain hidden if only covariation cases are used.

How does Data become Evidence?

A fact that is often overlooked is: data are not evidence. Data must be linked to a conclusion by theory and argument before they can become evidence. If data is to become evidence, therefore, something must happen to it. "We may have data, but they only become evidence when (and if) we have some sort of model for when and how this takes place" (Miller & Fredericks, 2003). However, such models are varied and complex, and these factors make it difficult to often clearly demarcate what evidence is, not only for the qualitative case, but for any case. Achinstein, often considered to be the "Dean" of theorists on evidence, presents a comprehensive and exhaustive account of such theories (2001). My intent here is to utilize some of his general perspectives on the different types of evidence and on what makes evidence, evidence.

The following examples drawn from the work of Achinstein (2001), illustrate the complexity of this concept known as "evidence." In describing these examples, I will use the conventional notation of discussions of theories of evidence where "e" refers to evidence and "h" refers either broadly or specifically, to a hypothesis, claim, assertion, interpretation, or theory of some kind. The issue is clearly stated by Achinstein as follows:

> When the claim is made that something is evidence that a hypothesis is true, what exactly is being claimed? Is there some unique concept of evidence by reference to which we can understand what is being said? I will argue that there is not one

concept of evidence but several in use in the sciences. (2001, p. 13)

The term "sciences" can refer to any field where considerations of evidence are relevant.

Types of Evidence

What Achinstein's (2001) tried to develop are prerequisites for trying to understand what it means to say that one has evidence and that the evidence stands in some relationship to the claim. His earlier analysis (1983) laid down the groundwork for three types of evidence; Potential, Veridical and Subjective. Only the general framework of these three categories of evidence will be discussed, in turn, here.

Considering *potential evidence* first, "e" can be potential evidence that "h," even if "h" is false. Secondly, potential evidence is objective in that "h" does not depend upon anyone's beliefs about "e" or "h," or their relationship. Although theoretically speaking for potential evidence, "e" as well as "h" can be false, in practice however, when a concept of potential evidence is in use, it usually requires that "e" be true. Potential evidence, therefore, does not require "h" to be true, but does require "e" to be so.

Considering *veridical evidence,* "e" is veridical evidence that "h," only if "e" is potential evidence "h" and "h" is true. Veridical evidence requires not just that "h" and "e" both be true, but that "e's" truth be related in an appropriate manner to "h's". Putting it more general terms, there should be an explanatory connection between e's being true and h's being true. In other words, "e" is veridical evidence that "h" if and only if "e" is potential evidence that "h"; "h" is true; and there is an explanatory connection between "e's" being true and "h's" being true.

Subjective evidence is not evidence that is non-objective, or in some way inferior: it is simply a condition of evidence which has, as its defining characteristic, the fact that it is someone's evidence, whether it be an individual's or a group's. Subjective evidence is considered in the context of time, e.g., it is X's evidence at time *t*. In order for evidence to be "subjective evidence," it must, according to Achinstein (1983, p.23), fulfill the following conditions:

(1) X believes that "e" is evidence that "h";

(2) X believes "h" is true and probable; and

(3) X's reason for believing that "h" is true or probable is that "e" is true.

Subjective evidence, therefore, does not require "e" to be true, only that there are good reasons for so believing.

In order to illustrate the subtle differences between these three types of evidence, I would like to use the work of Heinrich Hertz and J. J. Thomson (Achinstein, 2001; Buchwald, 1994) with electric waves in the early part of the 19th century. The wave theorists argued against the particle theorists that, if cathode rays were negatively charged particles, then they should demonstrate electrical as well as magnetic effects. In 1833 Hertz conducted a series of experiments designed to find out whether electrical effects could be observed. In conducting these experiments he was trying to answer two questions:

- Do the cathode rays give rise to electrostatic forces in their neighbourhood?
- In their course, are cathode rays affected by external electrostatic forces?

To answer these questions Hertz employed a special piece of apparatus shown as Figure 2-4. This apparatus consisted of a glass tube, 250mm long and 25 mm wide, that contained an anode and a cathode. The anode consisted of three parts: (i) a brass tube that almost completely surrounded the cathode but had opposite it a circular opening 10mm in diameter through which the cathode rays passed; (ii) a wire gauze about 1 sq mm in mesh through which the cathode rays also passed; (iii) a protective metallic case which surrounded most of the cathode tube. This metallic case and metallic mantle which surrounded the entire apparatus were connected to an electrometer, which showed a deflection if electricity was present.

When the cathode in the above apparatus was connected to a source of electricity, cathode rays were produced which travelled from the cathode to the end of the tube and caused a green phosphorescence to appear on the glass. It is, however, important to bear in mind that there is also the ordinary electric current that flows from the cathode to the anode, which Hertz tried to separate from the cathode rays by using the wire gauze. This wire gauze captured the current, allowing what Hertz took to be pure cathode rays, unmixed with current, to enter the electrometer. If these "pure" cathode rays carried a charge, this would have been registered by the electrometer. In this experiment when cathode rays were produced and the electrometer was connected to the apparatus, the needle of the electrometer showed no deflection and remained at rest. Hertz, therefore, concluded: "As far as the accuracy of the experiment allows, we can conclude with certainty that no electrostatic effect due to the cathode rays can be perceived" (Hertz, 1896, p.249).

Another experiment was designed to determine an answer to the second question, i.e. whether cathode rays were affected by external electrostatic forces. In this experiment Hertz placed a cathode tube similar to the one in the first experiment between oppositely electrified plates. Hertz reported that "no effect could be observed in the phosphorescent image" (p. 252). If these cathode rays had been electrically charged they would have been deflected by these plates, and this would have resulted in a changed position of the phosphorescence. From both these experiments Hertz concluded: "these cathode rays are electrically indifferent, and amongst known agents the phenomenon most nearly allied to them is light" (p.254).

e = In Hertz's experiments, which were designed to show electrical effects of cathode rays, no such effects were produced.

h = Cathode rays are not electrically charged.

Hertz was making the following evidential claim: *e* is evidence that *h* is true.

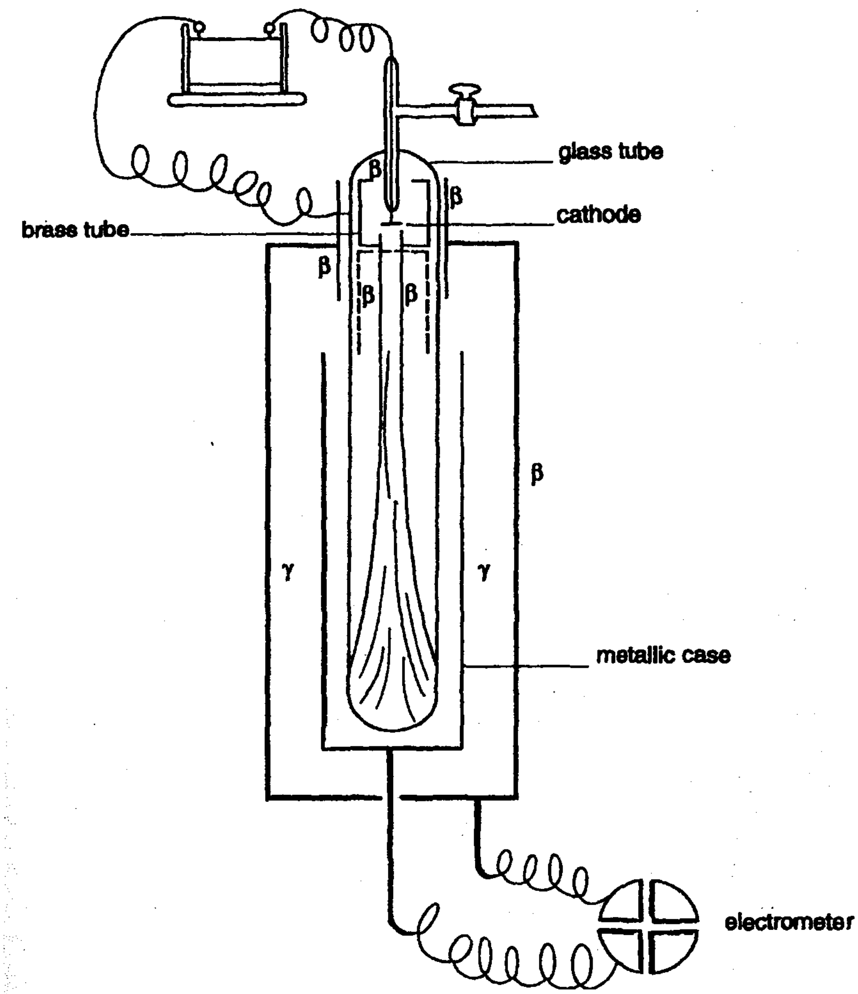
In 1895, a year after Hertz's death, the French physical chemist, Jean Perrin, carried out new experiments on cathode rays, from which he concluded that they are indeed electrically charged. In Perrin's experimental set-up, there was no device, such as the wire gauze, to separate the cathode rays from the ordinary electric current. If Hertz had been alive, he would have claimed that the charge that Perrin detected in the electrometer was not a charge on the cathode rays but one produced by the regular electric current in the tube.

Subsequently, two years later, in 1897, the British physicist, J. J. Thomson, conducted new cathode ray experiments that settled the issue definitively. Thomson repeated Perrin's experiment in a fashion that would circumvent the objection that the cathode rays and the ordinary electric current were being mixed together. Using a magnet, he deflected the cathode rays into a set of cylinders connected to an electrometer. He then showed that the electrometer registered a charge when, and only when the pure cathode rays were deflected toward it. In his discussion of this experiment Thomson (1897) wrote:

> An objection very generally urged against the view that the cathode rays are negatively electrified particles, is that hitherto no deflexion by the rays has been observed under a small electrostatic force ... Hertz made the rays travel between two parallel plates of metal placed inside the discharge tube, but found that they were not deflected when the plates were connected with a battery of storage-cells; on repeating the experiment I at first got the same result, but subsequently experiments showed that the absence of deflexion is due to the conductivity conferred on the rarefied gas by the cathode rays. On measuring this conductivity it was found that it diminished very rapidly as the exhaustion increased; it seemed then that on trying Hertz's experiment at very high exhaustion there might be a chance of detecting the deflexion of the cathode rays by a electrostatic force (p. 300).

In 1897, Thomson was technically able to produce a much higher vacuum in the tube than was Hertz in 1883. Thomson reasoned that if cathode rays are charged particles, when they pass through the gas in the gas tube, they ionize the gas molecules, thus producing both positive and negative charges that will neutralize the charge on the metal plates between which the cathode rays travel. Therefore, it stands to reason, that if the tube is not sufficiently evacuated, there will be no deflection of the cathode rays. Thomson's experiment clearly demonstrated that when the gas in the tube was much more completely evacuated than was done in Hertz's experiment, the cathode rays were

Figure 2-4. **Hertz's apparatus – the "Cathode Ray Purifier"**



deflected. From these two experiments Thomson concluded:

As the cathode rays carry a charge of negative electricity, are deflected by an electrostatic force as if they were negatively electrified, and are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body moving along the path of these rays, I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter. (p. 302)

The question to be considered was whether Hertz's experiments constitute evidence that cathode rays are not charged. The answer to that was that from 1883 to 1897 they were indeed evidence, but after 1897 they were no longer evidence. From 1883 to 1897 Hertz's experiments were considered evidence for the electric neutrality of cathode rays; after Thomson's 1897 experiments they were not. This answer suggests a concept of evidence that is relativized to a specific person or group: *e* is evidence that *h* for *such and such a person or group, or with respect to a certain set of background assumptions.* Besides, it may be evidence for that person or group *at one time but not another.* There are three conditions that must be satisfied if a concept of evidence is to be classified as *subjective evidence.*

First, this implies that the person or group at the time in question believes that *e* is evidence that *h*. Hertz fully believed this about his experimental results in 1883. Thomson, however, following his own more conclusive experiments in 1897, did not believe that Hertz's results were evidence that cathode rays are electrically neutral.

Second, the person or group believes that the hypothesis *h* is true or at least probable. After his 1883 experiments, Hertz believed that cathode rays are electrically neutral. Following his own experiments in 1897, Thomson believed that cathode rays are negatively charged. Thomson, in 1897, knew about Hertz's results and he was aware of Hertz's hypothesis that cathode rays are neutral. However, the results of Hertz's experiments were not *Thomson's* evidence that cathode rays are neutral, since Thomson did not believe that Hertz's hypothesis was true or probable.

Third, the person or group has as a reason for believing that *h* is true or probable, that *e* is true. Hertz's reason for believing that cathode rays are electrically neutral is that no electrical effects were produced in his experiments. If this had not been Hertz's reason for believing this hypothesis, then the experimental results in question, would not have been *his* evidence for that hypothesis. Analogous claims can be made for Thomson's evidence for the hypothesis that cathode rays are negatively charged. In subjective evidence, the evidence is seen as *X's evidence at time t,* where X is a person or group. We can say that *e* is X's evidence that *h* at time *t* if, and only if, the evidence meets the three conditions laid down earlier under *subjective evidence.*

Hertz's experiments were based on the mistaken assumption that the cathode tubes were sufficiently evacuated to permit the rays to be electrically deflected should they be charged. As Thomson's later experiments showed, Hertz's conclusion was false. Even though Hertz was justified in believing what he did, his experimental results do not in fact provide a good reason to believe his conclusion. Consequently, Hertz's results never were true, or genuine, or veridical evidence for his conclusion. The expression used here, "providing a good reason to believe," is vital in defining veridical evidence. In *veridical evidence*, if *e* is evidence that *h*, then *e* provides a good reason to believe *h*. In 1883 Hertz was justified in believing that cathode rays are neutral on the basis of his experimental results. However as Thomson demonstrated, these results do not "provide good reasons to believe" this hypothesis. Experiments and tests sometimes have flaws that call into question the conclusions that are drawn. As Thomson showed in 1897, there was a flaw in Hertz's 1883 experiments: Hertz's cathode tubes were not sufficiently evacuated to produce electrical deflection. We are supposing that this flaw could not have been expected by Hertz in 1883. Nevertheless, since his results were based on this flaw, however unexpected, they do not "provide a good for believing" Hertz's conclusion. Whether *e* is a good reason to believe *h* does not depend on whether in fact anyone believes *e* or *h*. Someone who knows Hertz's results is justified in believing his hypothesis only if that person's epistemic situation meets certain conditions (e.g. that person does not know, and is not in a position to know, Thomson's later results). Whether *e* is good reason to believe *h* can depend on empirical facts in addition to those reported in *e*. The empirical fact that Hertz's cathode tubes were not sufficiently evacuated to produce electrical deflection falsifies the claim that the absence of deflection in Hertz's experiments is a good reason to believe that cathode rays are electrically neutral.

The experiments that Thomson conducted in 1897, showed that Hertz's 1883 experiments contained a flaw, however unexpected. Thomson's experiments, however, showed something even more important, i.e. that Hertz's hypothesis that cathode rays are electrically neutral is false. The fact that it is false shows, that although Hertz's 1883 experimental results seemed to provide good reason to believe the hypothesis is true, they do not really do so. Because this CoE requires the truth of the hypothesis, *h,* Achinstein called this CoE that is based on the idea of "a good reason to believe," *veridical evidence.*

Potential evidence is related to a notion of "good reason to believe," that is weaker than that required for veridical evidence. Here there is no presupposition of the truth of the hypothesis. Some fact *e* can be evidence that *h*, and hence a good reason to believe *h*, even if *h* is false. This CoE is a fallibilist one. Whether some experimental, observational, or test results *e* constitute a good reason to believe *h* depends on whether there is some flaw in the design or execution of the experiment or test. Certain results may *appear* to be a good reason to believe a hypothesis, and someone in a given epistemic situation may be justified in believing they are a good reason, but in fact they are not. This was the case with Hertz's experimental results in 1883. Flawed experimental, observational, or test results do not constitute a good reason to believe a hypothesis, even in a sense of "good reason" that does not require the hypothesis to be true. Potential evidence is, therefore, very akin to veridical evidence.

As described in the initial classification, as well as elaborated on using the scientific example of Hertz's and Thomson's experiments, there are at least three different types of evidence. As seen from above, this classification system is very much dependent upon the context in which evidence occurs. This context could vary a lot depending on whether the students are carrying out an open-ended investigation, or a well-controlled lab or a science demonstration. Let me go on to examine how practical work can promote the understanding of evidence in science.

The Role of Practical Work in the Understanding of Evidence in Science

Practical work first became a part of science education in the mid-1850s (Gee & Clackson, 1992), when it was primarily carried out by the teacher as a means of illustrating concepts. By the turn of the 20th century, however, it came to be seen as a means of allowing students "to find out things for themselves." This led to the incorporation of individual practical work into the science curriculum. From then on, there has been an assumption that practical work carried out by the students themselves is a "good thing," and its inclusion in the science education curriculum has been accepted without serious question. Gott and Duggan (1996) chose an appropriate juncture to question whether practical work serves any useful purpose in today's science teaching. They found that the aims of experimental skills and techniques of doing science can be summarized as:

- Motivational aspects – linked to the promotion of interest and social skill;
- The application of substantive knowledge that have relevance in daily life;
- The development of experimental skills.

More recently, Millar (2004) emphasized how practical work is an essential component of science teaching and learning, both for the aim of developing students' scientific knowledge and that of developing students' knowledge *about* science. Practical work "seems to offer a way of holding up *evidence,* rather than authority, as the grounds for accepting knowledge" (p. 3). If practical work was to develop students' scientific knowledge, it would require students to make links between two domains of knowledge: the domain of objects and observables and the domain of ideas. Practical work which is more investigative and open-ended can enhance students' knowledge of scientific enquiry. It is difficult to argue that this kind of knowledge is necessary for scientific literacy, although it is of distinct value to students who wish to follow more advanced courses in science.

Substantive Knowledge and Procedural Knowledge

Experimental skills have a distinct knowledge base which is connected directly with the understanding of scientific evidence. According to Gott and Duggan:

> Being scientifically literate. . . . requires. . . that pupils have a sound knowledge base in the major substantive ideas of science and of ideas related to the collection, validation, representation and interpretation of evidence. Without such a repertoire, their ability to think scientifically about a whole range of issues will be impaired. (1996, p. 793)

In a later work, Roberts and Gott (2000) define "substantive ideas" as "the concepts which form the underpinning structure of the subjects": for example, ideas such as energy, force, photosynthesis, or solubility. Substantive knowledge refers to the understanding of the ideas in science which are based on facts, laws and principles: these are referred to as "substantive" or "declarative" knowledge.

Procedural knowledge, on the other hand, are the ideas "that are essential in the collection, understanding and evaluation of scientific evidence" (Roberts & Gott, 2000). Procedural understanding is the understanding of a set of ideas which is complementary to substantive understanding, but related to the "knowing how" of science. It is concerned with the understanding needed to put science into practice. "It is *the thinking behind the doing*" (Gott & Duggan, 1995). Lubben and Millar (1996) take the stand that procedural knowledge in science is"knowing how to carry out practical tasks." To illustrate this with an example: in a plant growth study, procedural understanding does not refer to the measuring itself, but to the decisions that have to be made about what to measure, how often, and over what period of time. It also includes an understanding of the notion of "fair test," as well as the nature of a line graph, and how it differs from a bar chart or how it illustrates patterns between variables.

The content of procedural understanding is not documented properly. Although procedural understanding *can* be a means of learning or learning about a concept, it is also a kind of understanding *in its own right* (Gott & Duggan, 1995), equivalent to substantive understanding. It has a knowledge base which is either not taught, or is taught with little in the way of system or rigour. The basic experimental *skills* used in empirical work are essential to the success of that work: but, on their own they are not enough. "There is a higher cognitive dimension, which has been called procedural understanding, necessary for the correct application of those skills" (Roberts & Gott,1999, p. 22). Just like the substantive ideas, procedural understanding has a detailed body of knowledge attached to it. The substantive body of knowledge is well known: the procedural one, however, is less familiar and often appears as fragmented elements in the science curriculum. It is important, therefore, to have a clearly laid out knowledge base for procedural knowledge, and to incorporate this into the science curriculum, so that it is addressed systematically and progressively during the teaching of school science.

Concepts of Evidence

Learning science requires the acquisition of both substantive knowledge and of skills. The substantive knowledge, as mentioned above, includes understanding the facts, concepts, and theories of science. Skills include knowing how to use a thermometer, how to draw a graph, how to set up distillation apparatus, and how to focus a microscope. It is not always recognized that these skills have a distinct knowledge base that is connected with the understanding of scientific evidence. The skills must be exercised within an understanding of ideas such as variables and their manipulation, accuracy, fair testing, and the validity and reliability of data. It is these ideas that Gott has termed "concepts of evidence." Gott et al. coined the phrase "concepts of evidence" (Duggan & Gott, 1994) to refer to the concepts which are associated with procedural understanding. According to him, concepts of evidence supply "the underpinning ideas about how evidence can be *collected*, *verified*, *analyzed* and *interpreted*" (Gott, 2004, p.11). Concepts of evidence are concepts underlying the doing of science, in relation to the evidence as a whole. They form a knowledge base akin to the substantive knowledge that traditionally, has been seen as the heart of science. These concepts of evidence were also termed "conceptions of scientific evidence" (Taylor & Dana, 2003).

Procedural ideas can be taught in practical as well as non-practical contexts (Roberts & Gott, 2000). The research at Durham University carried out by Gott et al. has shown that "certain ideas about the collection, analysis, and interpretation of data have to be well understood, before we can handle scientific evidence effectively" (Roberts, 2001, p. 114). It is the application and the synthesis of these CoEs that constitute the "thinking behind the doing" of science. Many students who take science courses, however, will not understand how to evaluate scientific evidence unless the underlying concepts of evidence are specifically taught (Roberts & Gott, 2000). If these concepts of evidence are to be taught, then they need to be carefully defined. CoEs are normally treated as *skills*, not conceptual knowledge. Since they are treated as skills, they do not often appear on typical achievement tests, and this therefore reduces the value placed on them in the curriculum.

The CoEs were developed by Gott and Duggan (1995) to describe the procedural understanding necessary for working in all science disciplines. At that time, in their first draft, the descriptors could be interpreted as being restrictive in that they were more closely allied to lab-based investigations, rather than being applicable to the many other types of science-based work, especially where relationships between naturally changing variables are studied, such as in the many types of biological fieldwork. More recently, Gott et al. (2003) have defined the CoEs in such a way that they could be much more readily applied to the range of contexts investigated by biologists, as well as scientists working in all branches of science. This compendium of CoEs developed is a comprehensive, but as yet tentative, definition of concepts of evidence ranging from the ideas associated with a single measurement to those which are associated with evaluating evidence as a whole.

The latest version of this compendium (Gott et al., 2003), obtained from their website, includes 21 categories of CoEs as shown in Figure 2-5. The 2003 version contains additional categories to those found in his 1995 version. However, in 1995 he provided a concise logic of his conceptualization, which I present in Figure 2-8. According to Gott's classification, the 21 categories of CoEs range in complexity from simple CoEs such as "Observation" and "Measurement" to more complex CoEs such as "Instruments: Calibration and Error" and "Reliability and Validity of a Single Measurement." Each of these 21 categories is sub-divided as well, although the sub-categories are not shown in Figure 2-5. This list of CoEs was informed by research and writing in primary and secondary science education, in science-based industry, and in the public understanding of science. Gott et al. point out (2003) that although some of these CoEs are fundamental and appropriate at any age, others may be necessary only for a student engaged in a particular branch of science.

If we accept that CoEs form an important part of scientific knowledge, and a part that is crucial to the citizen-scientist, then decisions must be made about the best way of teaching them. There is growing suspicion that CoEs are neither systematically nor comprehensively presented in the secondary school, or college and university science curriculum. There is also a widespread concern that the traditional substantive knowledge is so vast that there is insufficient time to cover it as well as more general issues such as understandings of evidence. As a consequence, suggestions have been proffered that the substantive content of science courses be reduced to make room for such knowledge as understanding of evidence, because it is this knowledge that students will use in coping with controversial socio-scientific issues that arise in their lives.

With this view on the understanding of evidence, Gott and Duggan have developed a simple model to describe problem-solving. If one is to have the ability to solve problems and judge evidence in science, then one must have an understanding of both the substantive ideas of science as well as a procedural understanding (Figure 2-6). Problem-solving in science, as shown in Figure 2-6, is represented as involving two kinds of understanding: substantive and procedural. The left strand in Figure 2-6, represents substantive understanding, while the right side stands for procedural understanding. When scientists solve problems, both substantive and procedural ideas are used (Roberts & Gott, 1999; Gott et al., 1999). The procedural knowledge is based on skills and concepts of evidence, while the substantive knowledge is founded on facts and substantive concepts.

Procedural knowledge was further investigated by Gott and Duggan (1995), and their building blocks - the CoEs of which there are 21 categories – were structured around the four main stages of investigative work: that is, concepts associated with the design of the task, measurement, data handling, and finally the evaluation of the complete task in terms of the reliability and validity of the resulting evidence. The stages in Figure 2-7 are not stages in time, because these stages may need to be revisited. At the data handling stage, for example, a decision may be made to increase the number of measurements. The final evaluation of the task necessitates an understanding of all three stages – design, measurement, and data handling – and this understanding of evaluation is needed as much at the beginning as at the end of the task. The kind of understanding that is required at each of these four stages is laid down in detail in Figure 2-8.

Included within the initial stage of investigative work were an understanding of: (a) the idea of variable and identifying the independent variable and measuring or assessing the dependent variable; (b) the fair test in terms of controlling the necessary variables and its importance in relation to the validity of any resulting evidence; (c) the significance of an appropriate sample size; (d) the distinction between categoric, discrete, continuous and derived variables. The second stage was comprised of CoEs associated with measurement and required an understanding of: (a) the need to choose sensible values for quantities so that the resulting measurements will be meaningful; (b) range and interval; (c) the relationship between the choice of instrument and the required scale, range of readings required, and their spread and accuracy; (d) the need for repeats to give reliable data; (e) the appropriate degree of accuracy that is required to provide reliable data. The third stage of investigative work was associated with data handling and enveloped and understanding of: (a) tables; (b) graph types; (c) patterns that represent the behaviour of variables and how they are represented in tables and graphs; (d) the nature of multivariate data. The final stage of investigative work was associated with the evaluation of the complete task and involved and understanding of: (a) reliability and (b) validity. These and the understandings introduced above are described in more detail in Figure 2-8.

Students' Perceptions about Evidence and Their Facility in Using it

In their paper, Lubben and Millar (1996), discussed students' responses to probes exploring the understanding of reliability of measured data. Using paper-and-pencil tests on a large mixed-ability group of students in the 11-16 age range, they suggested a model shown as Figure 2-9, depicting the progression in student ideas about the reliability of experimental data.

The procedural progression model shown in Figure 2-9, was tested to determine its applicability on older and educationally more advanced students by Allie et al. (1998). The sample on whom they did the test was first year undergraduate physics students on the Science Foundation Programme at the University of Cape Town. The demographics of the sample tested showed that a large proportion of the sample was second-language English speakers, and the majority have had little or no experience with practical work before entering university. They all have, however, completed a course in "Tools and Procedures for Physics 1" prior to their involvement in this study. The focus of this study was on three probes dealing with the reasons for repeating measurements and three probes concerned with how to deal with sets of experimental data, especially their ideas about the spread of a set of measurements. Most of the answers of the respondents were classified as falling into levels F, G or H described in Figure 2-9. There were only a few instances where A, C and D level reasoning could be identified. However, it was found that the scheme required an additional level to allow for the most sophisticated reasoners to be accommodated. Lubben-Millar model (Figure 2-9) should include a new category

Figure 2-5. **Categories of the Concepts of Evidence from Gott**

1. Fundamental ideas
2. Observation
3. Measurement
4. Instruments: underlying relationships
5. Instruments: calibration and error
6. Reliability and validity of a single measurement
7. The choice of an instrument for measuring a datum
8. Sampling a datum
9. Statistical treatment of measurements of a single datum
10. Reliability and validity of a datum
11. Design of investigations: Variable structure
12. Design: Validity, "fair tests" and controls
13. Design: Choosing values
14. Design: Accuracy and precision
15. Design: Tables
16. Reliability and validity of the design
17. Data presentation
18. Statistics for analysis of data
19. Patterns and relationships in data
20. Reliability and validity of the data in the whole investigation
21. Relevant societal aspects

called I-reasoners who "show an understanding that consistency of data sets can be judged by comparing the relative positions of their means in conjunction with their spreads."

Figure 2-6. **A Model for Problem Solving from Gott and Duggan**

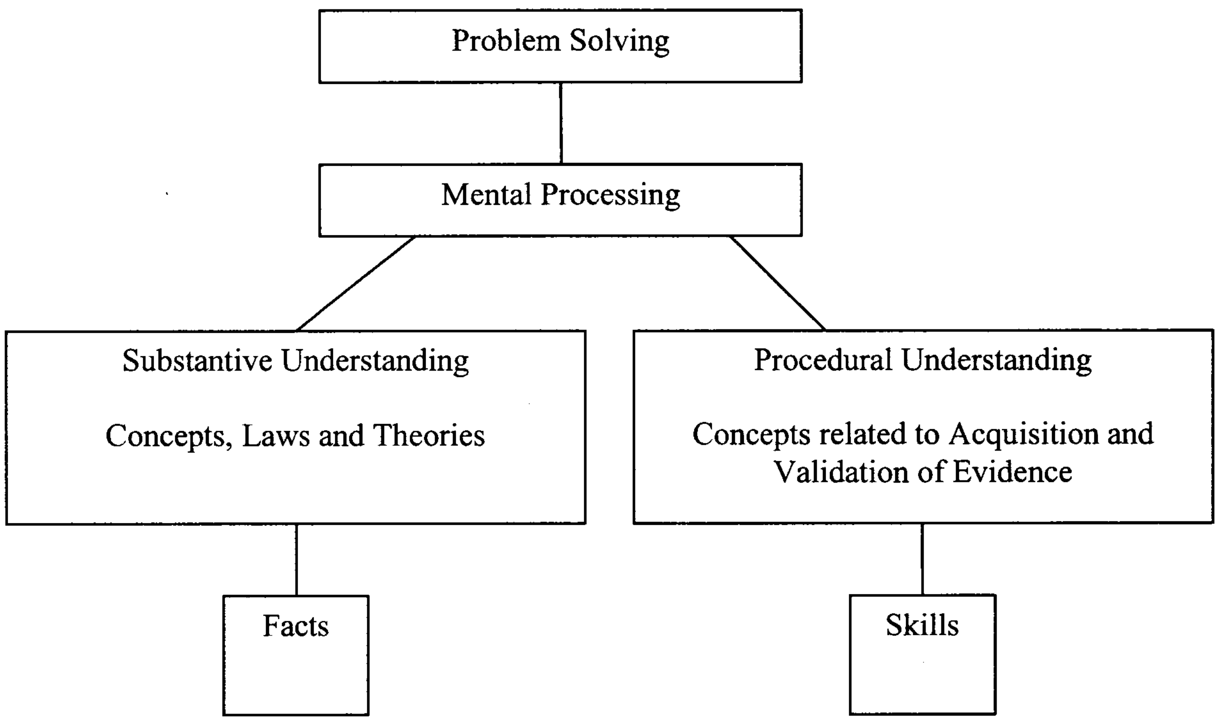


Figure 2-7. **Procedural Understanding and Concepts of Evidence**

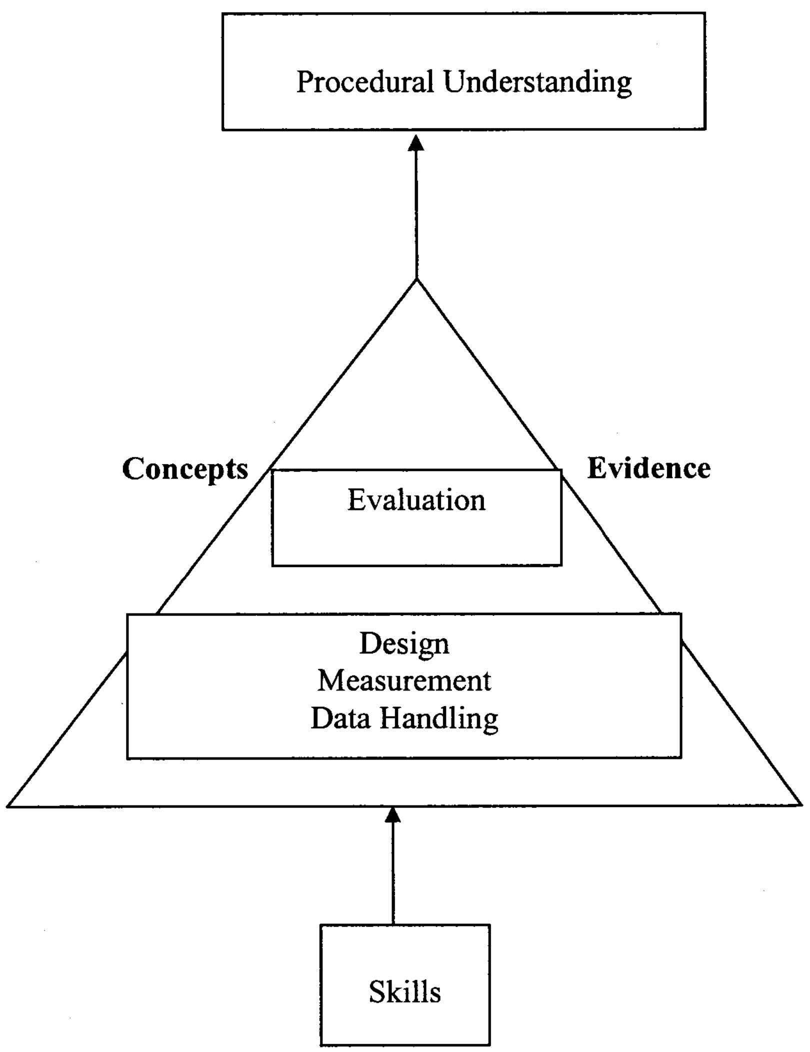


Figure 2-8. **Concepts of Evidence and Their Definition**

Concepts of Evidence		*Definition*
Associated with design	Variable	Understanding the idea of a variable and identifying the relevant variable to change (the independent variable) and to measure, or assess if qualitative (the (the dependent variable)
	Fair test	Understanding the structure of the fair test in terms of controlling the necessary variables and its importance in relation to the validity of any resulting evidence.
	Sample size	Understanding the significance of an appropriate sample size to allow, for instance, for probability or biological variation
	Variable types	Understanding the distinction between categoric, discrete, continuous and derived variables and how they link to different graph types
Associated with measurement	Relative scale	Understanding the need to choose sensible values for quantities so that resulting measurements will be meaningful. For instance, a large quantity of chemical in a small quantity of water causing saturation, will lead to difficulty in differentiating the dissolving times of different chemicals
	Range and interval	Understanding the need to select a sensible range of values of the variables within the task so that the resulting line graph consists of values which are spread sufficiently widely and reasonably spaced out so that the 'whole' pattern can be seen. A suitable number of readings is therefore also subsumed in this concept.
	Choice of instrument	Understanding the relationship between the choice of instrument and the required scale, range of readings required, and their interval (spread) and accuracy.
	Repeatability	Understanding that the inherent variability in any physical measurement requires a consideration of the need for repeats, if necessary, to give reliable data
	Accuracy	Understanding the appropriate degree of accuracy that is required to provide reliable data which will allow a meaningful interpretation
Associated with data handling	Tables	Understanding that tables are more than ways of presenting data after they have been collected. They can be used as ways of organising the design and subsequent data collection and analysis in advance

Figure 2-8 (Cont)

		of the whole experiment
	Graph type	Understanding that there is a close link between graphical representations and the type of variable they are to represent. For example, a categoric independent variable such as type of surface, cannot be displayed sensibly in a line graph. The behaviour of a continuous variable, on the other hand, is best shown in a line graph
	Patterns	Understanding that patterns represent the behaviour of variables and that they can be seen in tables and graphs
	Multivariate	Understanding the nature of multivariate data and how particular variables within those data can be held constant to discover the effect of one variable on another
Associated with the complete task	Reliability	Understanding the implications of the measurement strategy for the reliability of the resulting data; can the data be believed?
	Validity	Understanding the implications of the design for the validity of the resulting data; an overall view of the task to check that it can answer the question

From the responses obtained for the DMSS (i.e. different mean but a similar spread) probe which concentrates on the spread around a mean to compare whether the two sets of measurements are consistent with each other, 30% of the total sample may be regarded as "spread reasoners" (i.e. those who mention the calculation of a spread or uncertainty as reason for repeating measurements). These students "recognize that there are variations between the data points but do not synthesize this into a global measure which may be used together with the mean to characterize the data set."

Figure 2-9. **Model of Progression of Ideas Concerning Experimemtal Data**

Level	*Student's View of the Process of Measuring*
A	Measure once and this is the right value
B	Unless you get a value different from what you expect, a measurement is correct
C	Make a few trial measurements for practice, then take the measurements you want
D	Repeat measurements till you get a recurring value. This is the correct measurement
E	You need to take a mean of different measurements. Slightly vary the conditions to avoid getting the same results.
F	Take a mean of several measurements to take care of variation due to imprecise measuring. Quality of the result can be judged only by authority source.
G	Take a mean of several measurements. The spread of all the measurements indicates the quality of the result.
H	The consistency of the set of measurements can be judged and anomalous measurements need to be rejected before taking a mean.

In the study by Allie et al., it was found that significantly fewer students perceived the need for determining a mean for measurements of distance than for measurements of time. This meant that the use of procedural understanding depends on the measure being made. This reiterates the observation that the application of conceptual understanding is influenced by context. Consequently, it is important to document students' perception about procedural decisions for experimental work with a variety of measuring instruments (e.g. ammeters, thermometer, and pH meters, etc.) each with different resolution. It may also be added here that research is needed into the similarities and differences in the use of procedural understanding in various sciences, (e.g. chemistry, biology, physics, and astronomy) in order to determine more clearly the parameters for the transferability of procedural understanding.

Despite the fact that the students in this sample may be characterized as "advanced reasoners," their language use was haphazard. Terms concerning data such as *measurement, calculation, result,* and *value* are used interchangeably. There is also some confusion about terminology such as: *spread, error, range, uncertainty, precision* and *accuracy.* This may be attributed to linguistic problems of second-language English speakers. However, work done by Sere et al. (1993) show a similar loose usage of terminology for (first-language) French physics undergraduates. It appears, therefore, that the haphazard use of terminology is equally related to the lack of differentiation between systematic and random errors in the minds of the students. The vast majority of students in this study argued justifiably "that repeating is needed to limit the random error, and therefore to improve precision" (p. 427). Very few students referred (incorrectly) to repeating for the purpose of excluding systematic errors such as parallax. Out of the total responses, 51% indicated that students repeat in order to get closer to the 'real' or 'correct' value for the time/distance measurement. This implied that they intended to improve the accuracy, which depends on eliminating systematic errors and cannot be achieved by repeating measurements. This misconception may be remedied if teaching included short practical tasks challenging students to develop and execute techniques for reducing random errors and systematic errors, respectively. In addition, having a specific requirement for including separate notes on potential sources of random and systematic error in standard practical reports may remind students of appropriate terminology and relevant strategies.

The ability to handle variables is a significant component of children's logical development. In Gott's classification of CoEs, he placed the chief variables (Dependent and Independent) under the main category of "Design of Investigation: Variable Structure." Under this main category there are sub-categories such as: sub-category 1 – The Independent Variable; and sub-category 2 – The Dependent Variable. The paper by Donnelly (1987) reported findings on the variable-handling of pupils fifteen years old, in the context of scientific investigations. This paper used the survey data of the science programme of the UK Assessment of Performance Unit (APU). The study referred to a national pupil population of England, Wales and Northern Ireland. The data used refer to six questions used in the APU science programme for 1984. The questions are based on two physical situations called "Catalase" and "Springload." Both situations are investigated in terms of one dependent and two independent variables, in addition to an indeterminate number of variables requiring control. "Catalase" examined the effect of river sludge on hydrogen peroxide, while "Springload" studied the rate of bobbing of springs of various dimensions. Responses to each situation were investigated in practical, oral and pencil-and-paper modes. In the practical mode, pupils were allowed to investigate the situation using a range of apparatus. On completing the practical task pupils were asked how they had attempted to determine the effect of each variable, by using a standardized question.

The "fair test" is a more advanced CoE sub-category in which one independent variable is varied at a time, and the dependent variable is monitored. It would be interesting, therefore, to gauge students' perceptions about this CoE and their facility in using it. Almost all pupils in the Donnelly study answered in terms of the amounts of one or both of the independent variables used. For the "Catalase" situation, the data indicate that 47% of the pupil population could generate an effective design. It is noteworthy that only 58% of respondents identified the volume of hydrogen peroxide as a parameter to be varied. The oral performance was poorer and this may be related to linguistic problems, especially the inability to articulate an experimental design which had in fact been carried through. The written response allowed some pupils to offer a more effective design, while it inhibited others from expressing any idea of variable manipulation. In "Catalase", a more complex process of variable operationalisation is required, because pupils frequently appeared to manipulate *both independent variables at once* in a very deliberate way. These parallel some of the behaviours described by Piaget (Inhelder and Piaget, 1958, p.74) in the pendulum problem where students were varying both length of string and weight of the bob deliberately but simultaneously in the same experiment.

The pupils operationalised the independent variables with much greater ease in the "Springload" situation. 85% of pupils carried out a non-confounded test of *both* variables. The performance in the oral response again was substantially lower than that in the practical task, with only 33% suggesting non-confounded combinations of springs for both variables. It appears, therefore, that "fulfilling the behavioural criteria alone may not indicate adequate conceptualization of experimental design." Pupils could perform well on the practical task and still be deficient in understanding the principles underlying the manipulation of variables in a non-confounding manner. Approximately 23% of pupils who carried out a full, non-confounded test of both variables on the behavioural criteria, did not indicate in response to questioning a suitable combination of springs to test *either* length *or* width. One "logical" combination of springs was found to be heavily represented among pupils who had failed to give expected combinations. These pupils suggested the use of two or more of the longest springs to test for length, and the equivalent situation for width. When the abstract written environment was use, the performance on the "Springload" type question dropped below that in the equivalent "Catalase" type question. In both situations (i.e "Catalase" and "Springload"), however, the most notable point is the proportion of pupils not identifying an independent variable as requiring alteration during an investigation. Donnelly stated in conclusion:

> There is apparently an intermediate step between identifying characteristics of situations and understanding these as a set of independently manipulative "variables." In addition, there is an evident logical connection between appreciating manipulation of such a characteristic as a method of establishing knowledge of relationships and the need to hold constant characteristics other than that being investigated. (p. 145)

This work implied that to understand more fully 15-year-old pupils' approaches to the variable-based logic of investigation, we need to first investigate their underlying conception of a "variable," and of "instrumentality", as central criteria in the validation of scientific knowledge.

The study by Kanari and Millar (2004) was described earlier in this chapter. It explored the perceptions of data and measurement that 10 to 14-year-old students held and drew upon as they carried out a practical science inquiry task. It also delved into how students reason from data they have collected, or are presented with. This work shed light on the perceptions of students on approaching data collection and data interpretation, especially when investigating the relationships between variables. Most students controlled the second independent variable while collecting data of the effect of the independent variable they were investigating (i.e. the first independent variable). Therefore, the idea of the "Fair Test" was well perceived and understood, perhaps because of the fact that this concept had been strongly emphasized in school science teaching from primary school onward in England, in the past decade.

The most common data collection strategy used by the students in Kanari and Millar's study was trend-focused, i.e. increasing/decreasing the value of the independent variable in steps, and looking for a corresponding steady increase/decrease in the dependent variable. Only a small minority used a difference-based strategy, i.e. looking first for a difference in the dependent variable when small and large values of the independent variable were used, in spite of the fact that this strategy is often a more efficient way to proceed. The preference of the trend-focused strategy may reflect the usage of a "routine" for school science investigations of this sort. This may be because the English National Curriculum emphasized the relationships between variables in practical investigations, so that students from an early age, are familiar with such tasks. Most students in Kanari and Millar's study did not repeat measurements, and did not perceive the need to do so. When repeat measurements were taken, it was either because the students had been taught a "routine" that they followed rather automatically; or an unexpected result such as a value that does not obey an emerging trend; or a response to something unexpected such as a problem while making the measurements. This study, however, found that many children have some awareness of measurement error, because most do not show any surprise that repeat measurements of the same quantity are different. Only a minority, they felt, however, had developed this awareness of the variability of measurements into a notion of "measurement uncertainty" or "error." The students in this study had much greater problems in interpreting data; for example, deciding whether a variable has increased, decreased or stayed the same. The authors felt that non-covariation situations should definitely be included because here the inherent uncertainty of measurements (or error), and therefore the need to repeat them, can be clearly demonstrated and taught.

Ryder et al. (1999) described views about the nature of science held by a small sample of science students in their final year at the university. Students in their sample tended to view knowledge claims in science as provable beyond doubt using empirical data alone. In their professional lives (e.g., as research scientists, teachers, or journalists) many of these students would have to communicate new scientific ideas to those outside of science. In many instances these knowledge claims may be resting on tentative evidence and may be contentious within the scientific community itself. This makes it imperative that science teaching address and develop students' views about the relationship between knowledge claims and data. It was observed that even in the final-year project work, where science is being made in many cases, the emphasis is still on getting reliable data or evidence.

Students who decided to study further would have opportunities to develop their images of science during their doctoral studies through attendance at conferences, discussion with other scientists, and their own attempts to get their ideas accepted. It is, however, important to communicate a realistic image of science to those science undergraduates who will not become professional scientists. This study by Ryder et al. suggested that undergraduate students' images of science can be broadened and developed when they are placed in appropriate teaching contexts. In addition, the nature of the students' discipline and the type of project they were working on, were important influences on the development of the students' images of science. For example, it was found that students whose project had an epistemological focus (i.e. relating data to knowledge claims) tended to show developments in their epistemological reasoning. Conversely, students whose projects involved making experimental techniques work with novel materials, tended to show limited development in their reasoning about data and knowledge claims.

From the interviews in Ryder's study it was apparent that those students whose projects had involved working closely with professional scientists and research students had learned a great deal about the world of science, even though such issues were rarely discussed explicitly. For many students involved in this study, project work in their final year was a socialization into the practice of scientific research in universities. Students got to know professional scientists and research students as colleagues. They were able to express their own ideas and discuss them with experts in the field. Students whose project work involved working within an active research laboratory, learned about science through a form of apprenticeship. In addition, laboratory work can provide opportunities in which students can discuss with their peers or with more experienced science students the uncertainties and ambiguities present in measurements, the difficulties of absolute proof, and the conjectural nature of many knowledge claims.

Summary

This chapter, being a literature review, traced the history of science education from its genesis in the 19th century to its present stature as an intellectually demanding field with a great impact on people's lives especially in the contemporary, technology-driven society. It also conceptualized and characterized scientific literacy, following its growth and development through history. The four major themes of scientific literacy for analyzing science textbooks were listed and the one most relevant for describing evidence identified. The state of school science was discussed along with the role of textbooks in the teaching of science. Evidence was defined and the work done on understanding evidence tracked. The different types of evidence were then described with illustrative examples from the work of Hertz and Thomson. Substantive and procedural knowledge were traced as the two necessary arms of problem-solving in science. Concepts of evidence were then justified as the building-blocks of procedural knowledge. Finally, students' perceptions of evidence and their facility in dealing with it were discussed with suitable case studies as examples.

CHAPTER THREE

DESIGN AND METHODOLOGY

The study is primarily analytic and evaluative designed to see what CoEs are present, in what abundance, how, and with what curricular implications they are distributed. A detailed content analysis of the sampled portions within the selected textbooks was undertaken. Content analysis may be defined as "a summarizing, quantitative analysis of messages that relies on the scientific method .... and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented" (Neuendorf, 2002). The scientific method endorses paying attention to objectivity-intersubjectivity, a priori design, reliability, validity, generalizability, replicability and hypothesis testing. This chapter describes the sampling technique used; the methodology used in identifying situations in the body of the texts; spotting the CoEs contained in each situation and categorizing them according to Gott's classification; coding the situations; tabulating the data; and finally standardizing the data so that they are comparable across the different textbooks. Some simple, complex and interesting examples of situations were chosen to illustrate clearly the methodology ascribed in the data collection.

Sampling

The sampling technique employed resulted in three high school chemistry textbooks and one undergraduate chemistry textbook. Sampling is the process by which a subset of units are selected for study from the larger population (Neuendorf, 2002). For this study, sampling took place at three levels. At the first level of sampling, chemistry textbooks were selected using a purposive technique. In purposive sampling the researcher makes a decision as to what units he or she deems appropriate to include in the sample. In trying to decide upon which were the most commonly used chemistry textbooks, I checked with the Ministries of Education in various provinces (including Alberta, British Columbia and Ontario) as well as the publishers, Thomson Nelson and McGraw-Hill Ryerson, and the booksellers, Chapters Indigo, and based my selection on their experienced observation over a span of many years. Among all the high school chemistry textbooks currently authorized by provincial ministries of education, the three most commonly used across Canada were chosen as well as the undergraduate chemistry textbook *General Chemistry* (Petrucci et al., 1993), which is the prescribed textbook for the first year undergraduates in the Chemistry Department at the University of Alberta.

The high school textbook used in Alberta was *Nelson Chemistry* (Jenkins, van Kessel & Tompkins, 1993). Nelson Chemistry was found to be the most popular chemistry textbook used in Canada. Part of the reason for this popularity is that the Thomson Nelson publishers offered different editions of the textbook catering specifically for various regions of this country. For example, they have the Atlantic edition that caters only for the Atlantic provinces – New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island. There is also an Alberta edition, a British Columbia edition and a Quebec edition more suited to each of those provinces respectively. The chemistry content contained in each of these editions is the nearly the same, but they differ in the local examples selected, the scientists quoted and the career opportunities highlighted. The second most popular chemistry textbook in Canada was found to be *McGraw-Hill Ryerson Chemistry 11 and 12* (Mustoe et al., 2001) for grades 11 and 12 respectively. *Chemistry* by Chang (2005) was the third most commonly used textbook, catering specifically for students doing the International Baccalaureate (I.B.) program in high school. The undergraduate *General Chemistry* textbook was used by all students who took the general, fundamental first year chemistry courses, Chemistry 101 and Chemistry 102, as well as Chemistry 103 and 105 geared more for students in the Faculty of Engineering. All the four chosen chemistry textbooks were purchased.

At the second level of sampling, in order to facilitate selection of units within textbooks, the Alberta Chemistry-20 and Chemistry-30 curricula were used. Two prescribed units, one from each of the two grade levels were chosen purposively. The Chemistry-20 topic "Solutions" and the topic "Acids and Bases," covered in both Chemistry-20 and -30, were chosen from the chemistry curriculum because these were the topics most prone to having numerous CoEs. Generally, the topic "Solutions," as entailed in the curriculum, was covered in a single chapter in the four different textbooks. "Acids and Bases," however, (usually started in Chemistry-20 and completed in -30) was a topic that spanned about two-and-a-half chapters in most textbooks.

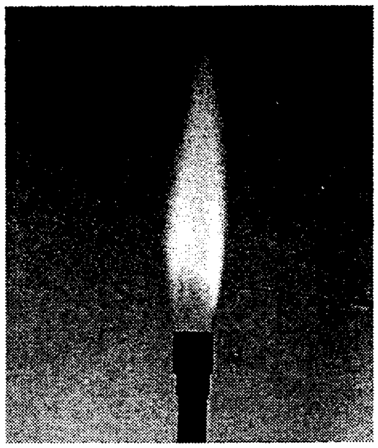
At the third level of sampling, all the pages under each topic were included, except for end of chapter summaries, exercises, pictures, tables, figures and their descriptive legends, samples of solved problems, molecular models, structural formulae showing mechanisms, biographies of scientists and accounts of their daily work, key words, and review questions. I focused exclusively on the prose aspects of the textbooks, recognizing that the books are multi-semiotic. The graphs, pictures, figures, tables, and the like are also in need of analysis, although I will not analyze them in this study.

Identifying Situations

The pages included in the sample were read and studied in order to identify "Situations." Situations were places in the text containing an observation or a diagram in which there was a change that could be observed. The observation may be one that is described in the body of the text, or a physical or chemical change that will be observed during the course of an experiment or investigation that students are asked to carry out, or that will be demonstrated to them by their teachers. A Situation could also be the description, or photographic representation, of an observable change in any form of matter within the body of the textbook. The excerpt, shown in Figure 3-1, is taken from the chapter on "Solutions." As I read through this page titled "Qualitative Analysis by Color," I came to the second sentence which reads: "For example, copper(II) ions produce a BLUE aqueous SOLUTION and usually a GREEN FLAME" (Jenkins et al., 1993, p. 123, emphasis added). The blue solution and the green flame are both observable and count as valid observations in this topic of qualitative analysis by color. That means this is a "Situation." I searched each Situation to see which of Gott's (2002) Concepts of Evidence were present. The Situations were consecutively numbered from one onwards. Each Situation number was placed conspicuously in a circle.

Figure 3-1. **An example of how a Situation in the text is spotted and the CoEs contained in it are categorized**

10



1 (3) **Figure 5.9**

Copper(II) ions usually impart a green colour to a flame.

Qualitative **Analysis** by Color

Some ions impart a specific color to a solution, a flame, or a gas discharge tube. For example, copper(II) ions produce a blue aqueous solution and usually a green flame. A *flame test,* a test for the presence of metal ions such as copper(II), is conducted by dipping a clean platinum or nichrome wire into a solution and then into a flame (Figure 5.9). The initial flame must be nearly colorless and the wire, when dipped in water, must not produce a color in the flame. Usually, the wire is cleaned by dipping it alternately into hydrochloric acid and then into the flame, until very little color is produced. The colors of some common ions in aqueous solutions and in flames are listed on the inside back cover of this book.

It is useful to see a case of what was ruled out as being a Situation. The non-example in Figure 3-2 is selected from *General Chemistry* by Petrucci et al. It shows a master brewer inspecting wort temperature and pH in the making of beer. Though it shows the master brewer in action, none of her observations are visible. Therefore, since there are *no observations*, this is *not* a Situation.

Figure 3-2. **A Master Brewer inspecting Wort Temperature and pH in the Making of Beer**

Coding Situations

When a major category of CoEs was identified, its corresponding number according to Gott's classification was placed in a square box in the empty space on the page where it was spotted. An example of this is illustrated, in detail, in Figure 3-1. Each major category in Gott's system was divided into sub-categories which were also depicted by numbers. These numbers were placed in round brackets next to the square, as clearly illustrated in the example shown as Figure 3-1. Comments, where necessary, about the various categories and sub-categories of CoEs in each Situation were carefully noted.

For example, for the Situation in Figure 3-1, the observation of blue solution and green flame is a "Fundamental Idea" and, therefore, falls under Category 1 according to Gott's classification. Since this is a major category, the number 1 is placed in a box as shown in Figure 3-1. Since the green flame color is caused by the presence of the copper(II) ions, the association here is causal, and therefore the sub-category here is 3 – "Association & Causation"- as per Gott's classification (Gott, 2002). The sub-category three is indicated by placing the three in round brackets as shown in Figure 3-1. Each CoE, therefore, is given a major category, placed in a square, and one or more sub-categories, placed in round brackets. If no suitable sub-category was found, then I suggested a new, additional sub-category that could be added to Gott's classification system. Appendix B illustrates the coding for a complex Situation (Situation 44) that included all 21 categories of CoEs.

Data Tabulation and Standardization

Once the Situations had been identified, and the CoEs in each had been categorized and sub-categorized, I collected all the resultant data. The frequencies of occurrence of each CoE by Situation were determined and compared by grade level and publisher. The total number of different concepts of evidence by grade and publisher were used to indicate the range of coverage of Gott's classification system. The occurrence of CoEs by grade was examined for development in sophistication. The total set of identified CoEs was examined in comparison to Gott's system to identify any gaps in the depth, accuracy, and appropriateness of the occurrences.

If any comparisons are to be made between textbooks, there has to be a way to standardize or normalize the number of CoEs observed, as number of CoEs/page, number of CoEs/set number of lines, or number of CoEs/set number of words. Although all three standardization methods mentioned above can be used, the third method of standardization, though tedious, is the most accurate. In this study, therefore, the third method was pursued for each of the five textbooks.

In order to implement the third method of standardization, the average number of words per line had to be determined for each textbook. This was done by counting the number of words in ten lines chosen from the various chapters examined in each textbook. This was repeated four times, and the average number of words per line determined.

The next step in standardization was to count the number of lines in every sampled page of each chapter. Although this was a tedious task, the accuracy achieved as the end-result made the undertaking worthwhile. Multiplying the number of lines obtained, by the number of words per line, gave the total number of words sampled in the chapter. The CoEs could then be expressed as a certain number per fixed number of words of the text. This effort made comparisons of the representation of CoEs between the texts comparable and meaningful.

CHAPTER FOUR

RESULTS

In this chapter, I present the data obtained from identifying CoEs within Situations. There are three basic tables for each publisher in which the data collected from the textbooks were entered, and each of these three tables will be discussed at length. There will be an analysis of how laden the individual situations are with concepts of evidence. The frequency of occurrence of situations, and of concepts of evidence within these situations, across grades, topics, and publisher will be mapped out as well. The data gathered from the textbooks will then be compared across publishers, across topics of study and across level of education.

Results by Publisher

Nelson "Chemistry"

Table 4-1 shows the "Frequency of CoE by Situation," in the Nelson Series under the topic "Solutions," at the Chemistry-20 level (i.e. Grade 11 level). Ignoring the details, let us consider what this table represents. The Situations in this chapter are listed in the first column under the rows 1 to 19. The Concepts of Evidence (CoEs) are listed in the columns 1 to 21. The last column lists the total number of CoEs in each situation. The general trend is that some situations have hardly any CoEs (e.g. Situation 10, with one CoE, or Situation 6 with two CoEs), while others have many CoEs (e.g. Situation 15 with 71 CoEs, or Situation 16 with 41 CoEs). Notice that there are many CoEs with zero representation. The mean number of CoEs/Situation in Table 4-1 is 22. The total occurrence of each CoE is in the last row of Table 4-1 under the heading "Total."

According to these values, the individual CoEs appear with frequencies ranging from 2 (for CoE #8, Sampling a Datum) to 51 (for CoE #5, Instruments: Calibration and Error). "Instruments: Calibration and Error" (i.e. CoE #5) is, by far, the CoE that occurred most predominantly in this chapter "Solutions."

Since the comparison between textbooks, as discussed in Chapter Three, can only be made if the CoEs are standardized to Number of CoEs/a set number of words, I proceeded to standardize the results in Table 4-1. Using the frequency of each of the CoEs from CoE #1 to CoE #21, and the total number of words in the text covering the topic "Solutions – Chemistry-20", the frequency of each CoE/500 words of text was calculated and listed. The last row of Table 4-1, therefore, gives the frequency of each individual CoE/500 words of text, in the topic "Solutions – Chemistry-20".

From the last row in Table 4-1, it is obvious that the frequencies of some individual CoEs fall at the high end of the spectrum while others fall at the low end of the spectrum. The CoE with the highest frequency is CoE #5 with a frequency of 3.79CoEs/500 words of text; the second highest frequency is for CoE #7 with a frequency of 3.57CoEs/500 words of text; and the third highest frequency if for CoE #6 with a frequency of 2.6CoEs/500 words of text. The CoE with the lowest frequency was found to be CoE #8 with a frequency of 0.15CoEs/500 words of text; the second lowest frequency of 0.3CoEs/500 words of text is for CoE #9; and the third lowest frequency of 0.52CoEs/500 words of text was for CoE #17. This is discussed further under Table 4-16, which looks at the frequency of CoEs/500 words of text across publishers, in the four textbooks studied.

Table 4-1

Frequency of CoEs by Situation in the Nelson Series (Solutions -- Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
2	1	1	1	2	3	3	3	0	0	2	2	2	3	1	1	2	1	2	2	0	32
3	1	1	1	2	3	3	3	0	0	2	2	2	3	1	1	2	1	2	2	0	32
4	2	1	1	2	4	3	3	0	0	2	2	2	3	1	1	2	1	2	2	2	36
5	1	3	2	1	4	3	3	0	1	1	2	2	3	2	1	2	1	1	2	2	37
6	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	1	1	0	0	0	0	0	0	0	0	2	2	3	1	1	2	1	2	2	2	20
10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	1	1	1	3	2	3	1	0	1	2	2	3	1	1	2	1	2	3	2	33
12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	2	2	2	2	4	3	5	0	0	4	0	0	0	0	0	4	0	0	0	0	28
14	1	2	2	2	2	3	3	1	1	2	2	2	3	1	1	2	1	2	2	0	35
15	5	6	7	5	14	5	12	0	0	10	0	0	0	0	0	6	0	1	0	0	71
16	2	4	6	2	7	3	5	0	0	4	2	1	0	1	1	2	0	1	0	0	41
17	2	1	1	1	2	3	4	0	1	0	3	2	0	0	0	0	0	1	0	0	21
18	0	2	1	1	4	3	3	0	1	0	0	0	0	0	0	0	0	0	0	0	15
19	1	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	9
Total	24	32	26	22	51	35	48	2	4	28	19	17	21	9	8	26	7	17	15	8	419
Frequency /500 Words of Text	1.78	2.38	1.93	1.63	3.79	2.6	3.57	0.15	0.3	2.08	1.41	1.26	1.56	0.67	0.59	1.93	0.52	1.26	1.11	0.59	31.11

Table 4-2

Frequency of CoEs by Situation in the Nelson Series (Acids and Bases -- Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
20	1	0	0	1	5	4	4	0	1	2	0	0	3	0	0	2	0	0	0	0	23
21	2	3	0	0	0	0	0	0	0	0	2	3	0	0	0	2	0	0	0	0	12
22	1	2	2	1	3	4	4	3	1	2	4	2	6	2	1	2	3	2	2	3	50
23	1	3	2	2	3	4	3	0	0	2	4	2	1	1	1	2	1	2	1	2	37
Total	5	8	4	4	11	12	11	3	2	6	10	7	10	3	2	8	4	4	3	5	122
Frequency /500 Words of Text	0.86	1.37	0.69	0.69	1.88	2.06	1.88	0.51	0.34	1.03	1.71	1.2	1.71	0.51	0.34	1.37	0.69	0.69	0.51	0.86	20.9

Table 4-2 depicts the "Frequency of CoEs by Situation" in the Nelson series, still at the Chemistry-20 level (i.e. Grade 11 level) but under a different topic, "Acids and Bases." The layout of Table 4-2 is identical to that of Table 4-1. It must be noted, however, that there are far fewer Situations in Table 4-2 (i.e. 4 Situations) than in Table 4-1 (i.e. 19 Situations), although the number of pages covered in the Nelson "Chemistry" textbook itself, range from 13 pages for Table 4-1, to 17 pages for Table 4-2. As in Table 4-1, the last column shows the total number of CoEs for each Situation. All Situations have a significant total number of CoEs, varying from 12 to 50 CoEs. The mean number of CoEs/Situation is 31 in Table 4-2. There is a small proportion of CoEs with zero representation in Table 4-2 (i.e. 23%), than in Table 4-1 (i.e. 45.6%). The total values for the individual CoEs have frequencies ranging from 2 for CoE # 9 (Statistical Treatment of Measurements of a Single Datum) to 12 for CoE #6 (Reliability and Validity of a Single Measurement). CoE #8 with a total occurrence of 3 is next in line to CoE #9 at the lowest end of the spectrum. Similar to what was observed in Table 4-1, in Table 4-2 (i.e. Acids & Bases- Chemistry-20), CoE #8 is still near the lowest frequency and CoE #5 persists near the highest frequency.

The third table in the Nelson series, Table 4-3, depicts the frequency of the CoEs by Situation, still in the topic "Acids and Bases," but at the Chemistry-30 level (i.e. Grade 12). The most prominent feature of this table is the dramatic six-fold increase in the number of Situations from 4 in Table 4-2 to 24 in Table 4-3. The corresponding number of pages in the Nelson textbook has only increased three-fold from 13 to 39. The number of CoEs/Situation, range from a minimum of 14 to a maximum of 59. All the

Table 4-3
Frequency of CoEs by Situation in the Nelson Series (Acids and Bases -- Chem 30)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
24	1	2	2	1	3	4	4	0	2	2	4	2	5	2	1	2	3	3	2	3	48
25	1	2	2	2	2	3	4	0	1	2	4	2	3	2	1	2	1	3	3	2	42
26	1	4	1	1	1	3	3	0	0	2	3	3	2	1	0	2	1	1	1	2	32
27	1	3	0	1	1	0	0	0	0	2	3	2	2	0	0	0	0	0	0	2	17
28	1	3	2	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	14
29	1	3	2	1	3	3	2	3	0	2	3	2	3	1	1	2	1	2	3	2	40
30	1	3	2	1	3	3	2	3	0	2	3	2	3	1	1	2	1	2	3	2	40
31	1	1	0	1	2	0	0	0	0	0	4	2	0	1	1	0	1	2	3	3	22
32	2	3	1	1	4	4	4	0	1	2	4	2	3	2	1	2	1	2	3	3	45
33	2	3	2	3	4	3	4	3	1	2	4	2	4	2	1	2	1	2	3	3	51
34	2	4	0	2	2	0	2	0	0	2	0	0	0	1	1	2	1	2	3	2	26
35	1	2	2	2	4	4	4	0	1	2	2	2	2	2	1	2	2	3	2	2	42
36	3	2	2	3	4	3	4	3	1	2	3	3	4	2	1	2	1	2	3	3	51
37	1	2	2	2	5	4	4	4	1	2	4	2	5	2	1	2	2	3	3	2	53
38	2	2	2	2	7	4	4	4	1	2	4	2	5	2	1	2	2	3	3	2	56
39	1	1	0	0	0	0	0	0	0	2	3	2	1	1	1	2	1	2	2	0	19
40	3	3	2	2	4	4	4	4	1	2	3	2	4	2	1	2	2	3	2	3	53
41	3	3	0	0	0	0	0	0	0	0	3	2	5	1	1	2	1	5	3	4	33
42	1	3	2	1	6	4	3	0	0	2	0	1	1	0	1	2	1	0	1	2	31
43	1	1	2	1	3	4	0	0	0	0	2	2	1	0	0	2	0	0	0	2	21
44	2	2	2	2	5	4	4	4	1	2	0	2	5	2	1	2	2	3	3	2	50
45	2	2	2	2	7	4	4	4	1	2	4	2	5	2	1	2	2	3	3	2	56
46	3	2	2	1	6	4	4	0	1	2	3	2	5	2	1	2	1	3	2	3	49
47	3	3	2	2	6	4	4	4	1	2	3	2	6	2	1	2	1	4	3	4	59
Total	40	59	36	36	85	66	64	36	14	40	66	45	74	33	20	42	29	53	54	58	950
quency) Words f Text	2	2.77	1.69	1.69	3.99	3.1	3.01	1.69	0.66	1.88	3.1	2.11	3.48	1.55	0.94	1.97	1.36	2.49	2.54	2.72	44.74

Situations listed in this table use a significant number of CoEs and generally they use more CoEs than the Situations in Table 4-2. The mean number of CoEs/Situation is 40 in Table 4-2. The individual CoEs range in total frequency from 14 for CoE #9 to 85 for CoE #5. CoE #9, with the lowest frequency, is "Statistical Treatment of Measurements of a Single Datum," while CoE #5, with the highest frequency, is "Instruments: Calibration and Error."

McGraw-Hill "Chemistry"

Table 4-4 lists the "Frequency of CoE by Situation," in the McGraw-Hill series under the topic "Solution," at the Chemistry-20 level. There is a total of 10 Situations in this table. All Situations have a substantial number of CoEs ranging from 20 to 58. The mean number of CoEs/Situation is 38.5. Although some CoEs have zero representation in certain Situations, such CoEs are fewer than in Table 4-1 under Nelson. The last row of Table 4-4 shows the summed up total occurrence of each CoE. The individual CoEs appear with frequencies ranging from 7 (for CoE #4, Instruments: underlying relationships; for CoE #9, Statistical Treatment of Measurements of a Single Datum; and for CoE #15, Design: Tables) to 37 (for CoE #5, Instruments: calibration and error).

Table 4-5 displays the "Frequency of CoEs by Situation," in the McGraw-Hill series, still at the Chemistry-20 level, but under a different topic – Acids and Bases. This table has a total of five Situations. Some of the CoEs here have zero representation: this is especially true of CoE #8 which is Sampling a datum. All Situations have a substantial number of CoEs, ranging from 46 to 60. The Situations average out with a mean of 53.8

Table 4-4

Frequency of CoEs by Situation in the McGraw-Hill Series (Solutions - Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
48	2	2	2	1	7	4	0	0	0	2	2	2	5	1	1	2	1	3	2	3	42
49	1	3	0	0	0	0	0	0	0	0	2	2	3	1	1	2	1	2	3	3	24
50	2	2	2	1	4	4	4	0	1	2	2	2	3	2	1	2	1	2	3	2	42
51	1	1	0	0	0	4	3	0	1	2	2	2	3	1	0	2	0	2	2	3	29
52	3	2	2	2	6	4	3	4	1	2	2	2	5	3	1	2	1	3	3	3	54
53	3	1	0	0	0	4	3	0	0	2	2	2	0	0	0	0	0	0	0	3	20
54	3	1	0	0	0	4	3	0	0	2	2	2	0	0	0	0	0	0	0	3	20
55	3	2	2	1	6	4	4	4	1	2	2	2	5	2	1	2	1	2	3	3	52
56	1	2	2	1	7	4	4	0	1	2	2	1	3	2	1	2	1	2	3	3	44
57	4	2	2	1	7	4	4	4	2	2	2	3	6	2	1	2	2	2	3	3	58
Total	23	18	12	7	37	36	28	12	7	18	20	20	33	14	7	16	8	18	22	29	385
Frequency of CoEs/ 500 Words of text	1.44	1.13	0.75	0.44	2.31	2.25	1.75	0.75	0.44	1.13	1.25	1.25	2.06	0.88	0.44	1	0.5	1.13	1.38	1.81	24.09

Table 4-5

Frequency of CoEs by Situation in the McGraw-Hill Series (Acids and Bases - Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
58	4	3	2	1	5	4	4	0	2	2	3	3	4	2	1	2	1	2	3	3	51
59	4	3	2	2	5	4	4	0	2	2	4	2	4	2	1	2	1	4	3	3	54
60	4	4	2	0	0	4	4	0	2	2	3	2	5	2	1	2	0	3	3	3	46
61	4	3	2	3	6	4	4	0	2	2	3	2	5	2	1	2	2	4	3	4	58
62	4	4	2	2	9	4	4	4	1	2	2	2	5	2	1	2	1	3	3	3	60
Total	20	17	10	8	25	20	20	4	9	10	15	11	23	10	5	10	5	16	15	16	269
Frequency of CoEs/ 500 Words of text	1.39	1.18	0.69	0.56	1.74	1.39	1.39	0.28	0.62	0.69	1.04	0.76	1.6	0.69	0.35	0.69	0.35	1.11	1.04	1.11	18.67

Table 4-6

Frequency of CoEs by Situation in the McGraw-Hill Series (Acids and Bases - Chem 30)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
63	4	4	2	1	6	4	4	4	2	2	3	2	6	2	1	2	1	4	3	3	60
64	4	4	2	2	8	4	4	4	2	2	3	2	6	2	1	2	2	5	3	3	65
65	3	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	5	3	3	58
66	3	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	5	3	3	58
67	3	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	5	3	3	58
Total	17	20	10	9	35	20	20	8	10	10	15	10	27	10	5	10	9	24	15	15	299
Frequency of CoEs/ 500 Words of text	1.3	1.53	0.77	0.69	2.69	1.53	1.53	0.61	0.77	0.77	1.15	0.77	2.07	0.77	0.38	0.77	0.69	1.84	1.15	1.15	22.93

CoEs/Situation. The individual CoEs appear with different frequencies ranging from 4 (for CoE #8, Sampling a datum) to 25 (for CoE #5, Instruments: calibration and error).

Table 4-6 depicts the "Frequency of CoEs by Situation" in the McGraw-Hill series at the Chemistry-30 level, but still in the topic Acids and Bases. There are only five Situations. The only CoE that is not represented is CoE #8. All the Situations listed in this table have a very substantial number of CoEs ranging from 58 to 65, with a mean of 59.8 CoEs/Situation. Individual CoEs appear with different frequencies ranging from 5 (for CoE #15, Design: Tables) to 35 (for CoE #5, Instruments: calibration and error). In all the tables (i.e. Tables 4-4, 4-5, and 4-6) of McGraw-Hill Chemistry the CoE, with the greatest frequency has consistently been CoE #5, Instruments: calibration and error.

Chang "Chemistry"

Table 4-7 shows the "Frequency of CoE by Situation," in the Chang series under the topic "Solutions," at the Chemistry-20 level. There is a total of 7 Situations in this topic. Some CoEs have varying degrees of zero representation e.g. CoEs #5, #7, #8 and #9. Of these, CoE #8 and CoE #9 are not represented at all, i.e. have zero representation, throughout. By studying the last column in the Table 4-7, it is evident that all the Situations in this topic have a substantial number of CoEs ranging from 25 to 52, with a

Table 4-7

Frequency of CoEs by Situations in the Chang Series (Solutions - Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
68	4	3	2	1	6	4	4	0	0	2	3	2	5	2	1	2	1	4	3	3	52
69	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
70	1	4	0	1	0	4	0	0	0	2	0	0	0	0	1	2	2	2	3	3	25
71	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
72	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
73	4	3	2	1	0	4	0	0	0	2	3	2	3	2	1	2	1	4	3	3	40
74	4	2	2	1	0	4	0	0	0	2	3	2	5	2	1	2	1	4	3	3	41
Total	25	21	12	7	6	28	4	0	0	14	15	12	22	12	7	14	8	26	21	21	275
Frequency of CoEs/ 500 Words of text	6.35	5.34	3.05	1.78	1.53	7.12	1.02	0	0	3.56	3.81	3.05	5.59	3.05	1.78	3.56	2.03	6.61	5.3	5.34	69.91

mean of 40 CoEs/Situation. As in the previous publishers, the last row entitled "Total" shows the total occurrence of each CoE. The individual CoEs appear with frequencies ranging from 0 (CoE #8, Sampling a datum, CoE #9, Statistical treatment of a single datum, and CoE #15, Design: Tables) to 28 (CoE #6, Reliability and validity of a single measurement).

Table 4-8 tabulates the frequency of CoEs by Situations in the Chang series still at the Chemistry-20 level, but under a different topic, "Acids and Bases." The layout of Table 4-8 is similar to that described for Table 4-7 earlier. There is a total of only 3 Situations under this topic, and this is far fewer than the 7 Situations present in Table 4-7 although the number of pages covered in the Chang "Chemistry" textbook itself, ranges from 5 pages for Table 4-7, to 12 pages for Table 4-8. Some CoEs have zero representation, especially CoE #8, Sampling a datum. All the Situations have a substantial number of CoEs, ranging from 39 to 55, and with a mean of 45 CoEs. The individual CoEs appear with different frequencies ranging from 0 (for CoE #8, Sampling a datum) to 13 (for CoE # 13, Design: Choosing values).

Table 4-9 depicts the "Frequency of CoEs by Situations," in the Chang series at a different level, i.e. Chemitry-30, but still in the topic "Acids and Bases". There is a total of 7 Situations in this topic at the Chemistry-30 level (Table 4-9) compared to the 3 Situations at the Chemistry-20 (Table 4-8). The number of pages covered in the Chang "Chemistry" textbook itself range from 29 pages in Table 4-9 to 7 pages in Table 4-8.

Table 4-8

Frequency of CoEs by Situations in the Chang Series (Acids and Bases - Chem 20)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
75	4	2	2	1	0	4	0	0	0	2	3	2	4	2	1	2	1	2	3	4	39
76	4	3	2	1	0	4	0	0	0	2	3	2	4	2	1	2	1	3	3	3	40
77	4	3	2	1	8	4	4	0	1	2	3	2	5	2	1	2	1	4	3	3	55
Total	12	8	6	3	8	12	4	0	1	6	9	6	13	6	3	6	3	9	9	10	134
Frequency of CoEs/ 500 Words of text	2.86	1.9	1.43	0.71	1.9	2.86	0.95	0	0.24	1.43	2.14	1.43	3.1	1.43	0.71	1.43	0.71	2.14	2.1	2.38	31.89

Table 4-9

Frequency of CoEs by Situations in the Chang Series (Acids and Bases - Chem 30)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
78	4	3	2	1	0	4	0	0	0	2	4	2	5	2	1	2	1	3	3	3	42
79	4	3	2	1	0	4	0	0	0	2	3	2	5	2	1	2	1	3	3	3	41
80	3	2	0	0	0	4	0	0	0	2	3	2	0	0	0	0	0	2	3	7	28
81	4	2	2	1	0	4	0	0	0	2	3	2	4	1	1	2	1	3	3	3	38
82	4	5	2	1	0	4	0	0	0	2	3	2	5	2	1	2	1	2	3	3	42
83	4	3	2	1	7	4	4	0	2	2	2	2	5	2	1	2	2	4	3	3	55
84	1	3	2	2	3	4	3	0	0	2	3	1	5	2	1	2	1	2	3	3	43
Total	24	21	12	7	10	28	7	0	2	14	21	13	29	11	6	12	7	19	21	25	289
Frequency of CoEs/ 500 Words of text	1.72	1.51	0.86	0.5	0.72	2.01	0.5	0	0.14	1	1.51	0.93	2.08	0.79	0.43	0.86	0.5	1.36	1.5	1.79	20.72

All Situations have a very substantial number of CoEs, ranging from 28 to 55, but with a mean of 41 CoEs/Situation. The total number of the individual CoEs appear with frequencies ranging from 0 (for CoE #8, Sampling a datum) to 29 (for CoE #13, Design: Choosing values). CoE #6 follows closely at the high end of this spectrum with a frequency of 28.

Petrucci "General Chemistry"

Table 4-10 shows the "Frequency of CoEs by Situation," in the Petrucci Series under the topic "Solutions," at the Chemistry-20 level. As enumerated in the first column, there are 7 Situations (listed as Situations 85 to 91) in this topic. The general trend is that most Situations have a significant number of CoEs, ranging on the whole from 39 to 49; Situation 88, however, is an exception with only 5 CoEs. The mean number of CoEs/Situation is 36 CoEs. At the highest end of the spectrum, CoE #89 had a total of 49 CoEs. The individual CoEs appear with different frequencies ranging from 2 (for CoE #9, Statisical treatment of measurements of a single datum) to 24. The frequency of 24 is shared by three CoEs which are: CoE #1, Fundamental ideas; CoE #6, Reliability and validity of a single measurement; and CoE #13, Design: Choosing values.

In Table 4-11 we are looking at the "Frequency of CoE by Situation" in the Petrucci series under the topic "Acids and Bases" which is covered in Chapter 17 of this textbook. Chapter 17 is the introductory chapter on the wide topic "Acids and Bases." There are approximately half the number of Situations in Table 4-11 when compared to Table 4-10, in this series (i.e. 4 in Table 4-11 compared to 7 in Table 4-10) although the actual page numbers vary from 12 in Table 4-10 compared to 33 in Table 4-11. All Situations, however, have a significant number

Table 4-10

Frequency of CoEs by Situation in the Petrucci Series (Solutions)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
85	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
86	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
87	4	3	2	1	0	4	0	0	0	2	2	2	3	2	1	2	1	4	3	3	39
88	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5
89	3	3	2	1	3	4	4	0	1	2	3	2	5	2	1	2	1	4	3	3	49
90	4	3	2	1	0	4	0	4	1	2	3	2	5	1	1	2	1	3	3	3	45
91	4	3	2	2	0	4	0	0	0	0	3	2	5	1	1	2	1	3	3	3	39
Total	24	19	12	7	3	24	4	4	2	10	15	12	24	10	6	12	6	22	18	21	255
Frequency of CoEs/ 500 Words of text	3.24	2.57	1.62	0.95	0.41	3.24	0.54	0.54	0.27	1.35	2.03	1.62	3.24	1.35	0.81	1.62	0.81	2.97	2.43	2.84	34.45

Table 4-11

Frequency of CoEs by Situation in the Petrucci Series (Acids and Bases - Chapter 17)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
92	4	3	2	1	7	4	4	0	2	2	2	2	5	2	1	2	2	3	3	3	54
93	4	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	3	3	3	57
94	3	2	0	0	0	4	0	0	0	2	3	1	0	1	1	0	1	3	3	7	31
95	4	4	2	2	7	4	4	0	1	2	3	2	5	2	1	2	2	3	3	3	56
Total	15	13	6	5	21	16	12	0	5	8	11	7	15	7	4	6	7	12	12	16	198
Frequency of CoEs/ 500 Words of text	0.85	0.74	0.34	0.28	1.19	0.91	0.68	0	0.28	0.45	0.63	0.4	0.85	0.4	0.23	0.34	0.4	0.68	0.68	0.91	11.24

Table 4-12

Frequency of Coe by Situation in the Petrucci Series (/ (Acids and Bases - Chapter 18)

	Concepts of Evidence																				
Sit. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	Total
96	4	4	2	1	4	4	4	0	1	2	2	2	4	2	1	2	1	4	3	3	50
97	4	4	2	2	7	4	4	0	1	2	3	2	5	2	1	2	2	3	3	3	56
98	4	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	3	3	3	57
99	4	4	2	2	7	4	4	0	2	2	3	2	5	2	1	2	2	3	3	3	57
100	4	4	1	1	6	4	4	0	1	2	3	2	5	1	1	2	1	3	3	3	51
101	4	3	2	1	3	4	0	0	1	2	3	2	4	1	1	2	1	3	3	3	43
Total	24	23	11	9	34	24	20	0	8	12	17	12	28	10	6	12	9	19	18	18	314
Frequency of CoEs/ 500 Words of text	1.99	1.91	0.91	0.75	2.83	1.99	1.66	0	0.66	1	1.41	1	2.33	0.83	0.5	1	0.75	1.58	1.5	1.5	26.1

of CoEs ranging in number from 31 to 56. The mean number of CoEs/Situation in this table is 50 CoEs. The individual CoEs have various frequencies, ranging from CoE #8 which has a frequency of 0 (i.e. zero representation) to CoE #5 which has the highest frequency of 21.

Table 4-12 is compiled using Chapter 18 on "Acids & Bases – Advanced" from the Petrucci "General Chemistry." Chapter 18 is a development of the acid-base concepts introduced in Chapter 17, and therefore, may be held comparable to Chemistry-30 which builds on the acid-base concepts introduced in Chemistry-20. This Table 4-12 has more Situations than Table 4-11 (i.e. 6 Situations in Table 4-12 compared to 4 Situations in Table 4-11) leading to a 3:2 ratio, although the pages covering these two topics are in the ratio 22:32 which works out to approximately 7:10. All the Situations in this table have a significant number of CoEs ranging from 43 to 57, and averaging out with a mean of 52 CoEs/Situation. The individual CoEs have frequencies that differ from 0 (which is the frequency of CoE #8, Sampling a datum) to 34 (which is the frequency of CoE #5, Instruments: calibration and error).

Comparisons Among Publishers, Topics and Grades

The data examined so far in Tables 4-1 to 4-12 have been publisher-specific, dealing with each publisher/author in turn. I have compiled the final results from the various publishers/authors to make comparisons, detect trends, and thereby address some of our research questions.

Figure 4-1 compares the total number of Situations/500 words of text in the four textbooks studied. It is quite evident, that Nelson has the highest number of

Figure 4-1. **Total number of situations/500 words of text in the four textbooks**

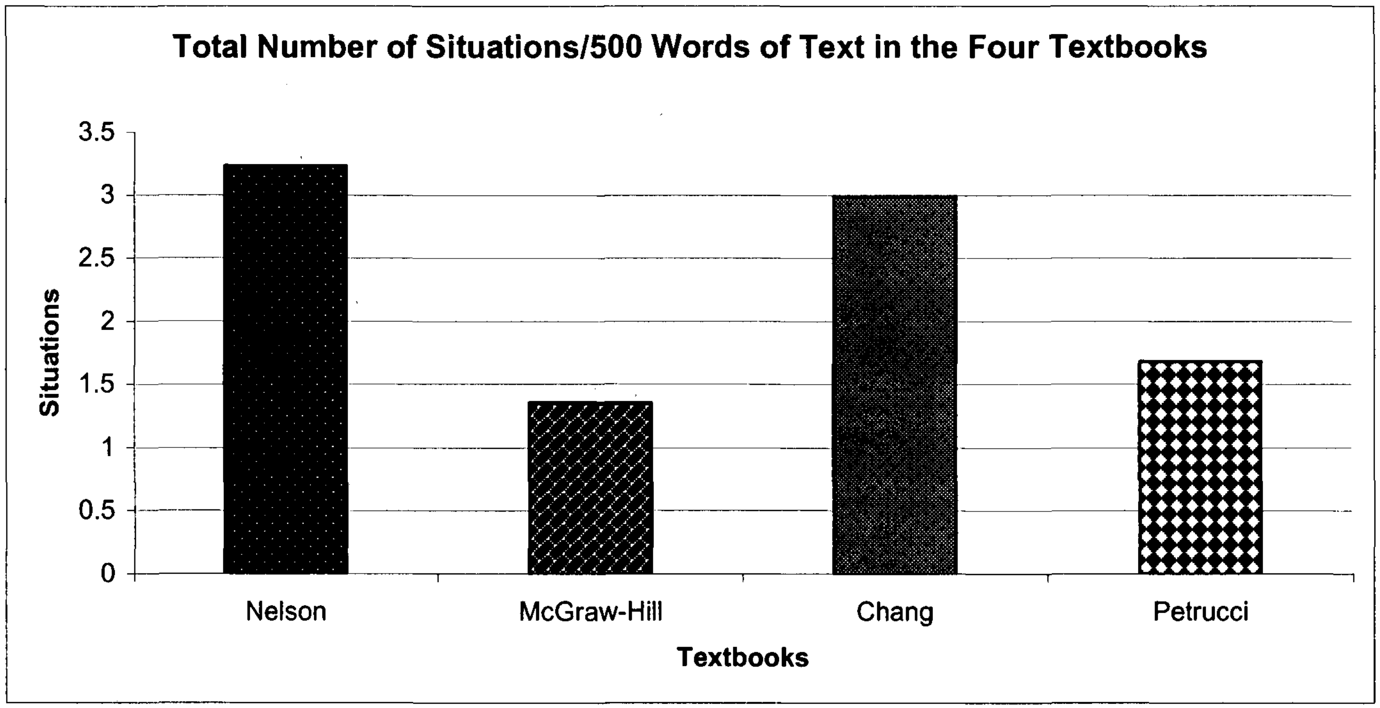
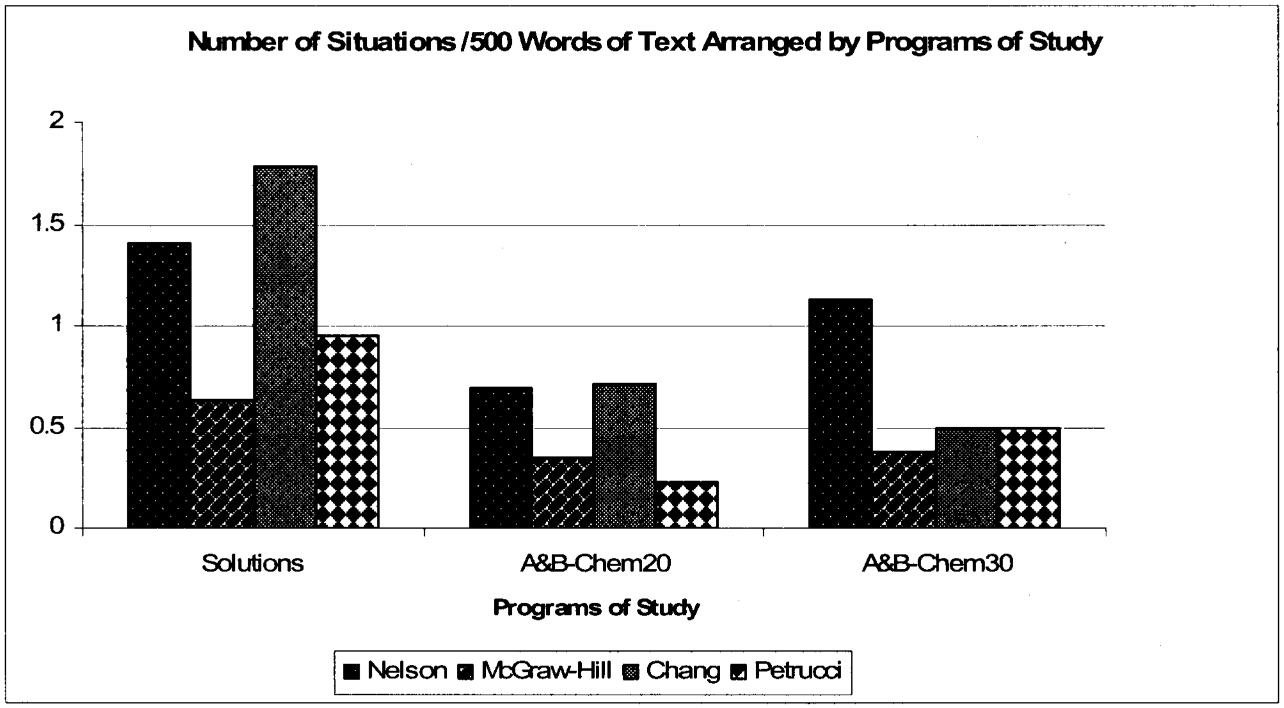


Figure 4-2. **Number of situations/500 words of text arranged by programs of study**



Situations/500 words of text, followed closely in second place by Chang. At the other end of the distribution, is McGraw-Hill with the lowest number of Situations/500 words of text, with Petrucci in close proximity, having about the same number of Situations/500 words of text).

Figure 4-2 depicts the number of Situations/500 words of study arranged according to the three topics studied which were "Solutions," "Acids & Bases – Chemistry-20," and "Acids & Bases – Chemistry-30." Within each topic, the textbooks have the same shading scheme as in Figure 4-1. The greatest number of Situations in Figure 4-2 was found in the topic "Solutions": this was followed by "Acids & Bases – Chemistry-30". Under the topic "Solutions" itself, the Situations/500 words of text showed variance according to the publisher/author. Chang had the highest number of Situations/500 words of text, followed in second place by Nelson. Petrucci was in third position and the lowest number was displayed by McGraw-Hill.

Under the topic "Acids & Bases – Chemistry-30" it is Nelson that has the highest number of Situations/500 words of text, and the second place, in this topic, is jointly held by Chang and Petrucci, both of which have 0.5 Situations/500 words of text. McGraw-Hill was in third place. Chemistry-30 is usually taught in Grade 12. The topic "Acids & Bases – Chemistry-20" displays Chang as having the highest number of Situations, but is followed closely by Nelson. The third place in this topic is held by McGraw-Hill and the lowest position is shown by Petrucci. Chemistry-20 is usually covered in Grade 11.

Figure 4-3. **Mean number of CoEs/Situation in the four textbooks**

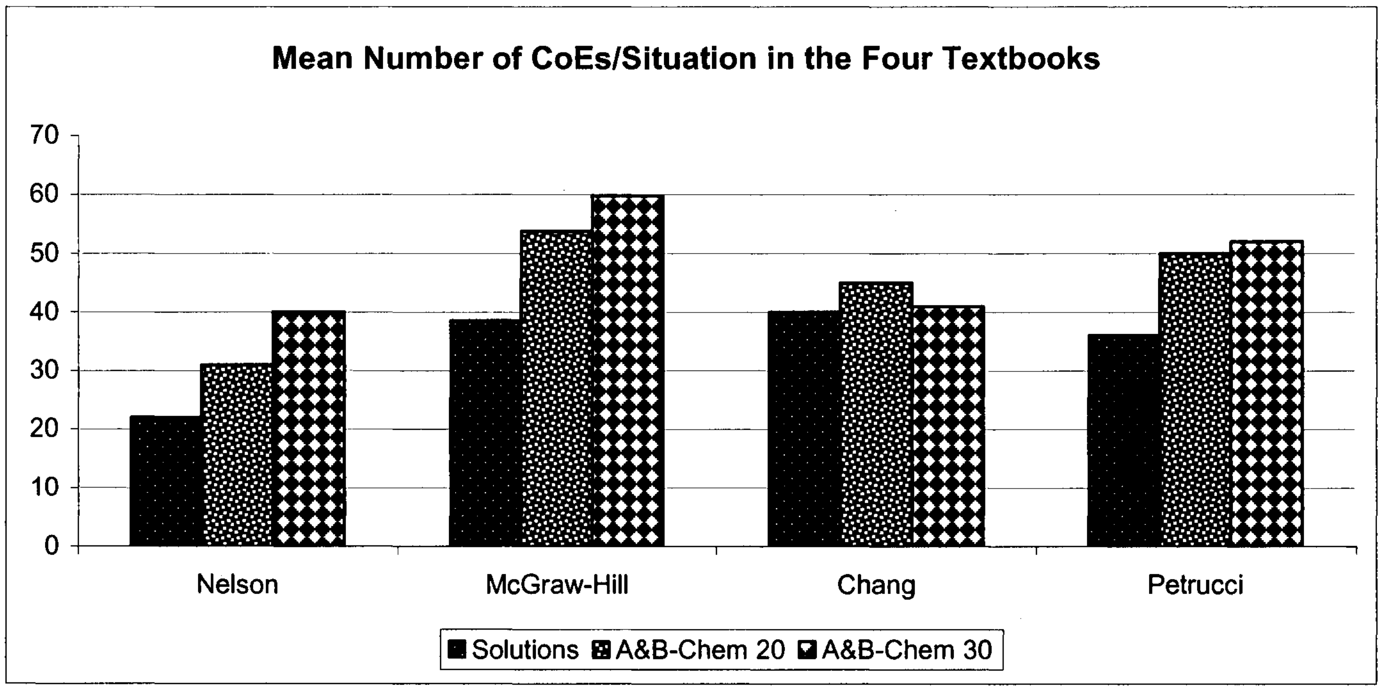


Figure 4-3 looks at the mean number of CoEs/Situation, in each of the four textbooks. In other words, it is looking at how laden the Situations are with CoEs, or the density of CoEs within the Situations. On closely examining Figure 4-3, a trend is perceived that persists in three of the textbooks examined. This trend is that "Acids & Bases – Chemistry-30" has the greatest number of CoEs/Situation, followed in second place by "Acids & Bases – Chemistry-20", and in third place by "Solutions – Chemistry-20". Therefore, in Nelson McGraw-Hill and Petrucci, the density of CoEs within the Situations, arranged by topics of study as well as by grades, in descending order is: "Acids & Bases – Chemistry-30", "Acids & Bases – Chemistry-20", and "Solutions – Chemistry-20".

The exceptional case in Figure 4-3 is Chang, where the trend observed above does not exist. In Chang, the highest density of CoEs is observed in "Acids & Bases – Chemistry-20", and "Acids & Bases – Chemistry-30", which follows closely in second place, has Situations which are less laden with CoEs. The third place in Chang is taken by "Solutions – Chemistry-20".

Table 4-13(A) shows the frequency of individual CoEs/500 words of text under the topic "Solutions – Chemistry-20," in all the four textbooks examined. The CoEs that occur with the first highest, second highest and third highest frequencies were noted. Table 4-13(A) lists these CoEs and their corresponding frequencies. From this list, under each category, the CoE that appears most often (i.e. the mode) was selected and listed in Table 4-13(A). The mode for the first highest position in the four textbooks falls to CoE #5, which is "Instruments: calibration and error." The second highest frequency that occurs most frequently (mode) falls to CoE #6, which is "Reliability and validity of a single measurement." The mode for the third highest frequency is CoE #13, which is "Design: choosing values."

Table 4-13(B) also looks at the frequency of CoEs/500 words of text, under the same topic "Solutions – Chemistry-20," but this table picks out the first, second and third lowest frequencies, and thereby concentrates on the other end of the distribution, as compared to Table 4-13(A). The mode for the CoE with the lowest frequency falls to CoE #8, which is "Sampling a datum". The mode for the second lowest frequency goes to CoE #9, "Statistical treatment of measurements of a single datum." The mode for the third lowest frequency of CoEs/500 words of text in all the four textbooks rests on CoE #7, which is "The choice of an instrument for measuring a datum."

In Table 4-14(A) the frequency of individual CoEs/500 words of text under the topic "Acids & Bases – Chemistry-20," is examined in all the four textbooks. The CoEs with the highest, second highest, and third highest frequency were noted. From the first highest category in this table, the CoE/500 words that appears most frequently in the four textbooks is CoE #5, which is "Instruments: calibration and error." There are four CoEs/500 words of text that appear in the second highest category: these are CoE #1, "Fundamental ideas"; CoE #5, "Instruments: calibration and error"; CoE #6, "Reliability and validity of a single measurement"; and CoE #13, "Design: choosing values". All four are chosen because no clear mode is expressed in this category. The most frequently occurring CoEs/500 words of text for the third highest frequency are CoE #6, which is "Reliability and validity of a single measurement," and CoE #7, which stands for "The choice of an instrument for measuring a datum."

Table 4-13 (A)

Highest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Solutions - Chemistry-20

	Highest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	3.79	5	3.57	7	2.6	6
McGraw-Hill	2.31	5	2.25	6	2.06	13
Chang	7.12	6	6.61	19	6.35	1
Petrucci	3.24	1	3.24.	6	3.24	13
Mode of CoEs		5		6		13

Table 4-13 (B)

Lowest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Solutions - Chemistry-20

	Lowest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	0.15	8	0.3	9	0.52	17
McGraw-Hill	0.44	4	0.44	9	0.44	15
Chang	0	8	0	9	1.02	7
Petrucci	0.27	9	0.41	5	0.54	7& 8
Mode of CoEs		8		9		7

Table 4-14 (A)

Highest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Acids & Bases - Chemistry-20

	Highest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	2.06	6	1.88	5	1.88	7
McGraw-Hill	1.74	5	1.6	13	1.39	6&7
Chang	3.1	13	2.86	1	2.86	6
Petrucci	1.19	5	0.91	6	0.91	21
Mode of CoEs		5		N/A		6&7

Table 4-14 (B)

Lowest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Acids & Bases - Chemistry-20

	Lowest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	0.34	9	0.34	15	0.51	8&20
McGraw-Hill	0.28	8	0.35	15	0.35	17
Chang	0	8	0.24	9	0.71	4&15&17
Petrucci	0	8	0.23	15	0.28	9
Mode of CoEs		8		15		17

Table 4-14(B) also looks at the frequency of CoEs/500 words of text under the same topic, "Acids & Bases – Chemistry-20," but this table focuses on the opposite end of the spectrum compared to Table 4-14(A), in that it picks out the first, second and third lowest frequencies of CoEs/500 words of text. The most frequently occurring CoE for the first lowest frequency is CoE #8, which is "Sampling a datum." The mode for the second lowest frequency is CoE # 15, which is "Design: Tables." The mode for the third lowest frequency of CoE/500 words of text in all the four textbooks is CoE #17, classified as "Data presentation."

Table 4-15(A) considers the frequency of individual CoEs/500 words of text under the topic "Acids & Bases – Chemistry-30," in all the four textbooks. The CoEs with the first highest, second highest, and third highest frequency were noted. From the first highest category in Table 4-15(A), the CoE/500 words that appear most frequently in the four textbooks is CoE #5, which is "Instruments: calibration and error." The mode for the second highest frequency of CoEs/500 words of text falls to CoE #13, which is "Design: choosing values." The mode for the third highest frequency in CoE/500 words of text rests on CoE #6, representing "Reliability and validity of a single measurement."

In Table 4-15(B) the frequency of CoEs/500 words of text under the same topic, "Acids & Bases – Chemistry-30," is examined in all the four textbooks. The CoEs/500 words of text occurring with the first, second and third lowest frequencies are noted. The first lowest frequency, falls to CoE #8, which is "Sampling a datum." The mode for the second lowest frequency is shown by CoE #15, which stands for "Design: Tables." CoE #17 which is "Data Presentation" is the mode for the third lowest frequency.

Table 4-15 (A)
Highest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Acids & Bases - Chemistry-30

	Highest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	3.99	5	3.48	13	3.1	6&11
McGraw-Hill	2.69	5	2.07	13	1.84	19
Chang	2.08	13	2.01	6	1.79	21
Petrucci	2.83	5	2.33	13	1.99	1&6
Mode of CoEs		5		13		6

Table 4-15 (B)
Lowest Frequency of CoEs/500 Words of text in the four textbooks studied under the topic Acids & Bases - Chemistry-30

	Lowest Frequency of CoEs/500 Words of text					
	First		Second		Third	
	Frequency	CoE	Frequency	CoE	Frequency	CoE
Nelson	0.66	9	0.94	15	1.36	17
McGraw-Hill	0.38	15	0.69	4	0.69	17
Chang	0	8	0.14	9	0.43	15
Petrucci	0	8	0.5	15	0.66	9
Mode of CoEs		8		15		17

Table 4-16

The CoEs with Highest and Lowest Frequency of Occurrence by Publisher

Textbook	Highest Frequency	Lowest Frequency
Nelson	5	4
McGraw-Hill	5	4,8,15
Chang	13	8
Petrucci	5	8

The findings of Tables 4-13, 4-14 and 4-15 for the highest and lowest frequency of occurrence of CoEs/500 words of text may be summarized in Table 4-16. By studying Table 4-16 it is evident that the most prevalent CoE/500 words of text, with highest level of occurrence in all the four textbooks examined is CoE #5, representing "Instruments: calibration and error." At the other end of the spectrum, the most common CoE/500 words of text, with the lowest level of occurrence, in all the four textbooks looked at, is CoE #8, which is "Sampling a datum."

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND IMPLICATIONS

Having presented in Chapter Four the results of examining the concepts of evidence under the two topics "Solutions" and "Acids and Bases" in the four chemistry textbooks used in this study, I intend to devote this chapter to: a summary of the results obtained, further questions that can be raised, and final comments on implications for practice.

According to their vision of what a scientifically literate person is, the *National Science Education Standards* (National Research Council, 1996) stated:

> A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately. (p. 22)

This outlines the vital role of evidence as the basis on which arguments are laid out and weighed in life situations. One develops the capability to draw conclusions from such evaluations of evidence and to apply them rigorously, only if one is scientifically literate. Being scientifically literate enables one to evaluate scientific information based on where it comes from and how it was generated. It is important, therefore, that students get introduced to scientific evidence through the concepts of evidence that they encounter in their science classes via their lessons, their textbooks, their investigations and their laboratory work. Since one of the main sources where students encounter these concepts of evidence is their textbook, it is very important to study how concepts of evidence are laid out in them. Chemistry textbooks at the high school and undergraduate level in

Canada have not been examined so far for their representation of evidence, making this study unique and essential.

Summary of Results

The leading question addressed in this study was: "How is evidence represented in the university undergraduate and high school chemistry textbooks?" This study addressed a perennial problem in the high school chemistry curriculum, which is the failure to treat in a systematic and comprehensive way the nature of scientific evidence. I have completed a thorough examination of the three most widely used high school chemistry textbooks in Canada and a widely used first-year university textbook. I have found that the treatment of CoEs varies widely across programs, within programs across topics, and within and across programs by the concepts of evidence themselves. Some concepts of evidence were laid out explicitly in the textbooks while the majority was only implied. In the chapter "Solutions" from Nelson "Chemistry," for example, 22.5% of the total number of CoEs were explicitly and clearly laid out, while the majority (i.e. 77.5%), of CoEs had only implicit reference. These data will, I hope, point us the directions that most need addressing.

Question 1 – Comparison by Grade, Topic and Publisher

Besides the leading question, I also raised more specific questions. The first specific question I raised was: "How frequently do these CoEs occur by grade level, by topic and by publisher?" The similar grade level comparison occurs between Solutions which is covered in Chemistry-20 and the initial part of Acids and Bases, which is also done as part of Chemistry-20. In all publishers studied, there are more Situations in Solutions Chemistry-20 than in Acids and Bases Chemistry-20.

The comparison across grades in the same topic occurs between Acids and Bases as covered in both Chemistry-20 and Chemistry-30. There are more Situations in Chemistry-30 than in Chemistry-20 in all publishers except McGraw-Hill, which has the same number of Situations in both Chemistry-20 and -30. Chemistry-20 covers basic concepts under this topic such as the effect of acids and bases on aqueous systems, pH, using indicators, calculating hydrogen ion and hydroxide ion concentrations for strong acids and bases, etc. In Chemistry-30, however, the topics covered are more complex such as, explaining pH/pOH scale in terms of logarithms; defining Kw, Ka and Kb; calculating hydronium ion concentrations and hydroxide ion concentrations; pH and pOH for solutions using the ionization constant for water, Kw; performing a titration and related calculations to determine the concentration of an acid or base solution, etc. The CoEs identified under the Situations in Chemistry-30 material draw upon the fundamental knowledge gained in Chemistry-20 and build upon it. The concepts covered, and the Situations identified, are more complex in Chemistry-30.

When comparing across publishers it is evident that Nelson has highest number of situations/500 words of text followed in second place by Chang. The third position is taken by Petrucci and McGraw-Hill has the lowest number of Situations/500 words of text.

There are more Situations/500 Words of Text in Solutions than in the other topic examined in this study. A related question that arises is whether this preponderance of CoEs in the chapter on Solutions is an intended pattern or just an occurrence brought about by chance. The chapter on Solutions lends itself to more labs and investigations and demonstrations partly because a lot of topics are covered in it. Topics that are covered include: classification of solutions as electrolytes and non-electrolytes; chemical analysis for identifying solutions; the behaviour of acidic, basic and neutral solutions; the behaviour of substances in water; acid nomenclature; qualitative and quantitative analysis of unknown solutions; the iodine clock reaction; sequential qualitative analysis of unknown solutions; qualitative analysis by colour and flame test; communicating concentrations and concentration ratios of solutions; dilution; solution preparation; communicating hydrogen ion concentration; solubility rules and examples; the effect of temperature on solubility; solubility equilibrium; and testing the theory of dynamic equilibrium. Since such a large array of topics are covered, and since most topics have relevant investigations, laboratory experiments and demonstrations, there are a large number situations in this chapter, more than in any other chapter examined in this study. In the chapter, "Solutions," however, the number of CoEs/Situation varies and in some there are only a few CoEs/Situation (i.e. the lowest being 1CoE/Situation) whereas in others there are as many (i.e. the highest being 71CoEs/Situation).

Question 2 – Range and Density of CoEs

The second question that was raised was: "From the different types of CoEs employed, what is the range in the density of CoEs/Situation?" In three out of the four textbooks studied – Nelson, McGraw-Hill and Petrucci - the density of the CoEs/Situation is greatest for "Acids & Bases – Chemistry-30," followed in second place by "Acids & Bases – Chemistry-20," and in third place by "Solutions – Chemistry-20. One would expect this because "Acids & Bases – Chemistry-30" has experiments such as titrations in which all 21 different CoEs are represented. The Situations in "Acids & Bases – Chemistry-30" were, on the average, more laden with CoEs than the Situations in "Solutions – Chemistry-20."

Question 3 – Gaps in the occurrences of the CoEs

The third question was: "Are there any gaps in the depth, accuracy and appropriateness of the occurrences of the CoEs in textbooks when compared to Gott's classification system?" As mentioned earlier under the summary of results in Chapter 5, in the chapter "Solutions" from Nelson "Chemistry," only 22.5% of the total number of CoEs were laid out clearly and explicitly. This suggests that there are gaps in the depth, accuracy and appropriateness of the occurrences of the CoEs in the textbooks. The remaining 77.5% of CoEs were given implicit reference by the authors. Although all 21 categories of CoEs are represented in the textbooks, there are some sub-categories that are not represented in them. These include: [**1**](5), [**2**](6), [**5**](6), [**11**](7), [**12**](3)(4), [**17**](2)(3), [**19**](4)(8), [**21**](3)(4)(5). The sub-categories shown above constitute gaps in the occurrence of CoEs in the textbooks when compared to Gott's classification system. Some of these sub-categories are absent because they are inappropriate for chemistry at this level. For example, [**2**](6) is in the main category of "Observation" and in the sub-category "Observation and Map Drawing." Map drawing is not done in chemistry and therefore has not been represented in the textbooks. Therefore, this does not signal a gap in the appropriateness of the occurrences of CoEs, but rather an unimportant sub-category when the field of study is chemistry.

Questions 4 – Additions or Deletions to Gott's classification system

The fourth question was closely related to the third and could be posed as: "Does Gott's classification system of CoEs require any additions or deletions?" Though Gott's compendium of CoEs is very extensive and thorough, there are some omissions. Under the first category of CoEs, "Fundamental Ideas," there is a subcategory termed "Types of measurement." This sub-category includes 'Interval Data,' 'Ordinal Data,' and 'Categoric Data' but it does not mention 'Ratio Data.' I think this is an omission because students need to be familiar with this type of measurement especially when dealing with pH in the chapters on "Acids and Bases." I think it would be necessary, useful and appropriate to include a fourth type of data termed 'Ratio Data' under "Types of measurement."

In the CoE 11 which is "Design of investigations: variable structure," various variables are listed such as "Independent variable" and "Dependent variable". "Controlled variables," however, are not enlisted as a sub-category, although these variables occur in most experiments and investigations. Students need to be aware of, and be able to pick out the controlled variables because these are an essential part of most experimental or investigational designs. There is a sub-category 'Control Variables in the Lab,' but I think a mention of them under "Variable Structure" is essential.

Besides, under "Design: Validity, 'fair tests' and controls" which is CoE 12, there is no provision for "Controls" in an experiment/investigation. If, for example, one is testing for sulphate in a given sample, one has to add barium chloride solution to the sample. How can we be sure that it is not some other substance present in the barium chloride solution that was causing the formation of the white precipitate. To address this question appropriately, one has to run a control with all the other conditions the same, except that the barium chloride solution is substituted with distilled water. This would be one "Control" experiment run at the same time and under exactly the same conditions as the main experiment. There is a sub-category "Control Group Experiments" under Gott's classification, but suppose there is only one control experiment in the investigation: this is not provided for with a sub-category of its own, or a special mention under "Control Group Experiments."

One of the possible CoEs not mentioned in Gott's compendium is *"Comparisons."* Sometimes data collected are made evidence by comparing with a standard, a standardized colour range or a standard trace. Such comparisons can have many forms, some of which are:

a) Melting Point standards for determining the melting point of an unknown
b) Universal Indicator Paper for pH determinations
c) Infra Red traces for identifying unknowns

In determining the melting point of an unknown, the observed melting point is compared against the melting points of some known standards, and the unknown is thereby identified. The degree of variance of the observed melting point from that of the standard is also a measure of the unknown's purity.

Universal Indicator Paper is often employed in determining the pH of a solution. One has to dip this paper in the solution whose pH is to be determined, and observe the colour the solution imparts to the paper. The colour obtained, or the datum, is then compared against a standard colour range of the Universal Indicator Paper with corresponding pHs. The result of this comparison provides the evidence that the solution has a particular pH, determined by matching the colour of the paper obtained with a particular colour from the standard colour range of the Universal Indicator Paper that is provided. The pH of the matching colour is chosen as the pH of the solution under investigation.

In Organic Chemistry infra red traces of unknowns are often used to identify them. After the initial analysis of the peaks obtained, the IR trace of the unknown is compared against the traces of standard known compounds, and the initial identification of the unknown is thereby confirmed and established. The IR trace obtained of the unknown is the data, and the match obtained with a standard is the evidence that the unknown is the same compound as the standard. When the match is obtained, therefore, the data become evidence that the composition of the unknown is the same as the standard.

Question 5 – Development in the Sophistication of CoEs across Grades and University

The fifth question that was posed was: "Was there a development in the sophistication of the CoEs as one moved from Grade 11 to Grade 12 to university?" Yes, there was an increase in the complexity and sophistication of the CoEs as one moved from Grade 11 to Grade 12 through to university. For example, in the topic Acids and Bases in Grade 11 there were Situations with just one or two of the simpler CoEs such as "Fundamental Ideas" or "Observations," but in the topic Acids and Bases in Grade 12 Situations such as the titrations involve all 21 categories of CoEs, and include all the more complex CoE categories as well. At the University level the topic Acids and Bases included more complex titrations such as those of polyprotic acids which have more than one dissociation constant. More advanced skills are therefore required for carrying out these titrations successfully.

Question 6 – Making the teaching of CoEs more effective

The sixth question posed had two parts: a) "Can CoEs be taught or learnt?" and if so, b) "How would you improve/modify the teaching of these CoEs to make it more effective?" CoEs can most definitely be taught and attempts must be made at the high school and undergraduate level to teach them systematically. If the teaching of procedural knowledge, and of CoEs in particular, is to become effective, the content that needs to be taught must be clearly laid out in the form of a curriculum and teachers must have access to this. This is very essential because procedural knowledge has a knowledge base akin to substantive knowledge, and that has to be clearly specified in an unambiguous manner preferably in a hierarchical order as suggested under the "Grouping of the 21 Categories of CoEs from Gott." For example, under the group "Design," the basic concepts such as variable structure including dependent/responding variables and independent/manipulated variables have to be taught specifically first with examples. This could be incorporated into the beginning of a Lab or Investigation assigned to the students, to be done individually or in small groups. Even a Demonstration carried out by the teacher could act as a point of entry for teaching these CoEs. Students could be given training to recognize and manipulate the variables that occur. Eventually students could be trained to recognize these CoEs in whatever branch of science they occur. This could then be

Figure 5-1. **Grouping of the 21 categories of CoEs from Gott**

Categories of CoEs	*Groups*
11. Design of investigations: Variable structure 12. Design: Validity, "fair tests" and controls 13. Design: Choosing values 14. Design: Accuracy and precision 15. Design: Tables 16. Reliability and validity of the design	Design
3. Instruments: underlying relationships 4. Instruments: calibration and error 6. The choice of an instrument for measuring a datum	Instruments
1. Fundamental ideas 2. Observation	Basics
5. Measurement 7. Reliability and validity of a single measurement	Measurement
8. Sampling a datum 9. Statistical treatment of measurements of a single datum 10. Reliability and validity of a datum 17. Data presentation 18. Statistics for analysis of data 19. Patterns and relationships in data 20. Reliability and validity of the data in the whole investigation	Datum/Data
21. Relevant societal aspects	Society

followed by the teaching of a more difficult CoE under "Design" such as "fair tests" and controls. Such a teaching of the fundamentals of the CoEs, with periodical revisitation either in theory or in practicals, would have a strengthening and a "snow-balling effect," I think, in laying the foundations of procedural knowledge.

Question 7 – The hierarchical arrangement of CoEs

"Can the CoEs be arranged hierarchically?" was the seventh question that was asked. Yes, the 21 categories can be classified into *six* groups and the members of the groups can be arranged hierarchically as indicated in Figure 5-1. When one studies the members of each group, they demonstrate an increase in their complexity spirally. This sequential order of the six groups can be followed even when attempting to find the answer to a research question.

This arrangement of the 21 categories of CoEs into six groups may be compared to Gott's own grouping into three groups. The three groups were: 1) "Making a single measurement"; 2) "Measuring a datum"; and 3) "Data in investigations – looking for relationships." The third group includes a vast array of tasks such as: the design of practical investigations; data presentation, patterns and relationships in practical investigations; reliability and validity of data in whole investigation; data to evidence; and societal issues. This final group has a collection of tasks that do not really belong together sequentially, and are just jumbled together haphazardly without any logical connection. The design of practical investigations is required at the beginning of a laboratory investigation whereas data presentation and pattern assignment are tasks employed at the end. The classification of the 21 categories of CoEs into six groups as I have suggested is much more logical, systematic and sequential.

On examining the CoEs within each of the six groups, there is a hierarchical arrangement according to their order within the group. For example, within the group of "Design":

a) The initial task is to decide upon the variables. One has to decide which the independent variables are, and which the dependent or responding variables are. The responding variable is the one that you measure.

b) For the design to be valid, there has to be a "fair test;" in other words, the independent variables have to be varied one at a time. When each independent variable is altered, the corresponding value of the responding variable is noted.

c) To ensure that one is measuring what one intends to measure and not artifacts, controls have to be put in place as well. The number and complexity of the controls depend on the research question that is at hand.

d) Then one has to decide which values are to be measured from which range to which range.

e) One has to then make a decision about the accuracy (i.e. how close the measured value is to the "true" value), and the precision (i.e. how close the multiple measured values of one factor are to each other).

f) Once decisions have been made about the accuracy and the precision, one has to design suitable tables that can comfortably accommodate, and clearly lay out, the values being measured.

g) Further points to consider would be this design's reliability and validity. Reliability refers to the reproducibility of the experiment or investigation. In other words, can the same experiment be done in another lab, by different investigators, and yield the same result? Validity queries whether the investigation is really measuring what it claims to measure.

A similar hierarchical breakdown can be made within each of the six groups.

Question 8 – Curriculum objectives regarding learning about Evidence

The eighth question was: "What curriculum objectives would you recommend regarding learning about evidence?" The curriculum objectives are those listed under "Skills" and "STS Connections" in the High School Curriculum. For learning about evidence, the curriculum objectives could be laid down, for example, as:

a) Designing of investigations:
   - Variable structure
   - Validity, "fair tests" and controls
   - Choosing values
   - Accuracy and precision
   - Tables
   - Reliability and validity of the design

b) Using Instruments:
   - Underlying relationships
   - Calibration and Error, including Zero Error
   - Choosing an instrument for measuring a datum

c) Conveying fundamental ideas:

* Linking two or more variables
* Association and Causation
* Types of Measurement
* Observation

d) Measuring:

* Reliability and Validity of a single measurement

e) Handling a Datum/Data:

* Sampling a datum
* Statistical treatment of measurements of a single datum
* Reliability and validity of a datum
* Data presentation
* Statistics for analysis of data
* Patterns and relationships in data
* Reliability and validity of the data in the whole investigation

f) Relating to the role of evidence in society:

* Credibility of evidence
* Practicality of Consequences
* Experimenter Bias
* Power Structures
* Paradigms of Practice
* Acceptability of Consequences
* Status of Experimenters

* Validity of Conclusions

Further Questions

This study raises many more interesting questions, some of which need further research and investigation. One of the pressing questions it raises is: what are the methods by which the understanding of CoEs can be tested? We can never say with complete certainty that CoEs can be taught or learnt until we carry out a pre-teaching and a post-teaching testing. If we can show that there is a marked improvement in the performance of the students after the teaching as compared to before the teaching, then, and only then, is there warrant for belief that training/teaching can improve the way students handle evidence.

There are various methods of testing available. Which methods of testing would be most appropriate for testing the understanding of the CoEs? Would the use of "paper and pencil" testing be sufficient? Should this type of testing have to be followed up with "interview testing"? The differences between procedural and substantive knowledge call for differences in the testing methods employed as well. Should procedural knowledge be tested by hands-on performance in a laboratory environment when assigned tasks such as setting up a distillation apparatus or carrying out a titration successfully? The above questions can only be addressed adequately by more research and testing. There is an inherent advantage if research studies adopt both paper and pencil testing as well as interviewing in their methodology, because they can further probe the question: Do high school and university undergraduate students find it easier to express their procedural understanding in science verbally than in writing?

Another very interesting question to investigate would be: Are there any differences between the understanding of CoEs by men and women? It would also be interesting to find out whether the socio-economic status of individuals has any effect on their understanding of procedural knowledge. For such a study one would have to carry out the sampling carefully so that students at both ends of the socio-economic spectrum are selected. In a Canadian high school or university with a rich multicultural setting, one could also explore whether ethnic diversity affects the understanding of procedural knowledge. For example, could the understanding of procedural knowledge be a function of the value that individuals place on scientific literacy or attaining a firm foundation in science? This could be similar to the surmising done by Phillips and Norris (2002) that "those who value literacy because they see it as leading to a better life might acquire more literacy."

Curriculum Policy Implications

The school curriculum can be thought of as a type of policy – a plan, a rule or a guide concerning *what* shall be taught, *where*, *when* and *how*. It has, therefore, both a *rational* and a *political* character. "It is rational in that its *content* should be rationally defensible. It is political in that it should represent a commitment on the part of certain individuals to *act* in a particular way" (Orpwood, 1985, p. 479).

Curriculum policy could be established in a number of different ways ranging from the two extremes of the "top-down" central control by government bureaucrats, to the "grass-roots" populist control by stakeholders (Hart and Robottom, 1990). Many curriculum policies, however, develop by way of collaboration and they attain some balance between top-down and grass-roots extremes. The balance is often governed by the cultural context: some are top-down oriented as in Japan (Ogawa, 1997); while others are more grass-roots oriented (e.g., in North America).

Science curriculum policy is normally formulated more smoothly through consultation with different stakeholders (Orpwood, 1985), such as government officials, the scientific community, science teachers, university science educators, parents, as well as other representing groups and institutions in society. Some of these stakeholders are politically accountable for the choices that are made in science education. These are called the *"internal stakeholders"* comprising ministers of education and school trustees, who have political responsibility and accountability, and their officials, administrators, and teachers, who are accountable by virtue of their employment in ministries or school systems. All others are *"external stakeholders"* and these include university professors, persons in business and industry, parents, members of the public, and the Science Council of Canada. The external stakeholders also have stakes in the curriculum, but they are not politically accountable for it.

During curriculum deliberations in a democratic society, it is essential that the full range of stakeholders participate in seeking a consensus over new directions. The most elaborate, consultative methodology is *deliberative inquiry*, which combines top-down and grass-roots approaches. The term "deliberative inquiry" can be found in the work of Dewey (1938, p. 181). It was from Dewey's philosophy of inquiry and Schwab's (1978) curriculum writings that the strategy was developed. The term "deliberative inquiry" suggests a tension between the practical deliberations of policymakers and the theoretical inquiries of researchers. The deliberative inquiry strategy is represented as two wheels, shown in Figure 5-2, which can turn independently or together. Of the two wheels, one represents inquiry and the other deliberation. The art of successfully conducting deliberative inquiry is that of ensuring that the wheels turn together, mutually supporting one another. Of the two processes, deliberation has got a more vital role in determining new directions for action, while the role of inquiry is more supportive and informing.

Deliberative inquiry generates three essential products which are:

A) A continuing commitment to deliberation and possible change on the part of all stake-holders.

B) A reliable database about the context in which any proposed change must take place.

C) A range of issues and alternative courses of action to form the substance of deliberations.

These three products are generated by the cycle of deliberative inquiry which has six strategic features or stages which are:

1. Issue identification – Real curriculum problems originate in schools not in the research literature. Different stakeholders will identify different frictions or see evidence of different inadequacies. Before clear research questions are formulated, some process of negotiation or preliminary deliberation must take place in which both internal and external stakeholders come to recognize the legitimacy of each other's points of view.
2. Development of research questions – In the early stages of deliberative inquiry, most of the issues identified are based on the perceptions of critics, hunches about problems and other inadequately documented allegations. These issues have to be analyzed and

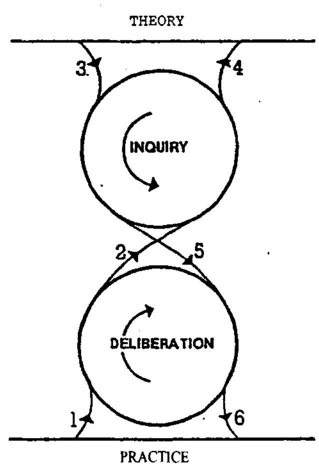
interpreted in order to yield empirical questions of a type that research can be mobilized to answer.

3. Uses of theory – To develop a research program, an array of methodological alternatives must be considered from the collection of possible research techniques.
4. Payoff for theory – The principal outcomes of deliberative inquiry are in terms of changes in policy and practice.
5. Communication for deliberation – In order to get research reports to function adequately they should raise questions as often as answer them. They should also be focussed on the specific context under consideration.
6. Payoff for practice – Finally it is the deliberators who have to decide what to do. They must " . . . choose, not the *right* alternative, for there is no such thing, but the *best* one" (Schwab, 1978, p. 318). At this stage all the stakeholders must be involved – those internal as well as those external to the system – so that they develop a "sense of proprietorship in the others' problems" (Schwab, 1978, p. 295).

Interspersing these six stages as shown in Figure 5-2, the processes of deliberation and inquiry can operate in harmony for the maximum amount of benefit to the field of practice.

To sum up, deliberative inquiry is a structured and informed conversation among stakeholders, who by their face to face encounter with government officials help them reach a decision on curriculum policy. The stakeholders achieve this by discussing and reexamining their own values along with their reading of relevant research (Orpwood, 1985). Curriculum evaluation research consistently shows that the teacher has more influence on student outcomes than does the government's choice of curriculum (Welch, 1995). The science teacher is, therefore, a key stakeholder and usually holds a central role during deliberative inquiry meetings.

Figure 5-2. **The deliberative inquiry model**



The deliberative inquiry model

(Orpwood, 1985)

This deliberative inquiry model could also be used in reaching a consensus about what, where, when and how to include the teaching of CoEs in the chemistry curriculum at the Chemistry-20, Chemistry-30 and undergraduate General Chemistry levels. Teachers who have been informed of this study could act as the key stakeholders in these deliberative inquiry meetings. In this way a strong curriculum policy about the teaching of CoEs, applicable by teachers in the classrooms at these levels, could be developed.

I had a few meetings with teachers belonging to the Edmonton Public School Board, and I shared the findings of this research study with them. The teachers expressed a great interest in identifying the Situations and the CoEs in them, and were of the opinion that this information should be made available to them at the beginning of the year before they plan their experiments and classes. They felt that if they are short of time, and cannot cover all laboratory experiments, they can choose ones with the most CoEs and thereby try to cover all the 21 CoEs. If they are under a time-constraint, and if two labs cover the same CoEs, they can teach one lab and leave out the second.

The teachers also expressed a great need for technical questions to test their students' understanding of the CoEs. They felt that a question bank would be ideal. They also thought that it would be appropriate to use evidence to solve real life problems, and this would be a worthwhile exercise for students in dealing with real-life situations. Another area where this study could be put to practical use, according to the teachers, was when schools choose textbooks. If this study was available to them, they could use its findings when trying to choose a textbook that handles evidence, and CoEs specifically, in the best possible manner.

We have, in Canada, a curriculum that is well laid out for the substantive content knowledge in science. However, for the procedural content knowledge we do not have a clearly stipulated curriculum. The embryonic curriculum that exists for procedural knowledge is lacking in systematization and comprehensiveness. This should be noted by curriculum developers and they should take steps to address this shortcoming.

High school and university science teachers need to learn the language for seeing and recognizing concepts of evidence. They may need to be taught to recognize CoEs at first. They also need to learn the language for teaching CoEs. Workshops could be arranged for teachers to gain the literacy required for seeing CoEs, as well as for teaching them.

This study has extensively explored in depth the representation of evidence in university undergraduate and high school chemistry textbooks. This work has the potential to be of use to chemistry teachers as they strive to make their teaching of this core science subject more evidence-based. As the textbooks were studied, many instances of the representation of Gott's CoEs were found. The treatment of these CoEs, however, was neither systematic nor fully comprehensive. In order to promote a systematization of procedural knowledge, the first step has to be to lay it down explicitly in the curriculum as objectives just as the substantive content is laid down. There must then be a dedication on the part of the teachers to these curriculum objectives of procedural knowledge similar to the devotion they show to the curriculum objectives of the substantive content.

References

Achinstein, P. (2001). *The book of evidence.* Oxford: Oxford University Press.

Achinstein, P. (1983). Concepts of evidence. In P. Achinstein (Ed.), *The concept of evidence.* Oxford: The Oxford University Press.

Aikenhead, G.S. (2004). Science-based occupations and the science curriculum: Concepts of evidence. *Science Education, 89,* 242-275.

Allie, S., Buffler, A., Kaunda, L., Campbell, B, & Lubben, F. (1998). First-year physics Students.' Perceptions of the quality of experimental measurements. *International Journal of Science Education, 20,* 447-459.

American Association for the Advancement of Science. (1989). *Project 2061 : Science for all Americans.* Washington, DC: Author.

Aufshnaiter, v. S., et al. (1989). Play orientation in physics education. *Science Education, 73*(4), 467-479.

Bonell, C. (1999). Evidence-based nursing: A stereotyped view of quantitative and Experimental Research could work against professional autonomy and authority. *Journal of Advanced Nursing, 30,* 18-23.

Bordt, M., De Broucker, P., Read, C., Harris, S., and Zhang, Y. (2001). Determinants of science and technology skills: Overview of the study. *Education Quarterly Review, 8*(1), 8-12.

Brickhouse, N.W., Dagher, Z.R., Shipman, H.L. & Letts, W.J. (2002). Evidence and warrants for belief in a college astronomy course. *Science and Education, 11,* 573-588.

Bucat, R.B. & Cole, A.R.H. (1988). The Australian Academy of Science School Chemistry *Journal of Chemical Education, 65,* 777-779.

Buchwald, J. Z. (1994). *The creation of scientific effects: Heinrich Hertz and electric waves.* Chicago: The University of Chicago Press.

Bybee, R. (1997). *Achieving scientific literacy: from purposes to practical action.* Portsmouth: Heinemann.

Chang, R. (2005). *Chemistry.* New York: McGraw-Hill.

Chiappetta, E.L., Sethna, G.H., & Fillman, D.A. (1991). A quantitative analysis of high School chemistry textbooks for scientific literacy themes and expository learning aids. *Journal of Research Science Teaching, 28*(10), 939-951.

Chiappetta, E.L., Sethna, G.H., & Fillman, D.A. (1993). Do middle school life science textbooks provide a balance of scientific literacy themes? *Journal of Research in Science Teaching, 30*(7), 787-797.

Closs, J.S., & Cheater, F.M. (1999). Evidence for nursing practice: A clarification of the issues. *Journal of Advanced Nursing, 30,* 10-17.

Collette, A.T. & Chiappetta, E.L. (1986). *Science instruction in the middle and secondary schools.* Columbus, OH: Charles Merrill.

Council of Ministers of Education Canada. (1997). *Common Framework of Science Learning Outcomes K-12.* URL: http://www.cmec.ca/science/framework.

Dagher, Z.R. & BouJaoude, S. (1997). Scientific views and religious beliefs of college students: The case of biological evolution. *Journal of Research in Science Teaching, 34,* 429-445.

DeBoer, G. (1991). *A history of ideas in science education: Implications for practice.* New York: Teachers College Press.

DeBoer, G. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*(6), 582-601.

Dewey, J. (1938). *Logic: The theory of inquiry,* New York: Holt, Rinehart & Winston.

Dewey, J. (1966/1916). *Democracy and education.* New York: The Free Press.

DiGisi, L.L., & Willett, J.B. (1995). What high school biology teachers say about their textbook use: A descriptive study. *Journal of Research in Science Teaching, 32*(2), 123-142.

Donnelly, J. F. (1987). Fifteen-year-old pupils' variable handling performance in the context of scientific investigations. *Research in Science and Technological Education, 5(*2), 135-147.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science.* Buckingham: Open University Press.

Duggan, S., & Gott, R. (2002). What sort of science education do we really need? *International Journal of Science Education, 24,* 661-679.

Durant, J. R. (1993). What is scientific literacy? In Durant, J. R. & Gregory, J. (eds.), *Science and Culture in Europe, (*p. 129-137). London: Science Museum.

EKOS Research Associates Inc. (2004). *Rethinking Science and Society* (pp. 63). Ottawa: Author.

Fensham, P.J. (1983). A research base for new objectives of science teaching. *Science Education, 67,* 3-12.

Garcia, T.D. (1985). *An analysis of earth science textbooks for presentation of aspects of Scientific literacy.* Unpublished dissertation, University of Houston.

Gee, B. and Clackson, S. G. (1992). The origin of practical work in the English school Science Curriculum. *School Science Review, 73(265),* 79-83.

Glynn, S.M., & Muth, D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching, 31*(9), 1057-1073.

Gott, R. & Duggan, S. (1995). *Investigative work in the science curriculum.* Buckingham, UK: Open University Press.

Gott, R. & Duggan, S. (1996). Practical work: its role in the understanding of evidence in science. *International Journal of Science Education, 18*(7), 791-806.

Gott, R., Duggan, S., and Johnson, P. (1999). What do practicing applied scientists do and what are the implications for science education? *Journal of Research in Science and Technology Education, 17,* 97-107.

Gott, R. & Johnson, P. (1999). Science in schools: time to pause for thought? *School Science Review, 81(295),* 21-28.

Gott, R., Duggan, S., & Roberts, R. (2002). Concepts of Evidence: the thinking behind the doing (GCSE version). Retrieved 08/24/2005, from www.dur.ac.uk/richard.gott/Evidence/cofev.htm

Gott, R., Duggan, S., Roberts, R., & Hussain, A. (2003). Research into understanding Scientific Evidence. Retrieved 04/07/2005 http://www.dur.ac.uk/richard.gott/Evidence/cofev.htm

Gott, R. (2004). A written test for procedural understanding: a way forward for assessment in the UK science curriculum? *Research in Science and Technological Education, 22*(1), 5-21.

Hart, E. P. & Robottom, I. M. (1990). The science-technology-movement in science education: A Critique of the reform process. *Journal of Research in Science Teaching, 27,* 575-588.

Hertz, H. (1896). *Miscellaneous Papers.* London: Macmillan.

Hinman, R.L. (1998). Who is scientifically literate anyway? *Phi Delta Kappan.* 79(7), 540-544.

Holliday, W.G., Yore, L.D., & Alvermann, D.E. (1994). The reading-science learning-Writing connection: Breakthroughs, barriers, and promises. *Journal of Research in Science Teaching, 31*(9), 877-893.

Hurd, P. (1958). Science literacy: Its meaning for American schools. *Educational Leadership, 16*, 13-16, 52.

Husen, T., Tuijnman, A. & Halls, W.D. (Eds.). (1992). *Schooling in modern European society. A Report of the Academia Europeae.* London: Pergamon Press.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* Basic Books, Inc., Publishers.

Jenkins, E. (1997). Towards a functional public understanding of science. In R. Levinson & J. Thomas (Eds.), *Science today: Problem or crisis?* (pp. 137-150). London: Routledge.

Jenkins, E. W. (1994). Scientific literacy. In Husen, T. & Postlethwaite, T. N. (Eds.), *The International Encyclopedia of Education,* 2nd Edn., Volume 9 (pp. 5345-5350). London: Pergamon.

Jenkins, F., Van Kessel, H., & Tompkins, D. (1993). *Nelson Chemistry.* Ontario, Canada: Nelson Canada.

Kanari, Z. & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748-769.

Kirkham, R.L. (1997). *Theories of truth: A critical introduction.* London: Blackwell.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago, IL: University Press.

Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills.* London: Academic Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

Layton, D. (1975). *Science for the People.* London: George Allen & Unwin.

Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education, 18,* 955-968.

Masnick, A.M. & Morris, B.J. (2002). Reasoning from data: The effect of sample size and variability on children's and adults' conclusions. In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the 24th annual conference of the Cognitive Science Society* (p. 643-648). Mahwah, NJ: Lawrence Erlbaum.

Mautner, T. (Ed.). (1999). *The Penguin dictionary of philosophy.* London: Penguin Books.

Mayer, R.E. (1983). What we have learned about increasing the meaningfulness of science prose? *Science Education, 67*(2), 223-237.

Millar, R. (1996). Towards a science curriculum for public understanding. *School Science Review*, 77(280), 7-18.

Millar, R. (2004). *The role of practical work in the teaching and learning of science.* Paper prepared for the Committee: High School Science Laboraories: Role and Vision, National Academy of Sciences,Washington, DC.

Millar, R. (2006). Twenty First Century Science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, *28*(13), 1499.

Millar, R. & Driver, R. (1987). Beyond processes. *Studies in Science Education, 14,* 33-62.

Millar, R. & Osborne, J.F. (Eds.). (1998). *Beyond 2000: Science education for the future.* London: King's College London.

Miller, S. & Fredericks, M. (2003). The nature of "Evidence" in qualitative research methods. International Journal of Qualitative Methods, 2(1). Article 4. Retrieved 27/12/2006 from http://www.ualberta.ca/~iiqm/backissues/2_1/html/miller.html

Mustoe, F., Jansen, M.P., Doram, T., Ivanco, J., Clancy, C., & Ghazariansteja, A. (2001). *Chemistry 11.* Toronto, Canada: McGraw-Hill Ryerson.

Mustoe, F., Jansen, M.P., Doram, T., Ivanco, J., Clancy, C., & Ghazariansteja, A. (2001). *Chemistry 12.* Toronto, Canada: McGraw-Hill

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Science Teachers Association (1982). *Science, technology, society – Science education for the 1980s: An NSTA position statement.* Washington, DC: Author.

National Society for the Study of Education. (1947). Science education in American schools: Forty-sixth yearbook of the NSSE. Chicago: University of Chicago Press.

National Society for the Study of Education. (1960). Rethinking science education: Fifty-ninth Yearbook of the NSSE. Chicago: University of Chicago Press.

Neuendorf, A.K. (2002). *The content analysis guidebook.* Thousand Oaks: Sage Publications.

Norris, S. (1983). *Perceptions of the reorganized high school programme for Newfoundland and Labrador schools.* St. John's Newfoundland: Institute for Educational Research and Development, Memorial University of Newfoundland, 1983.

Norris, S. & Phillips, L. (1994a). The relevance of a reader's knowledge within a perspectival view of reading, *Journal of Reading Behavior, 26,* 391-412.

Norris, S. & Phillips, L. (1994b). Interpreting pragmatic meaning when reading popular reports of Science. *Journal of Research in Science Teaching, 31,* 947-967.

Norris, S., & Phillips, L. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224-240.

Ogawa, M. (1997). Toward an epic description of science education: Cultural history of Science education in Japan. In M. Ogawa (Ed.), *Effects of traditional cosmology on science education* (pp. 96-125). Ibaraki, Japan: Ibaraki University, Faculty of Education.

Orpwood, G. (1985). Toward the renewal of Canadian science education. I. Deliberative Inquiry model. *Science Education, 69,* 477-489.

Orpwood, G.W.F., & Alam, I. (1984). *Background Study 52, Science education in Canadian schools* Ottawa, ON: Science Council of Canada.

Orpwood, G.W.F., & Soque, J.P. (1984). *Summary of Background Study 52, Science Education in Canadian Schools.* Ottawa, Ontario, Canada: Science Council of Canada.

Osborne, J., Driver, R. & Simon, S. (1998). Attitudes to science: issues and concerns. *School Science Review, 79*(288), 27-33.

Pella, M.O., O'Hearn, G.T. & Gale, C.W. (1966). Scientific literacy – Its referents. *The Science Teacher, 33*(5), 44.

Petrucci, R.H., Harwood, W.S., & Herring, F.G. (1997). *General Chemistry: Principles* *and Modern Applications.* New Jersey: Prentice Hall.

Phillips, L. M. & Norris, S. P. (1999). Interpreting popular reports of science: what happens when the readers's world meets the world on paper? *International Journal of Science Education, 21*(3), 317-327.

Phillips, L.M. & Norris, S.P. (2002). Literacy policy and the value of literacy for individuals. In Y.M. Hebert (Ed.), *Citizenship in transformation in Canada* (pp.209-227). Toronto: University of Toronto Press.

President's Scientific Research Board. (1947). *Science and public policy (vols. 1 and 4).* Washington: Government Printing Office.

Rakow, S.J. (1985). Excellence in middle/junior high science teaching: The teacher's Perspective. *School Science and Mathematics, 85*(8), 631-632.

Roberts, D.A. (1983). *Scientific literacy towards balance in setting goals for schools science programs* (Cat. No. SS21-5/1983-2E). Toronto, ON: The Publication Office, Science Council of Canada.

Roberts, R. (2001). Procedural understanding in biology: the 'thinking behind the doing.' *Journal of Biological Education, 35*(3), 113-117.

Roberts, R. & Gott, R. (1999). Procedural understanding: its place in the biology curriculum. *School Science Review, 81*(294), 19-25.

Roberts, R. & Gott, R. (2000). Procedural understanding in biology: how is it characterized in texts? *School Science Review, 82*(298), 83-91.

Rockefeller Brothers Fund (1958). The pursuit of excellence: Education and the future of America. In *Prospect for America: Report Number V of the Rockefeller Panel Reports.* Garden City, NY: Doubleday.

Ruis, S.P. (1988). Something's wrong with chemistry textbooks. *Journal of Chemical Education, 65,* 720-721.

Rutherford, J., & Ahlgren, A. (1990). *Science for all Americans.* New York: Oxford University Press.

Ryder, J., Leach, J. & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching, 6*(2), 201-220.

Samarapungavan, A. (1992). Children's judgments in theory choice tasks: Scientific rationality in Childhood. *Cognition, 45*(1), 1-32.

Schwab, J. J. (1978). The practical: A language for curriculum. In *Science, curriculum,* *and liberal Education,* . Westbury, I. and Wilkhof, N.J. (Eds.), Chicago: University of Chicago Press.

Scruggs, M.M. (1988). What research says about textbooks. *Science and Children, 25*(4), 24-25.

Sere, M-G., Journeaux, R. and Larcher, C. (1993). Learning the statistical analysis of Measurement Error. *International Journal of Science Education, 15*(4), 427-438.

Shamos, M. (1995). *The Myth of Scientific Literacy.* New Brunswick, NJ: Rutgers University Press.

Showalter, V.M. (1974). What is united science education? (Part 5): Program objectives and scientific Literacy. *Prism, II*(2), 3-4.

Solomon, J. (1993). *Teaching Science, Technology and Society.* Buckingham: Open University Press.

Stake, R.E., & Easley, J.A. (1978). *Case Studies in Science Education.* University of Illinois Center for Instructional Research and Curriculum Evaluation, Urbana.

Statistics Canada. (2000). *Education in Canada, 2000.* (Catalogue no. 81-229, pp. 62-64). Ottawa: Author.

Statistics Canada. (2003). *Measuring up: Canadian Results of the OECD PISA Study* (pp.36-41). Ottawa: Author.

Stinner, A. (1992). Science textbooks and science teaching: from logic to evidence. *Science Education, 76*(1), 1-16.

Sutman, F.X. (1996). Science literacy: A functional definition. *Journal of Research in Science Teaching, 33*(5), 459-460.

Taylor, J.A., & Dana, T.M. (2003). Secondary school physics teachers' conceptions of Scientific evidence: An exploratory case study. *Journal of Research in Science Teaching, 40,* 721-736.

Thomas, G. and Durant, J. (1987). Why should we promote the public understanding of science? *Scientific Literacy Papers*, Summer, 1-14. Oxford: Department of External Studies, University of Oxford.

Thomson, J. J. (1897). Cathode rays. *Philosophical Magazine, 44,* 300-303.

Welch, W. W. (1995). Student assessment and curriculum evaluation. In Fraser, B. J. & Walberg, H. J. (Eds.), *Improving science education* (pp. 90-116). Chicago: The National Society for the Study of Education (University of Chicago Press).

Whewell, W. (1850). *Of a Liberal Education.* London: Parker, Rare Books.

Wright, J.C., & Wright, C.S. (1998). A commentary on the profound changes envisioned by the national Science standards. *Teachers College Record, 100*(1), 122-143.

Yager, R.E. (1983). The importance of terminology in teaching K-12 science. *Journal of Research in Science Teaching, 20,* 577-588.

Appendix A

Concepts of Evidence: the Thinking Behind the Doing (GCSE version)

© R Gott, S Duggan, R Roberts

We are indebted to Glen Aikenhead of the University of Toronto for his detailed comments on this version and for many of the examples we use to illustrate the ideas.
December 2002.
Please quote as:
Gott, R, Duggan, S and Roberts, R. University of Durham, http://www.dur.ac.uk/richard.gott/Evidence/cofev.htm

The section numbers are not sequential. They are references to a more extensive list. Only those ideas which have been validated as being appropriate to GCSE are included here.

1 Fundamental ideas

0 ***Introduction*** Investigations must be approached with a critical eye. What sort of link is to be established, with what level of measurement and how will opinion and data be weighed as evidence?

This first category pervades the entire scheme and sets the context in which all that follows needs to be judged.

1 ***Opinion and data*** ...it is necessary to distinguish between opinion based on scientific evidence and ideas on the one hand, and opinion based on non-scientific ideas (prejudice, whim, hearsay . .) on the other.

Distinguishing between the measurable energy emitted from a mobile phone mast and the 'energy' associated with 'crystals'.

2 ***Links*** ...a scientific investigation seeks to establish links (and the form of those links) between two or more variables

3 ***Association and causation***...links can be causal (change in the value of one variable CAUSES a change in another), or associative (changes in one variable and changes in another are linked to some third, and possibly unrecognised, third (or more) variable)

4 ***Types of measurement*** ...interval data (measurements of a continuous variable) are more powerful than ordinal data (rank ordering) which are more powerful than categoric data (a label)

Being able to say that 2 wavelengths are 670nm. and 460nm. is more useful than saying one is longer than the other or that one is red and the other blue.

5 ***Extended tasks*** ...measurements, for instance, can be very complicated and constitute a task on their own, but they are only meaningful when set within the wider investigation(s) of which they will form a part

The measurement of 'the absorbancy of a papertowel' involves a 'method' which, all together, contributes to the final measure of absorbancy.

2 Observation

0 ***Introduction*** Observation of object and events can lead to informed description and the generation of questions to investigate further.

Observation is one of the key links between the 'real world' and the abstract ideas of science. Observation, in our definition, does not include 'measurement' but rather deals with the way we see objects and events through the prism of our understanding.

1 ***Observing objects*** ... objects can be 'seen' differently depending on the conceptual window used to view them.

...a low profile car tyre can be seen as nothing more than that, or it can be seen as a way of increasing the stiffness of the tyre, thus giving more centripetal force with less deformation and thus improving road holding.

2 ***Observing events*** ... events can similarly be seen through different conceptual windows.

...the motion of a parachute is seen differently when looked at through a framework of equal and unequal forces and their corresponding accelerations.

3 ***Using a key*** ... the way in which an object can be 'seen' can be shaped by using a key *e.g. a branching key gives detailed clues as to what to 'see'. It is, then, a heavily guided concept-driven observation*

4 ***Taxonomies*** ... taxonomies are a means of using conceptually driven observations to set up classes of objects or organisms that exhibit similar/different characteristics or properties with a view to using the classification to solve a problem.

....organisms observed in a habitat may be classified according to their feeding characteristics (to track population changes over time for instance) or a selection of materials classified into efficient conductors identified from inefficient conductors

5 ***Observation and experiment***... observation can be the start of an investigation, experiment of survey.

... noticing that shrimp populations vary in a stream leads to a search for a hypothesis as to why that is the case, and an investigation to test that hypothesis.

6 ***Observation and map drawing***

... technique used in biological and geological fieldwork to map a site based on conceptually driven observations that illustrate features of scientific interest

....an ecologist may construct a map of a section of a stream illustrating areas of varying stream flow rate or composition of the stream bed.

3 Measurement

0 ***Introduction*** Measurement must take into account inherent variation due to uncontrolled variables, human error and the characteristics of the instruments used. *This section lies at the very centre of our model for measurement, data and evidence and is fundamental to it.*

1 ***Inherent variation*** ...the measured value of any variable will never repeat unless all possible variables are controlled between measurements - circumstances which are very difficult to create

Repeated bounces of a squash ball under ostensibly identical conditions will result in varied data.

2 ***Human error*** ...the measured value of any variable can be subject to human error which
can be random, or systematic

In the case of the squash ball, human error could result from a shaky hand when the ball was released (random) or the bounce height could always appear higher than it really is if the observer was below the height looking up at the bounce against the rule

4 Instruments: underlying relationships

R Gott, S Duggan, R Roberts, A Hussain Page 2 of 14

1 ***Linear relationships*** ...most instruments rely on an underlying and preferably linear relationship
between two variables.

e.g. A thermometer relies on the relationship between the volume of a liquid and temperature.

2 ***Non-linear relationships*** ...some 'instruments', of necessity, rely on non-linear relationships.

e.g. Moving iron ammeter, pH.

3 ***Complex relationships*** ...the relationship may not be straightforward and may be confounded by other factors.

e.g. The prevalence, or size, of a species of lichen is an indicator of the level of pollution but other environmental factors such as aspect, substrate, or air movement can also affect the distribution of lichen.

4 ***Multiple relationships*** ...sometimes several relationships are linked together so that the measurement of a variable is indirect.

e.g. Medical diagnosis often relies on indirect multiple relationships.
Also, braking distance is an indirect measure of frictional force.

5 Instruments: calibration and error

0 ***Introduction*** Instruments must be carefully calibrated to minimise the inevitable uncertainties in the readings

All instruments must be calibrated so that the underlying relationship is accurately mapped onto the scale. If the relationship is non-linear, the scale has to be calibrated more often to map that non-linearity. All instruments, no matter how well made, are subject to error. Each instrument has finite limits on, for example, its resolution and sensitivity.

1 ***End points*** ...the instrument must be calibrated at the end points of the scale.

e.g. A thermometer must be calibrated at 0 °C and 100 °C.

2 ***Intervening points*** ...the instrument must be calibrated at points in between to check the linearity of the underlying relationship.

e.g. A thermometer must be calibrated at a number of intervening points to check, for instance, for non-linearity due to non-uniform bore of the capillary.

3 ***Zero Errors*** ...there can be a systematic shift in scale and that instruments should be checked regularly.

e.g. If the zero has been wrongly calibrated, if the instrument itself was not zeroed before use, or if there is fatigue in the mechanical components, a systematic error can occur.

4 ***Overload, limitingsensitivity / limit of detection***
...there is a maximum (full scale deflection) and a minimum quantity which can be measured reliably with a given instrument and technique.
e.g. change in mass when Mg burns in air could not be detected on scales that measure to only whole grammes.

5 ***Sensitivity**** ...the sensitivity of an instrument is a measure of the amount of error inherent in the instrument itself.
e.g. An electronic voltmeter will give a reading that fluctuates slightly.

6 ***Resolution and error*** ...the resolution is the smallest division which can be read easily. The resolution can be expressed as a percentage.
e.g. If the instrument can measure to 1 division and the reading is 10 divisions, the error can be expressed as 10±1 or as a percentage error of 10%.

7 ***Specificity**** ...an instrument must measure only what it purports to measure.
e.g.false positives on drug tests due to detection of a similar naturally occuring substance.

8 ***Instrument use*** ...there is a prescribed procedure for using an instrument which, if not followed, will lead to systematic and / or random errors.
e.g. When measuring the temperature of a liquid, if one takes the thermometer out of the liquid to read the thermometer, this will lead to systematically low or high readings, compared to reading the thermometer immersed in the liquid. More specifically, there is a prescribed depth of immersion for some thermometers that takes account of the expansion or contraction of the glass and the mercury (or alcohol) that are not in the liquid being measured.

9 ***Human error*** ...even when an instrument is chosen and used appropriately, human error can occur.
e.g. Scales on measuring instruments can easily be misread.

6 Reliability and validity of a single measurement

0 ***Introduction*** Any measurement must be reliable and valid.
A measurement, once made, must be scrutinised to make sure that it is a valid measurement; it is measuring what was intended, and that it can be relied upon. Repeating readings and triangulation, by using more than one of the same type of instrument or by using another type of instrument, can increase reliability.

1 ***Reliability*** ...a reliable measurement requires an average of a number of repeated readings; the number needed depends on the accuracy required in the particular circumstances
e.g. the height from which a ball is dropped could be checked if it was important that the drop height was accurate.

2 ***Reliability*** ...instruments can be subject to inherent inaccuracy so that using different

instruments can increase reliability.
e.g. Measurement of blood alcohol level can be assessed with a breathalyser and cross checked with a blood test. Also, temperature can be measured with mercury, alcohol, and digital thermometers to ensure reliability.
3 ***Reliability*** ...human error in the use of an instrument can be overcome by independent, random checks.
e.g. Spot checks of measurement techniques by co-workers are sometimes built into routine procedures.
4 ***Validity*** ...measures that rely on complex or multiple relationships must ensure that they are measuring what they purport to measure.
e.g. is the colour change a measure of bacterial activity or might something else have caused it?

7 The choice of an instrument for measuring a datum

0 ***Introduction*** Measurements are never entirely accurate for a variety of reasons.
Of prime importance is choosing the instrument to give the accuracy and precision required; a proactive choice rather than a reactive discovery that it wasn't the right instrument for the job!
1 ***Trueness or accuracy**** ...trueness is a measure of the extent to which repeated readings of the same quantity give a mean that is the same as the 'true' mean.
e.g. If the mean of a series of readings of the height of an individual pupil is 173 cm and her 'true' height, as measured by a clinic's instrument is 173 cm, the measuring instrument is 'true'.

2 ***Non-repeatability*** ...repeated readings of the same quantity with the same instrument never give exactly the same answer.
e.g. Weighing yourself on a bathroom scale in different places on the bathroom floor, or standing in a slightly different position on the scales, will result in slightly differing measurements. It is never possible to repeat the measurement in exactly the same way.
3 ***Precision*** ...(sometimes called "imprecision" in industry) refers to the observed variations in repeated measurements from the same instrument. In other words, precision is an indication of the spread of the repeated measurements around the mean. A precise measurement is one in which the readings cluster closely together. The less the instrument's precision, the greater is its uncertainty. A precise measurement may not necessarily be an accurate or true measurement (and vice versa). The concept of precision is also called "reliability" in some fields. A more formal descriptor or assessment of precision might be the range of the observed readings, the standard deviation of those readings, or the standard error of the instrument itself.
e.g. For bathroom scales, a precise set of measurements might be: 175, 176, 175, 176, and 174 pounds.
4 ***Reproducibility*** ...whereas repeatability (precision) relates to the ability of the method to give the same result for repeated tests of the same sample on the same equipment (in the same laboratory), reproducibility relates to the ability of

the method to give the same result for repeated tests of the same sample on equipment in different laboratories.
e.g. 'Round Robins' are often used to check between different laboratories. A standardised sample is sent to each lab and they report their measurment(s) and degree of uncertainty. Labs are then compared.
5 ***Outliers in relationships*** ...outliers, aberrant or anomalous values in data sets should be examined to discover possible causes. If an aberrant measurement or datum can be explained by poor measurement procedures (whatever the source of error), then it can be deleted.
e.g. an anomalous bounce would be deleted if the cause of the anomaly was known, but if it could not be explained, then further bounces would be needed to see if it was part of the inherent variation.

8 Sampling a datum

0 ***Introduction*** A series of measurement of the same datum can be used to determine the reliability of the measurement
We use the term 'sampling' to mean any sub-set of a population. The population might be a species of animal or plant, or even the possible sites where gold might be found. We shall also take the population to mean the infinite number of repeated readings.
1 ***Sampling*** ...one or more measurements comprise a sample of all the possible measurements that could be made.
e.g. The measurement of a single blade of grass is a sample of all the blades of grass in a field. Also, a single measurement of the bounce height of a ball is a sample of the infinite number of such bounces that could be measured.
2 ***Size of sample*** .. the number of measurements taken. The greater the number of readings taken, the more likely they are to be representative of the population.
e.g.repeated readings on a ammeter in a particular circuit are a sample of all possible readings. The more readings taken, the more the sample represents the population of all possible readings.

3 ***Reducing bias in sample / representative sampling***
...measurements must be taken using an appropriate sampling strategy, such as random sampling, stratified or systematic sampling so that the sample is as representative as possible.
To find the height of college students, tables of random numbers can be used to select students.
4 ***An anomalous datum*** ...an unexpected datum could be indicative of inherent variation in the data or the consequence of a recognised uncontrolled variable
e.g. Continuing the above examples, a very small height may have been recorded from a child visiting the college and should not be part of the population being sampled; whereas a very low rebound height from a squash ball may occur as a result of differences in the material of the ball and is therefore part of the sample.

9 Statistical treatment of measurements of a single datum

0 ***Introduction*** A group of measurements of the same datum can be described in various mathematical ways.

The statistical treatment of a datum is concerned with the probability that a measurement is within certain limits of the true measurement. The following are some basic statistics associated with a single datum.

1 ***Range*** …the range is a simple description of the distribution and defines the maximum and minimum values measured.

e.g. Measuring the height of carbon dioxide bubbles on successive trials in a yeast experiment, the following measurements were recorded and ordered sequentially: 1.8, 1.9, 2.1, 2.1, 2.1, 2.3, 2.4, 2.4, 2.5, 2.5 and 2.6 cm. The range is 0.8 cm (2.6 - 1.8)

4 ***Mean*** ...the mean (average) is the sum of all the measurements divided by the number of measurements.

e.g. Continuing the example above, the mean is 2.2 cm

10 Reliability and validity of a datum

0 ***Introduction*** A datum must have a known (or estimated) reliability and validity before it can be used in evidence.

Any datum must be subject to careful scrutiny to ascertain the extent to which it:

- is valid: that is, has the value of the appropriate variable been measured? Has the parameter been sampled so that the datum represents the population?

- is reliable: for example, des the datum have sufficient precision? The wider the confidence limits (the greater the uncertainty), the less reliable the datum.

Only then can the datum be weighed as evidence. Evaluation a datum also includes evaluating the validity of the ideas associated with the making of a single measurement.

1 ***Reliability*** ...a datum can only be weighed as evidence once the uncertainty associated with the instrument and the measurement procedures have been ascertained.

e.g. the reliability of the volume of water absorbed by a papertowel should be judged in terms of the uncertainty of the instruments used in the method and any slops and spillages that might have occurred.

2 ***Validity*** ...that a measurement must be of, or allow a calculation of, the appropriate datum.

e.g. the time it takes for a shoe to be pulled 50 cm across a surface is an invalid measure of the force required to move the shoe.

11 Design of investigations: Variable structure

0 ***Introduction*** The design of an investigation requires variables to be identified (as Independent, dependent and controlled) and measured.

An investigation is an attempt to determine the relationship, or lack of one, between the independent and dependent variables, or between two or more sets of data. Investigations take many forms but all have the same underlying structure. By identifying and understanding the basic structure of an investigation in terms of variables and types of

variables, we can begin to evaluate the validity of data.
1 ***The independent variable***...the independent variable is the variable for which values are changed or selected by the investigator.
e.g. length of resistance wire is the independent variable changed by the investigator to see what affect it has. The aspect of a slope is the independent variable selected in the field to investigate whether aspect affects vegetation.
2 ***The dependent variable*** ...the dependent variable is the variable the value of which is measured for
each and every change in the independent variable.
e.g. Continuing the examples above, the resistance of the wire or the density of bluebells.
3 ***Correlated variables*** ...in some circumstances we are looking for a correlation only rather than
any implied causation
e.g. the number of birds in the garden and the number of cats in the area.
4 ***Categoric variables*** ...a categoric variable has values which are described by labels. Categoric variables are also known as nominal data.
e.g. "type of sugar" has data values of "icing", "castor", "granulated", etc.
5 ***Ordered variables*** ...an ordered variable has values which are also descriptions, labels or categories but these categories can be ordered or ranked. Measurement of ordered variables results in ordinal data.
e.g. The variable of "age of bird" has ordered values "fledgling", "immature" and "adult". The variable "density of barnacles" can be assigned numbers according to a scale, but the numbers are not intervals but an ordered variable.
6 ***Continuous variables*** ...a continuous variable is one which can have any numerical value and itsmeasurement results in interval data.
e.g. height, current, time.
7 ***Discrete variables*** ...a discrete variable is a special case in which the values of the variable are restricted to integer multiples.
e.g. The number of layers used to insulate a cup.

12 Design: Validity, 'fair tests' and controls
R Gott, S Duggan, R Roberts, A Hussain Page 7 of 14

0 ***Introduction*** Uncontrolled variation can be reduced through a variety of techniques.
Fair tests and controls aim to isolate the effect of the independent variable on the dependent variable. Laboratory-based investigations, at one end of the spectrum, involve the investigator changing the independent variable and keeping all the controlled variables constant. This is often called 'the fair test', but it is no more than one of several valid designs. At the other end of the spectrum are field studies where many naturally changing variables are measured and correlations sought. For example, an ecologist might measure many variables in a habitat over a period of time. Having collected the data, correlations might be sought between variables such as day

length and emergence of a butterfly, using statistical treatments to ensure validity. The possible effect of other variables can be reduced by only considering data where the values of other variables are the same or similar. In between these extremes, are many types of valid designs that involve different degrees of manipulation and control. Fundamentally, all these investigations have a similar structure; what differs are the strategies to ensure validity.

1 ***Fair test*** ...a fair test is one in which only the independent variable has been allowed to affect the dependent variable.

e.g. A laboratory experiment about the effect of temperature on dissolving time, where only the temperature is changed. Everything else is kept exactly the same.

2 ***Control variables in the laboratory***...other variables can affect the results of an investigation unless their effects are controlled by keeping them constant.

e.g. In the above experiment, the mass of the chemical, the volume of liquid, the stirring technique, and the room temperature are some of the variables that should be controlled.

3 ***Control variables in field studies***...some variables cannot be kept constant and all that can be done is to make sure that they change in the same way.

e.g. In a field study on the effect of different fertilisers on germination, the weather conditions are not held constant but each experimental plot is subjected to the same weather conditions.

4 ***Control variables in surveys***...the potential effect on validity of uncontrolled variables can be reduced by selecting data from conditions that are similar with respect to other variables.

e.g. In a field study to determine whether light intensity affects the colour of 'dog's mercury leaves (Mercurialis perrenis, a woodland plant), other variables are recorded, such as soil nutrients, pH, and water content. Correlations are then sought by selecting plants growing where the value of these variables is similar. Also, in a survey investigating if blood pressure increased with age, other variables such as diet, fitness, weight to height ratio, etc. are recorded.

5 ***Control group experiments***...control groups are used to ensure that any effects observed are due to theindependent variable(s) and not some other unidentified variable. They are no more than the default value of the independent variable.

e.g. seedlings are grown in normal soil (control group) as well as in a variety of composts with different nutrients to ensure that the batch of seeds was viable and the temperature and watering regime etc were suitable for normal growth.

13 Design: Choosing values

R Gott, S Duggan, R Roberts, A Hussain Page 8 of 14

0 ***Introduction*** Choosing the values of the variables in an investigation.

The values of the variables need to be chosen carefully. This is possible in the majority of investigations, prior to the data being collected. In field studies where data are collected from variables that change naturally, some of these concepts can only be applied *retrospectively.*

1 ***Trial run*** …a trial run can be used to establish the broad parameters required of the experiment (scale, range, number) and help in choosing instrumentation and other equipment

e.g. paper helicopters of different dimensions are dropped from different heights to determine which is likely to be suitable for the investigation.

2 ***The sample*** …issues of sample size and representativeness apply in the same way as in sampling a datum (Measurement section).

e.g. This cannot be determined in advance, or be a set number, if the inherent variation is unknown.

3 ***Relative scale*** ...the choice of sensible values for quantities is necessary if measurements of the dependent variable are to be meaningful.

e.g. dropping a large paper helicopter from 0.5m doesn't allow it to start to turn which will invalidate the results.

4 ***Range*** ...the range over which the values of the independent variable is chosen is important in ensuring that any pattern is detected.

e.g. An investigation into the effect of competition for light with wheat seeds placed between 5 and 10cm apart would show little effect of crowding.

5 ***Interval*** ...the choice of interval between values determines whether or not the pattern in the data can be identified.

e.g. An investigation into the effect of temperature on enzyme activity would not show the complete pattern if 20°C intervals were chosen.

6 ***Number*** ...a sufficient number of readings is necessary to determine the pattern.

e.g. The number is determined partly by the range and interval issues discussed above, but in some cases for the complete pattern to be seen, more readings may be necessary in one part of the range than another. This applies particularly if the pattern changes near extreme values, for example, in a spring extension experiment at the top of the range of the mass suspended on the spring.

14 Design: Accuracy and precision

0 ***Introduction*** Ensuring appropriate accuracy and precision.

The design of the investigation must provide data with sufficiently appropriate accuracy and precision to answer the research question. This consideration should be built into the design of the investigation. Different investigations will require different levels of accuracy and precision depending on their purpose.

1 ***Determining differences*** ...there is a level of precision which is sufficient to provid data which will allow discrimination between two or more means.

e.g. Which is the most absorbant 'economy paper towel'? When the differences are potentially small the degree of precision in the measurement method must be sufficient to enable differences to be attributed to the paper towel rather than uncertainty in the measurements.

2 ***Determining patterns*** ...there is a level of precision which is required for the trend in a pattern to be determined.
e.g. Large error-of-measurement bars on the points of a line graph may not allow discrimination between an upward curve or a straight line.

15 Design: Tables

1 ***Tables*** ...tables can be used as organisers for the design of an experiment by preparing the table in advance of the whole experiment. A table has a conventional format.
e.g. An experiment on the effect of temperature on the dissolving time of calcium chloride:

16 Reliability and validity of the design

0 ***Introduction*** An evaluation of an investigation must consider reliability and validity.
In evaluating the design of an investigation, there are two overarching questions:
- will the measurements result in sufficiently reliable data to answer the question?
- will the design result in sufficiently valid data to answer the question?
Evaluation the design of an investigation included evaluating the reliability and validity of the ideas associated with the making of a single measurement and with each and every datum.
1 ***Reliability of the design*** ...the reliability of the design includes a consideration of all the ideas associated with the measurement of each and every datum.
e.g. Factors associated with the choice of the measuring instruments to be used must be considered, for instance, the error associated with each measuring instrument. The sampling of each datum and theaccuracy and precision of the measurements should also be considered. This includes the sample size, the sampling technique,relative scale, the range and interval of the measurements, the number of readings, and the appropriate accuracy and precision of the measurements.
2 ***Validity of the design*** ...the validity of the design includes a consideration of the reliability (as above) and the validity of each and every datum.
e.g. This includes the choice of measuring instrument in relation to whether the instrument is actually measuring what it is supposed to measure. This includes considering the ideas associated with the variable structure and the concepts associated with the fair test. For instance, measuring the distance travelled by a car at different angles of a ramp will not answer a question about speed as a function of angle.

17 Data presentation

0 Data can be presented in a number of ways.
Having established that the design of an investigation is reliable and valid, what do we need to understand to explore the relationship between one variable and another? Another way of thinking about this is to think of the pattern between two variables or two sets of data. What do we need to understand to know that the pattern is valid and reliable? The way that data are presented allows patterns to be seen. There is a close link between graphical representations and the type of variable they represent.

1 ***Tables*** ...a table is a means of reporting and displaying data. But a table alone presents limited information about the design of an investigation e.g.control variables or measurement techniques are not always overtly described.
e.g. a table showing type of sugar and dissolving time does not indicate the mass of sugar used, the volume or temperature of the water and whether it was stirred and how vigorously.
2 ***Bar charts*** ...bar charts can be used to display data in which the independent variable is categoric and the dependent variables is continuous.
e.g. The number of pupils who can and cannot roll their tongues would be best presented on a bar chart.
3 ***Line graphs*** ...line graphs can be used to display data in which both the independent variable and the dependent variable are continuous. They allow interpolation and extrapolation.
e.g. The length of a spring versus the force applied would be best displayed in a line graph.
4 ***Scatter graphs (or scatter plots)...*** are used to display data in which both the independent variable and the dependent variable are continuous. Scatter graphs are often used wherethere is much fluctuation in the data because they can allow an association to be detected. Widely scattered points can show a weak correlation, points clustered around, for example, a line can indicate a relationship.
e.g. The dry mass of the aerial parts of a plant and the dry mass of the roots.
5 ***Histograms*** ...histograms can be used to display data in which a continuous independent variable has been grouped into ranges and in which the dependent variable is continuous.
e.g. On a seashore, the distance from the sea could be grouped into ranges, i.e. lower, middle and upper shore, and the number of limpets in each range plotted in a histogram.
8 ***Other forms of display*** ...data can be transformed, for example, to logarithmic scales so that they meet the criteria for normality which allows the use of parametric statistics.
e.g. coloured shading or symbols could be used on a map to indicate the density of plants in an area of woodland. A pictorial scaled drawing could be used to show the amount of emissions from different forms of transport.

19 Patterns and relationships in data

0 ***Introduction*** Data must be inspected for underlying patterns. *Patterns cannot be treated in isolation from the physical system that they represent, because patterns represent the behaviour of variables in that system. Patterns can be seen in tables or graphs or can be reported by using the results of appropriate statistical analysis. The interpretation of patterns and relationships must respect the limitations of the data. For instance, there is a danger of overgeneralizing or of implying causality when there may be a different, less direct type of association.*

1 ***Types of patterns*** ...there are different types of association such as causal, consequential, indirect or chance associations. "Chance association" means that observed differences in data sets, or changes in data over time, happen simply by chance alone. We must sceptically be open to possibility that a pattern has emerged by chance alone. Statistical tests give us a rational way to estimate this chance.
e.g. In any large multivariate set of data, there will be associations, some of which will be chance associations. Even if x and y are highly correlated, x does not necessarily cause y: y may cause x or z may cause x and y. (See 1.3.) Also, changes in students' understanding before and after an intervention may not be significant and/or may be due to other factors.
2 ***Linear relationships*** ...straight line relationships (positive slopes, negative, and vertical and horizontal as special cases) can be present in data in tables and line graphs and that such relationships have important predictive power ($y = mx + c$)
e.g. Height and time for a falling object.
3 ***Proportional relationships***... direct proportionality is a particular case of a straight line relationships with consequent predictive characteristics. The relationship is often expressed in the form ($y = mx$).
e.g. Hooke's law: the length of a spring is directly proportional to the force on the spring.
4 ***'Predictable' curves*** ...patterns can follow predictable curves (y=x2 for instance), and that such patterns are likely to represent significant regularities in the behaviour of the system (velocity against time for a falling object for instance)
e.g. Velocity against time for a falling object. Also, the terminal velocity of a parachute against its surface area.
5 ***Complex curves*** .. some patterns can be modelled mathematically to give approximations to different parts of the curve (Hooke's law for a spring taken beyond its elastic limit for instance)
e.g. Hooke's law for a spring taken beyond its elastic limit.
6 ***Empirical relationships*** ... patterns can be purely empirical and not be easily represented by any simple mathematical relationship.
e.g. number of greenfly on a rose bush over time.
7 ***Anomalous data*** ... patterns in tables or graphs can show up anomalous data points which require further consideration before excluding them from further consideration.
e.g. A 'bad' measurement or datum due to human error.
8 ***Line of best fit*** ...for line graphs (and scatter graphs in some cases) a 'line of best fit' can be used to illustrate the underlying relationship, 'smoothing out' some of the inherent (uncontrolled) variation and human error

20 Reliability and validity of the data in the whole investigation
R Gott, S Duggan, R Roberts, A Hussain Page 12 of 14

0 ***Introduction*** An overall solution to a problem can included repeated experiments and triangulation from other data sources.
So far we have considered the data within a single investigation. In reality the results of an investigation will usually be compared with evidence from other investigations.
In evaluating the whole investigation, all the foregoing ideas about evidence need to be considered in relation to the two overarching

questions:
- are the data reliable?
- are the data valid?
In addressing these two questions, ideas associated with the making of single measurements and with each and every datum in an investigation should be considered. The evaluation should also include a consideration of the design of an investigation, as well as ideas associated with measurement, with the presentation of data, and with the interpretation of patterns and relationships.

1 ***A series of experiments*** ...a series of experiments can add to the reliability and validity of evidence even if, individually, their precision does not allow much weight to be placed on the results of any one experiment alone.

2 ***Secondary Data*** ...data collected by others is a valuable source of additional evidence, provided its value as evidence can be judged.
e.g. comparing the results with data reported from other sources

3 ***Triangulation*** ...triangulation with other methods can strengthen the validity of the evidence.

21 Relevant societal aspects

0 ***Introduction*** Evidence must be considered in the light of personal and social experience and the status of the investigators.
If we are faced with evidence and we want to arrive at a judgement or decision that leads to action, other factors outside the domain of science may become relevant, some of which are listed here.

1 ***Credibility of evidence*** ... credibility has a lot to do with face validity: consistency of the evidence with conventional ideas, with common sense, and with personal experience. Credibility increases with the degree of scientific consensus on the evidence or on theories that support the evidence. Credibility can also turn on the type of evidence presented, for instance, statistical versus anecdotal evidence.
e.g. Evidence showing low emissions of dioxins from a smokestack is compromised by photos of black smoke spewing from the smokestack (even though dioxins are relatively colourless). Also, concern for potential health hazards for workers in some industries often begins with anecdotal evidence, but is initially rejected as not being scientifically credible.

2 ***Practicality of consequences***... the implications of the evidence may be practical and cost effective, or they may not be. The more impractical or costly the implications, the greater the demand for higher standards of validity and reliability of the evidence.
e.g. The negative side effects of a drug may outweigh its benefits, for all but terminally ill patients. Also, when judging the evidence on the source of acid rain, Americans will likely demand a greater degree of certainty of the evidence than Canadians who live down wind, because of the cost to American industries to reduce sulphur dioxide emissions.

3 ***Experimenter bias*** ...evidence must be scrutinized for inherent bias of the experimenters. Possible bias may be due to funding sources, intellectual rigidity, or an

allegiance to an ideology such as scientism, religious fundamentalism, socialism, or capitalism, to name but a few. Bias is also directly related to interest: Who benefits? Who is burdened?
e.g. Studying the link between cancer and smoking funded by the tobacco industry; or studying the health effects of genetically modified foods funded by Green Peace. Also, the acid rain issue (above) illustrates different interests on each side of the Canadian/American border.
4 ***Power structures*** ...evidence can be accorded undue weight, or dismissed too lightly, simply by virtue of its political significance or due to influential bodies. Trust can often be a factor here. Sometimes people are influenced by past occurrences of broken trust by government agencies, by industry spokespersons, or by special interest groups.
e.g. Studies published in the New England Journal of Medicine tend to receive greater weight than other studies. Also, the pharmaceutical industry's negative reaction to Dr. Olivieri's research results that were not supportive of their drug Apotex at Toronto's Hospital for Sick Children in 2001
5 ***Paradigms of practice*** ...different investigators may work within different paradigms of research. For instance, engineers operate from a different perspective than scientists. Thus, evidence garnered within one paradigm may take on quite a different status when viewed from another paradigm of practice.
e.g. Theoretical scientists tend to use evidence to support arguments for advancing a theory or model, whereas scientists working for an NGO, for instance, tend to use evidence to solve a problem at hand within a short time period. Theoretical scientists have the luxury of subscribing to higher standards of validity and reliability for their evidence.
6 ***Acceptability of consequences***...Evidence can be denied or dismissed for what may appear to be illogical reasons such as public and political fear of its consequences. Prejudice and preconceptions play a part here.
.e.g. During the tainted blood controversies in the mid 1980s, the Canadian Red Cross had difficulty accepting evidence concerning the transmission of HIV in blood transfusions. BSE and traffic pollution are examples in Europe.
7 ***Status of experimenters*** ...the academic or professional status, experience and authority of the experimenters may influence the weight which is placed on the evidence.
e.g. Nobel laureates may have their evidence accepted more easily than new researchers' evidence. Also, A botanist's established reputation affects the credibility of his or her testimony concerning legal evidence in a courtroom.
8 ***Validity of conclusions*** ...conclusions must be limited to the data available and not go beyond them through inappropriate generalisation, interpolation or extrapolation
e.g. The beneficial effects of a pharmaceutical may be limited to the population sample used in the human trials of the new drug. Also, evidence acquired from a male population concerning a particular cardiac problem may not apply as widely to a female population.
R Gott, S Duggan, R Roberts, A Hussain Page 14 of 14

APPENDIX B

Titration Analysis of Vinegar

Situation 44

Context: The objective of this Laboratory Exercise is to find the $[CH_3COOH]$ (aq) in a sample of vinegar. Vinegar which is a weak acid is conveniently analyzed by titration with sodium hydroxide which is a strong base.

Experimental Design: Several (at least 4) samples of CH_3COOH are titrated with a *standard* 0.202 mol/L solution of NaOH (aq). Phenolphthalein indicator, which changes from colorless to pink, at the end-point of the reaction, is used to detect the equivalence point. The reaction may be summarized as:

$CH_3COOH(aq)$ + $NaOH(aq)$ ---> $CH_3COO^- Na^+$ (aq) + H OH(l)

10.00 mL 13.00 mL

0.202 mol/L

Lab Exercise 7H Titration Analysis of Vinegar

Vinegar is conveniently analyzed by titration with sodium hydroxide. Complete the Analysis of the investigation report.

Problem

What is the concentration of acetic acid, $CHaCOOH_{1aq)}$, in a sample of vinegar?

Experimental Design

Several 10.00 mL samples of acetic acid are titrated with a standard 0.202 mol/L solution of $NaOH_{1aq)}$. Phenolphthalein indicator, which changes from -colorless to pink at the endpoint of the reaction, is used to detect the equivalence point.

Evidence

TITRATION OF 10.00 mL OF ACETIC ACID WITH 0.202 mol/L $NaOH_{1aq)}$

Trial	1	2	3	4
Final buret reading (mL)	13.9	26.9 13.9	39.8 26.9	13.6
Initial buret reading (mL) Volume	0.2	13.0 light	12.9 light	0.5
of NaOH(aq) added (ml) Color at	13.7 dark	pink	pink	13.1
endpoint	pink			light pink

(Jenkins et al., 1993, p. 191)

From the "Evidence" under Lab Exercise 7H, taken from Jenkins et al., (1993, p. 191), using the relationship $c = n/V$ where c = concentration in moles/litre

n = number of moles

V = volume in litres

$$\text{Number of moles of NaOH} = c \times V = \frac{0.202\ \text{mol} \times 13.0\ \text{mL}}{1\text{L}} = 2.63\ \text{mmol}$$

$$\text{Number of moles of } CH_3COOH = 2.63\ \text{mmol} \times \frac{1}{1} = 2.63\ \text{mmol}$$

$$\text{Concentration of } CH_3COOH = \frac{2.63\ \text{mmol}}{10.00\ \text{mL}} = 0.263\ \text{mol/L}$$

According to the evidence given about the titration, the molar concentration of CH_3COOH in the sample of vinegar is 0.263 mol/L.

The above titration analysis of vinegar was coded as Situation 44 in my analysis of Situations. The CoEs identified in this Situation are tabulated below.

Coding of CoEs in Situation 44

Coding	*Justification*
1 (3)(4) - <u>Fundamental Ideas</u>	
	(3) - Association and Causation – Link between the concentrations of [CH_3COOH] and [NaOH] are linked to a third variable, the color of the indicator, Phenolphthalein.
	(4) - Type of Meausurement – Ratio data
2 (2)(5) - <u>Observation</u>	
	(2) – Observing Events – The event observed here is the neutralization of CH_3COOH by the NaOH "seen" through the conceptual window of detecting equivalence points through colour changes in the indicator.
	(5) – Observation and Experiment – Observation can be the start of an investigation or experiment or survey.

3 (1)(2) - <u>Measurement</u>

(1) – Inherent Variation – The titre values will never repeat exactly unless all possible variations are controlled between measurements – circumstances which are very difficult to create.

(2) – Human Error – The non-equivalence of titre values may be subject to human error.

4 (1)(3) - <u>Instruments: Underlying Relationships</u>

(1) – Linear Relationships – There may be a linear relationship between the concentrations of [CH_3COOH] and [NaOH].

(3) – Complex Relationships – The relationship of these two variables i.e. [CH_3COOH] and [NaOH] is linked by a complex relationship to the indicator present (i.e. Phenolphthalein).

5 (1)(3)(7)(8)(9) - <u>Instruments: Calibration and Error</u>

(1) – End-Points – The burette and pipette must be calibrated at the end-points of the scale.

Calibration of a pipette is a separate task by itself. The burette must be checked for flow and calibrated at 0mL and the highest point usually 50mLs.

(3) – Zero Errors – There can be a zero error, a systematic shift in the scale. Instruments such as burettes <u>should</u> be checked regularly.

(7) – Specificity – An instrument such as a pipette should measure what it professes to measure. Pipettes can be calibrated individually.

(8) – Instrument Use – There is a prescribed procedure for using an instrument, which if not followed, would lead to systematic and /or random error.

(9) - Human Error – Even when an instrument is chosen and used appropriately, human error can occur.

6 (1)(2)(3)(4) - <u>Reliability and Validity of a Single Measurement</u>

(1) – Reliability – A reliable measurement requires the average of a number of repeated readings. The average of the titre values is taken.

(2) – Reliability – Instruments are prone to inherent inaccuracies: so, using different burettes/pipettes can increase the reliability.

(3) – Reliability – Human error in the use of an instrument can be overcome by independent, random checks.

(4) – Validity – Measures such as equivalence point detection - which is a complex, multi relationship – must ensure that they are measuring what they profess to measure.

7 (1)(2)(3)(5) - <u>Choice of an Instrument for Measuring a Datum</u>

(1) – Accuracy – Measure of extent to which repeated readings of titre value give a mean that is the same as the "true" mean.

(2) – Non-repeatability – Repeated readings of the titre value never/seldom give exactly the same answer.

(3) – Precision – Is an indication of the spread of repeated measurements around the mean. In precise measurements of the titre values, the readings should cluster closely together.

(5) – Outliers in Relationships – Anomalous titre values should be examined to discover possible causes.

8 (1)(2)(1)(2)(3)(4) - <u>Sampling a Datum</u>

(1) – Sampling – Making approximately 4 measurements of titre values out of all possible measurements.

(2) – Size of Sample – Is number of measurements taken. The greater the number of readings taken, the greater the representativeness of the whole population.

(3) – Reducing Bias in Sample – Technique used here is random sampling.

(4) – Anomalous Data – Could be due to an inherent variation in data or could be due to a recognized, uncontrolled variable.

9 (2) - Statistical Treatment of Measurements of a Single Datum

(2) – Mean – Of titre values used in calculation.

10 (1)(2) - Reliability and Validity of a Datum

(1) – Reliability – The uncertainty associated with the burette and pipette must be ascertained along with the uncertainty in the measurement procedure. Only then can the datum be weighed as evidence.

(2) – Validity – The measurements made by burette and pipette must allow the calculation of the concentration of CH_3COOH in a sample of vinegar.

11 (1)(2)(3)(6) – Design of Investigations: Variable Structure

(1) – Independent Variable – 10.00mL of CH_3COOH

(2) – Dependent Variable – Volume of NaOH (0.202M) solution required to reach end-point.

(3) – Correlated Variable – Changes in colour of indicator at end-point.

(6) – Continuous Variable – Volume of NaOH and CH_3COOH can have any numerical value.

12 (1)(2) – Design: Validity, "Fair Tests" and Controls

(1) – "Fair Test" – Is one in which only one independent variable has been allowed to affect the dependent variable at one time.

(2) – Controlled Variable – Volume of CH_3COOH in this set of titrations; concentration of NaOH; and temperature.

13 (1)(2)(3)(4)(6) – Design: Choosing Values

(1) – Trial Run – Can be used to establish the broad parameters required of the experiment. A few drops of indicator (Phenolphthalein) must be added and the titre value roughly checked out in the first titration.

(2) – The Sample – Issues of sample size and representativeness apply as in **8**

(3) – Relative Scale – The choice of sensible values for volume of independent variable (i.e. volume of CH_3COOH) is necessary if the measurement of the dependent variable (i.e. volume of NaOH) is to be meaningful.

(4) – Range – Over which the values of the volume of CH_3COOH is chosen is important in ensuring that any pattern is detected. The volume of CH_3COOH pipetted out for the titration is, therefore, very important.

(6) – Number – A sufficient number of titre values are needed to determine the pattern.

14 (1)(2) – <u>Design: Accuracy and Precision</u>

(1) – Determining Differences – The level of precision chosen must be such as will allow a discrimination between two or more means.

(2) – Determining Patterns – A level of precision is required for the trend in a pattern to be

determined.

15 (1) - Design: Tables

(1) – Tables – Can be used to display and organize each of the trials and show how the titre value is arrived at in each titration.

16 (1)(2) – Reliability and Validity of the Design

(1) – Reliability of the Design – All the ideas associated with the measurement of each and every datum were considered. For example, factors associated with the choice of the measuring instrument; error associated with each measuring instrument; accuracy and precision of the measurements; the range of the interval of measurements; and the number of readings.

(2) – Validity of the Design – Includes choice of measuring instrument in relation to whether the instrument is actually measuring what it is supposed to measure.

17 (1)(4) – Data Presentation

(1) – Tables – A means of reporting and displaying the titre values and how they were arrived at

(i.e. showing any necessary calculations and derivations).

(4) – Scatter Graph – Not possible as not continuous pH values at different points in titration.

19 (1)(2)(6) – Patterns and Relationships in Data

(1) – Types of Patterns – The different types of associations here are causal and consequential.

(2) – Linear Relationship – Exists between concentration of CH_3COOH and NaOH used to titrate it.

(6) – Empirical Relationship – Patterns such as that between equivalence point of titration and end-point of indicator, can be purely empirical and not be easily represented by any mathematical relationship.

20 (1)(2)(3) – Reliability and Validity of Data in the Whole Investigation

(1) – A Series of Experiments – Can add to the reliability and validity of evidence.

(2) – Secondary Data – Data collected by others is a valuable source of additional evidence. Compiling results of different groups working on the same Lab Exercise could represent data by others.

21 (1)(8) – <u>Relevant Societal Aspects</u>

(1) – Credibility of Evidence – Has a lot to do with face validity. Credibility increases with degree of scientific consensus on the evidence and theories that support the evidence. There is scientific consensus on the result of an acid-base reaction and one can predict titre values of NaOH from the concentration and titre volumes of the acid, and concentration of NaOH used. This is a well-established scientific theory (i.e., $c_1V_1 = c_2V_2$).

(8) – Validity of Conclusions – Conclusions must be limited to the data available and not go beyond them through inappropriate generalization, interpolation or extrapolation.