CHAPTER 2

BACKGROUND TO THE TIMSS 2003 STUDY

2.1 INTRODUCTION

The International Association for the Evaluation of Educational Achievement (IEA), an international, independent, and non-profit organization, has conducted international comparative studies including mathematics, science, and language since the 1960s. The ultimate goal of such studies is to identify the factors likely to influence student learning and help policymakers or educational practitioners manipulate them to improve student achievement around the world.

One of studies conducted by the IEA, the Trends in International Mathematics and Science Study (TIMSS), is a large scale international comparative study of student achievement in mathematics and science, conducted every four years from 1995. The studies were initiated to develop cross-national achievement tests and administer these with various educational systems.

One of the predecessors of TIMSS, the First IEA Science Study (FISS) was conducted during the 1970-1971 school year in eighteen countries (Comber & Keeves, 1973). The Second International Science Study (SISS) collected data from 23 countries from 1983 to 1984 (Postlethwaite & Wiley, 1992). The IEA conducted FIMS (the First International Mathematics Study) in 1964 with 12 education systems taking part. SIMS (the Second International Mathematics Study) was conducted in 1980-82 with 20 education systems participating (Travers & Westbury, 1989). The third IEA study in science was combined with an assessment of mathematics, conducted from 1995 to 1996, and was known as the Third International Mathematics and Science Study (TIMSS).

In 1999, the IEA repeated TIMSS to estimate trends in student achievement from 1995 at Grade 8, and it was called, appropriately, TIMSS-Repeat. From 1995 onwards, TIMSS has been conducted in a four-year cycle, and the first word of the acronymic title changed from “Third” to “Trends in” (International Mathematics and Science Study). Nearly fifty countries participated in TIMSS 2003, and nearly seventy in the most recently conducted study TIMSS 2007 (see Table 2.1, below).

TIMSS provides participating countries with an opportunity to gain various and comparative perspectives about their learners’ achievement in mathematics and science as well as the educational system. First, the regular cycle of TIMSS studies allows the participating countries to measure progress in educational achievement of mathematics and science. Secondly, the comparisons between achievements of countries may suggest reasons for differences. Thirdly, TIMSS can help each country enhance evaluation of the efficacy of mathematics and science teaching and learning. Lastly, TIMSS highlights growth in mathematical and scientific knowledge and skills from Grade 4 to Grade 8 (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski & O’Conner, 2003).

Table 2.1 ***IEA Mathematics and science studies conducted from 1964-2007***

	Year	Number of countries	Population (grade)
FIMS	1964	12	8, final
FISS	1970-1971	18	4, 8, final[5]
SIMS	1980-1982	20	8, final
SISS	1983-1984	23 (Korea)	4, 8, final
TIMSS 1995	1994-1995	45 (Korea, SA)	4, 8, final
TIMSS 1999	1999	39 (Korea, SA)	8
TIMSS 2003	2003	49 (Korea, SA)	4, 8

---

[5] 4, 8, and final mean the grade level intended to represent four, eight, and final years of schooling respectively.

TIMSS 2007	2007	68 (Korea)	4, 8

Besides the assessment of students' achievement in mathematics and science, TIMSS collects contextual data in the form of questionnaires. The questionnaires are administered to the student, teacher, school, and National Research Coordinators (NRCs) to provide comprehensive information about the context as well as the intended and implemented curriculum within the education system.

Data[6] provided about students' achievement in relation to different types of curricula or education systems, instructional practices and school environments has been a resource of secondary analyses in educational research fields (Howie & Plomp, 2006). The results have created many debatable issues nationally and internationally (Bracey, 1998; Wang, 1998a; Cheng & Cheung, 1999), with participating countries reconsidering their own curricula and introducing educational reforms (Reynolds, Muijs & Treharne, 2003).

Although TIMSS is designed to evaluate science as well as mathematics, most secondary analyses tend to focus on the latter (Bos, 2002; Howie, 2002; Papanastasiou, 2002; Ramírez, 2004; O'Dwyer, 2005). Although science and mathematics are closely related, there is a need to focus on science uniquely and to suggest possible interventions for the improvements to science education. Ideally, disappointing results of TIMSS could contribute to the development of more effective science education in participating countries (Duit & Treagust, 2003).

The rest of the chapter provides a general overview of TIMSS, in particular design issues and logic, instruments, and data quality, with the aim of providing a brief insight into the topic. Design issues are discussed in Section 2.2, design logic of TIMSS in Section 2.3, and, based on the design logic, instruments are

---

[6] Although a Latin plural of datum, for grammatical purposes 'data' may also be used as an uncountable singular, as in this dissertation.

explored in Section 2.4. Finally, data transformation and data quality are explored in Section 2.5 and in Section 2.6 respectively.

2.2 DESIGN ISSUES REGARDING TIMSS

To address the TIMSS test, several global institutions were involved in the development of the instruments, administration of the test, and management of the data collected. This section shows briefly how the study was organized across 50 countries, how the objects were sampled, and how the data were collected.

2.2.1 ORGANIZATION OF TIMSS

Starting with 12 participating countries in 1964, there were 49 in TIMSS 2003, and 70 countries in TIMSS 2007, the latest. TIMSS is conducted under the auspices of the IEA, located in Amsterdam and controlled by three task forces, each responsible for a specific task. Firstly, the International Study Centre (ISC) is in charge of the design, development, and implementation of the study. More specifically, the Centre is responsible for the development of the assessment framework, assessment instrument and survey procedures, the certifying of the quality in data collection, the analysis of the data, and the reporting of the results. Secondly, the IEA Data Processing Centre takes charge of processing and verifying the data submitted by the participating countries, followed by the construction of an international database. Finally, Statistics Canada deals with collecting and evaluating the sampling documentation from the participating countries and calculating the sampling weights. In each participating country, a National Research Coordinator (NRC) and a national centre organize all aspects of TIMSS within that country (Martin, Mullis & Chrostowski, 2004).

2.2.2 SAMPLING

IEA studies mainly target all the students at the end of Grades 4 and 8, and the final year of formal schooling in the participating countries. Recently, the studies have focused on Grades 4 and 8 only. TIMSS 2003 had two target populations, but which grades participate in the test depends on each country's choice. The two target populations are defined as follows:

- Population 1: All students enrolled in the upper of the two adjacent grades that contain the largest proportion of nine-year-olds at the time of testing. This grade level was intended to represent four years of schooling, counting from the first year of primary or elementary schooling. It was Grade 4 in most countries.

- Population 2: All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling. It was Grade 8 in most countries (Martin Mullis & Chrostowski, 2004).

All participating countries were expected to define their *national desired populations* based on the definition of the *international desired populations* mentioned above. Each participating country used its national desired population to select its *national defined population*, which included at least 95 percent of the national desired populations, and the NRCs estimated the size of the target population to ensure it was as close as possible to the international target. In the process of sampling, there could be some exclusions, for instance, exclusions from national coverage; school-level exclusions, which could result from geographically remote regions or extremely small size; and within-school exclusions, which could occur due to intellectually disabled students or non-native language speakers (Martin, Mullis & Chrostowski, 2004).

At the first phase of sampling, stratification was made to group sampling units. Stratification improves the efficiency of the sample design, makes survey estimates more reliable, and ensures adequate representation in the sample of specific groups from the target population. TIMSS adopted a three-stage stratified cluster design, which selected a sample of schools from all those available, randomly selecting a science class from each sampled school, and sampling students within a sampled class. In addition, TIMSS involved explicit and implicit stratification. Explicit stratification involves separate sampling frames dependent on such stratification variables as geographic regions. This explicit stratification ensures disproportionate allocation of the school sample across strata. As opposed to explicit stratification, implicit stratification involves a single school sampling frame and sorts the schools in it according to a set of stratification variables. This stratification aims at ensuring proportional sample allocation, avoiding the complexity of explicit stratification as well as improving the reliability of survey estimates (Martin, Mullis & Chrostowski, 2004).

The selection of sampled schools was also carried out using a systematic probability-proportional-to-size (PPS) technique, as it is easy to implement and verify. The schools were listed by a measure of the size (MOS) of the sampling units corresponding to the number of students in the school in the target grade. The schools were sampled by the sampling interval given by dividing the total MOS by the number of schools to be sampled, and a random number in the range between 0 and the sampling interval. Sampled schools were all taken into consideration in terms of whether or not small they could increase sampling variance. Large schools could cause operational problems (Martin, Mullis & Chrostowski, 2004).

Once a school was selected, one classroom per school was sampled by means of PPS sampling within the schools. It should be noted that intact classes were sampled to analyze relationships between student achievement and teacher level data at the class level. When a sampled classroom was smaller than half the specified minimum cluster size, the classroom was combined with another classroom from the same grade and school. When a sampled class size was large, the fixed number of students was sub-sampled, using systematic sampling whereby all students in a sampled classroom were assigned equal selection probabilities (Martin, Mullis & Chrostowski, 2004).

2.2.3 DATA COLLECTION

TIMSS was administered near the end of the school year. Accordingly, countries in the Southern Hemisphere administered the test in October or November 2002, and countries in the Northern Hemisphere in April, May, or June 2003. The assessment booklets were organized into two sessions (Part I and II), having three item blocks respectively. These were administered to Grade 8 students in the sampled classroom for 90 minutes with a 20-minute break between the parts (Martin, Mullis & Chrostowski, 2004).

Each participating country carried out all aspects of the data collection using standardized procedures developed for the study and based on training manuals created for school coordinators and test administrators. A Quality Control Monitor (QCM) was appointed by the TIMSS & PIRLS International Study Centre to monitor compliance with standardized procedures for their countries. The QCM interviewed the NRC in each of the participating countries and visited the 15 sites (schools) sampled, observing the participants during the test administration.

After the administration of the TIMSS 2003 assessment, the NRC in each country dealt with the procedures of scoring the constructed-response items to ensure reliability of scoring. The data scored in each country was submitted to the IEA Data Processing Centre for verification, and the construction of an international database.

2.3 RESEARCH DESIGN

The design for TIMSS is based on a conceptual framework developed by TIMSS for the international studies. The framework is specific to mathematics and science in TIMSS and the instruments were developed according to it. The focus of exploration is placed on the TIMSS curriculum model in Section 2.3.1 and the science framework in Section 2.3.2, but not the mathematics framework.

2.3.1 THE TIMSS CURRICULUM MODEL

TIMSS has examined the schooling system from a curriculum point of view to explore how educational opportunities are provided to students, how students use these opportunities, and which factors operate across them. Since SIMS, TIMSS has developed a curriculum-based conceptual framework which includes three levels, viz., the intended, implemented, and attained curricula, as shown in Figure 2.1 (below). The three-dimensional curriculum model indicates what students need to learn, how educational systems should be arranged to promote student's effective learning, what is actually taught in classroom by whom, and how, and what students have learned and their attitudes towards mathematics and science (Mullis et al., 2003).

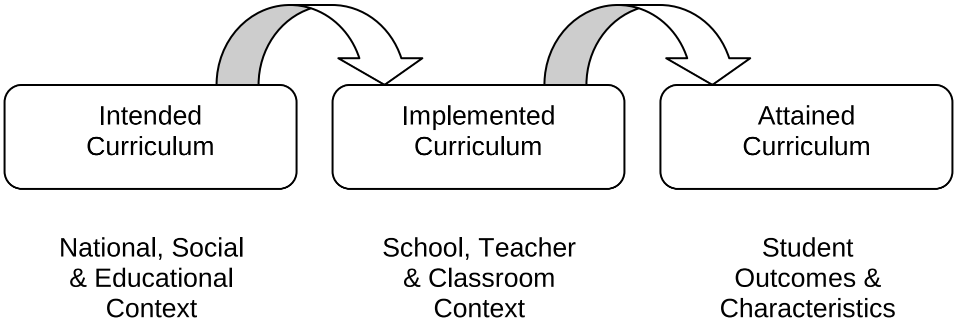

Figure 2.1 **TIMSS curriculum model (Mullis et al., 2003, p.3)**

The intended curriculum reflects society's request for teaching and learning in mathematics and science, and an educational system tends to plan it for a specific subject at the contextual level. The intended curriculum can be materialized in the form of curriculum documents that identify goal statements, prescribed textbooks, syllabi, evaluation policy, and other educational resources.

The implemented curriculum is about what is actually taught in the classroom, that is the intended curriculum as interpreted and translated by teachers in the classroom at school level. Teachers tend to carry out the intended curriculum according to their experience and beliefs regarding the subject. The classroom is the place where teaching and learning happens and teachers decide what is actually taught.

The attained curriculum is what students have learned, and includes the attitudes towards subjects. It may be evaluated by performance tests, the results of which ensure feedback to inform improvement of the intended or implemented curriculum. Ultimately, the attained curriculum is the main focus of many international comparative studies, such as TIMSS.

Based on the curriculum model described above, work to update the frameworks was carried out in line with a review of the TIMSS 1999 curriculum data to identify mathematics and science topics emphasized in the curricula of the TIMSS countries. In addition, the TIMSS framework includes contextual factors influencing students' learning in mathematics and science, and is discussed in the following section.

2.3.2 THE TIMSS SCIENCE FRAMEWORK

As stated above, TIMSS assesses mathematics and science separately. The starting point was mathematics, with science being built on the basis of the mathematics framework operation. Taking a brief look at the mathematics framework as a reference for science, the assessment framework for TIMSS 2003 was structured by two organizing dimensions, content and cognitive, corresponding to those used in the earlier TIMSS assessments. The content dimension consisted of five domains, namely number, algebra, measurement, geometry, and data. The cognitive dimension comprised four domains, i.e., knowing facts and procedures, using concepts, solving routine problems, and reasoning.

The science assessment framework for TIMSS 2003, as in the mathematics framework, includes the two organizing dimensions though here the five content domains are life science, chemistry, physics, earth science, and environmental science. The cognitive dimension encompasses three domains, namely factual knowledge, conceptual understanding, and reasoning and analysis (Martin, Mullis & Chrostowski, 2004). From 2003 on, TIMSS has placed more emphasis on questions that draw out students' analytical, problem-solving, and inquiry skills and capabilities (Mullis et al., 2003). The assessment framework explored above forms a basis for the instruments presented in Section 2.4.

2.4 INSTRUMENTS

Instruments addressed in TIMSS 2003 consisted of mathematics and science achievement test items, as well as questionnaires. The achievement test was designed to assess mathematics and science knowledge and skills based on school curricula for Grade 8 learners. The assessment items were developed, dependent on the contribution of NRCs during the entire process, and based firmly on the assessment frameworks and specifications to ensure validity and reliability (Martin, Mullis & Chrostowski, 2004). The survey questionnaires were based on many factors derived from research on effective schools.

2.4.1 SCIENCE ASSESSMENT

The assigned testing time for science content is as follows: 30% for life science, 25% for physics, and 15% each for chemistry, earth science, and environmental science. Each content area has several topics (Mullis et al., 2003).

Taking a brief look at the topics of each subject domain, *life science* includes the following topics: types, characteristics, and classification of living things; structure, function, and life processes in organisms; cells and their functions; development and life cycles of organisms; reproduction and heredity; diversity, adaptation and natural selection; ecosystems; human health. Even though TIMSS specifies a separate human biology topic area, the aforementioned are all related to human biology (Mullis et al., 2003).

While both *chemistry and physics* are incorporated in physical science at Grade 4, these two areas are assessed separately at the Grade 8 level. Chemistry assesses students on the following topics: classification and composition of matter; particulate structure of matter; properties and uses of water; acids and bases; chemical change. Physics places focus on the concepts related to energy and physical processes, with students being assessed on the following topics: physical states and changes in matter; energy types, sources, and conversions; heat and temperature; light; sound and vibration; electricity and magnetism; forces and motion (Mullis et al., 2003).

It is clear that *earth science* is focused on the earth and its place in the solar system and wider universe. However, earth science is complicated since it is related to various fields, such as geology, meteorology, physics, and astronomy. As such, some of the earth science topics are taught in subjects other than science. Although there is no single picture of the earth science curriculum, TIMSS seeks to assess such concepts common across countries, such as the earth's structure and physical features; the earth's processes, cycles, and history; the earth in the solar system and the universe (Mullis et al., 2003).

Environmental science is concerned with understanding related to the interaction of humans with ecosystems, changes in the environment from manmade or natural events, and protection of the environment. It emphasizes the roles and responsibilities of science, technology, and society to maintain the environment and conserving resources. The topics covered in the test are listed as follows: changes in population; use and conservation of natural resources; changes in environments (Mullis et al., 2003).

The cognitive dimension of the assessment focuses on student skills and abilities, defined as the sets of behaviours expected of students as they are involved in science content. There are three cognitive domains: factual knowledge, conceptual understanding, and reasoning and analysis. Firstly, *a factual knowledge* base of relevant science facts, information, tools, and procedures is fundamental to execute the more complex cognitive activities in science. In order to assess factual knowledge, items can ask students to recall or recognize science facts and concepts, demonstrate scientific terms, tools, and procedures, or describe scientific properties and relationships (Mullis et al., 2003).

Secondly, *conceptual understanding* is based on factual knowledge and can be indirectly assessed by asking students to use models to illustrate structures and relationships and demonstrate scientific concepts to solve problems. The activities measuring conceptual understanding are listed as follows: illustrate with examples; compare, contrast, and classify; represent and model; find relationship between underlying concepts and observed properties; extract and apply information (Mullis et al., 2003).

Lastly, *reasoning and analysis* requires more complex tasks than the two domains mentioned above. It involves some problem-solving situations unfamiliar to students and perhaps a little more complicated. Therefore, students may be requested to analyze the problems, to select and apply the appropriate equations or formulae to solve the situation, to hypothesize or predict. Activities related to reasoning and analysis are listed as follows: to analyze, to interpret, to solve the problems; to integrate and to synthesize; to hypothesize and to predict; to design and to plan; to collect; to draw conclusions; to generalize; to evaluate; to justify solutions found (Mullis et al., 2003).

Scientific inquiry has been emphasized in contemporary science since scientific literacy becomes important as technology develops. Scientific inquiry is associated with ‘doing science’ such as demonstrating, applying and using knowledge. Items that assess scientific inquiry ask students to involve the processes of scientific investigation and draw out some of the skills related to scientific inquiry in a practical context. Therefore, students are requested to explain cause and effect or relationships between variables. The items and tasks for scientific inquiry are set in content-based contexts without being classified separately. Specifically for Grade 8, scientific inquiry items were selected from topics such as ‘life in the oceans’ and ‘Galapagos islands’ from life science and ‘metal crown’ from the physics and chemistry domains.

In terms of question types, TIMSS uses two kinds of formats, viz., multiple-choice questions and constructed-response questions (Martin, Mullis & Chrostowski, 2004). Multiple-choice questions are assigned 54% of score points, and constructed-response 46%. It is expected that the latter questions are better suited than the former for asking students to explain or interpret data than for testing students’ knowledge or experience.

In addition to the 109 multiple-choice science questions are 80 constructed-response questions, consisting of 59 short-answer items and 21 extended-response items. All of these items are divided into 14 item blocks labelled S01 through S14. Six of the blocks contain trend items from 1995 and 1999 and eight blocks include new items developed for TIMSS 2003. Each block is composed of

8-9 multiple-choice items, 3-4 short-answer items, and 1-2 extended-response items, and accordingly, the total number of items per block ranges from 11 to 16 (Martin, Mullis & Chrostowski, 2004).

There are additional 14 item blocks for mathematics named M01 to M14 in the same way, making 28 item blocks in total. Among both the 14 mathematics and 14 science blocks, six item blocks form one student booklet, with 12 different student booklets consisting of six item blocks respectively. Participating students complete just one booklet.

2.4.2 QUESTIONNAIRES

TIMSS also aims to understand the context in which students learn, to improve students' learning in science and test their achievement in science. TIMSS designed questionnaires to provide a context for the performance scores, focusing on students' backgrounds and attitudes towards science, the science curriculum, teachers of science, classroom characteristics and instruction, school context and instruction (Martin, Mullis, Gonzalez & Chrostowski, 2004). The survey for the contextual information was based on factors identified from the findings of educational research.

All questionnaires relied on self-reported information based on Likert-type scales, and stratified on four levels: curriculum, school, teacher, and student. The purpose of these questionnaires was to gather information about five broad areas, viz., curriculum, school, teachers and their preparation, classroom activities and characteristics, and students at various levels of the educational system (Mullis et al., 2003).

The curriculum questionnaire has four versions, viz., mathematics and science for Grade 4 and for Grade 8 respectively, however all are very different in terms of structure and content. The curriculum-related questionnaire is based on the formulation and organization of the curriculum, defining its scope and content, the monitoring and evaluation of the implemented curriculum, and curricular materials and support. A curriculum formulated in a country tends to reflect the societal value or attitudes towards science education, the resources available for education, and the degree of attainment expected in conjunction with the economic level of a nation (Mullis et al., 2003).

Curricular documents define the scope and content of the curriculum in the form of the knowledge, skills and attitudes for students to be acquired through education offered in a country. However, the degree or the way the goals of curriculum are achieved varies across countries. In addition, organization of the curriculum, such as a decision to teach science as separate subjects or as a single subject, can influence the student learning experience. On the other hand, the curriculum implemented in schools can be monitored or evaluated by the way of standardized tests, school inspection, and audits. When implementing the curriculum, it can be supported by training teachers or by the development and use of teaching materials, such as textbooks (Mullis et al., 2003). Accordingly, the questionnaire related to curriculum seeks to assess all these points mentioned above.

The school questionnaire has two versions, one for Grade 4 and another for Grade 8, but they do not really differ. The school questionnaire covers the school-quality-related issues such as school organization, school goals, roles of the principal, resources to support science learning, parental involvement, and a disciplined school environment. Many factors identified from the research influence student learning and achievement at the school level, for example, whether or not schools are tracked, and if they have either an academic or vocational curriculum. The time allocated for science education at the school level can also influence student learning. Research indicates that schools articulating such goals as literacy, academic excellence, personal growth, good work habits, and self-discipline, tend to perform better than others. The leadership of the school principal is reported to be associated with student achievement. General resources like teaching materials, budget for supplies, school buildings and classroom space, and subject-specific resources including computers and laboratory equipment may influence student learning. A high degree of parental involvement, including checking homework, volunteering for field trips and fund raising, can influence academic performance. Similarly, a safe and orderly school environment is important, considering that being absent or late to class decreases time for study and reflects negative attitudes towards schooling (Mullis et al., 2003).

The teacher questionnaire is designed to be addressed to the classroom teacher of the sampled class. It has two parts, viz., information about teachers and their preparation, and classroom activities and characteristics (Mullis et al., 2003). Considering that teachers are the direct operators of curriculum implementation, teacher and classroom characteristics are the most important factors influencing student learning. Specifically, qualification of science teachers has been regarded as an important factor since science instruction is involved in many more counterintuitive scientific concepts than in other subjects (Brophy & Good, 1986). Items related to teachers and identified as important include academic preparation and certification, recruitment, assignment, induction, teacher experience, teaching styles, and professional development. Research shows that all of these factors are considered as influencing student achievement.

Also included are classroom activities and characteristics and include effective-learning-related issues such as curriculum topics taught, instructional time, homework, assessment, classroom climate, use of information technology, emphasis on scientific investigation, and class size (Mullis et al., 2003). Specifically, computers have changed the ways concepts are explored, which has not been the case in the past. Reflecting the importance of teachers' academic skills and the rapid growth in information technology (IT), teacher preparation and professional development, and the use of technology were added to TIMSS 2003 (Martin, Mullis & Chrostowski, 2004).

The student questionnaire is concerned with home background and resources for learning, prior experiences, and attitudes toward learning, all of which are recognized as influential factors emanating from research. Research shows that student background is most likely to influence student achievement. Home background factors influencing achievement can be indirectly measured by investigating the number of books in the home, availability of a study desk, the educational level of the parents, the presence of a computer, and the extent to which students speak the language of instruction. In addition, students' attitudes toward schooling or science are seen as important to their learning (Martin, Mullis & Chrostowski, 2004).

Some parallel questions are used to measure the same construct from different sources. Student questionnaires consist of 23 items, some of which also have sub-categories. Teacher and school questionnaires are made up of 34 and 25 items respectively and various sub-items constitute item sets. Student, teacher, and principal questionnaires for Grade-8 science, which are data for the current research, can be referred to in Appendix B, C, and D.

2.5 DATA TRANSFORMATION

TIMSS seeks to broadly cover the science curriculum and to measure trends across assessments, and thus necessitated a matrix-sampling booklet design, in which individual students respond to only a subset of items in the assessment rather than the entire set. For this purpose, TIMSS adopted Item Response Theory (IRT), and calculated the achievement scores using IRT methods with a scale of 800 points and a standard deviation of 100 points. Although different samples of students took different blocks of items, performance could be compared across countries, as the IRT analysis provided a common scale (Mullis et al., 2003).

IRT can be considered item-free person ability measures and person-free item difficulty measures. Accordingly, although all of test takers do not answer the same items, IRT can ensure that their results are comparable (Nakamura, 2001). Under IRT, the individual item of a test is highlighted as opposed to the raw test score focused on under classical test theory. IRT can be formulated with three item parameters, viz. difficulty, discrimination, and guessing parameter depending on a logistic function model used. Difficulty as a location index indicates a point on the ability scale where the probability of correct response is 0.5 as opposed to being relative to a group of examinees under classical test theory. The discrimination parameter indicates how well an item can differentiate between examinees having a latent trait tested in question and those not having. However, it is clear that high discrimination does not mean good validity of an item and it has nothing to do with ability itself (Baker, 2001). Lastly, the guessing parameter reflects the possibility of getting the item correct by guessing alone in multiple choices.

IRT has some basic principles compared to classical test theory. Firstly, these parameters rely on items themselves, not the group tested with them. The two groups, which are at different ability levels, produce the same values of the item parameter. However, under classical test theory, these parameters rely on the ability level of the examinees responding to the items. Secondly, the examinee’s ability is not dependent on the items used to determine it. Therefore, the examinee’s ability does not vary with respect to the items used. In contrast, under classical test theory an examinee tends to get a high score on the easy test and a low score on the difficult one (Baker, 2001).

TIMSS 2003 used three distinct IRT scaling models according to item format and scoring procedures when analysing the assessment data. A three-parameter model was used for the multiple-choice items and a two-parameter for constructed-response items with only two response options.

TIMSS used a matrix-sampling design that makes each respondent test part of all the items covering a wide range of contents. The matrix-sampling method makes it possible for population characteristics to be estimated more efficiently, but cannot make precise statements about individuals. In order to offset this drawback, plausible values methodology was used in TIMSS. Even though plausible values are not the best option to explain an individual's proficiency, they estimate population characteristics consistently. By having the students' responses to the items, the item parameters calibrated, and the conditioning variables, TIMSS produced the plausible values for student proficiency. TIMSS produced five plausible values for each sampled student, the variation indicating an uncertainty associated with proficiency estimates for individual students. These plausible values were offset by information about students' background gained through the process of conditioning, in order to enforce the reliability of the student scores.

In summary, TIMSS calibrated the achievement test items estimating model parameters for each item and created principal components from the questionnaire data for the conditioning procedure. Subsequently, IRT scale scores were generated for mathematics and science and for each content domain. Finally, the proficiency scale scores were placed on the metric used in the previous assessment and the average of the mean scores was set to 500 and the standard deviation to 100.

2.6 DATA QUALITY

Examining reliability and validity is very commonly accepted when quality in educational measurement is considered. Reliability concerns the consistency of measurements and implies internal consistency, equivalence, and stability, while validity involves the credibility of results and contains predictive and concurrent validity, content-related validity, and construct validity. These two criteria, viz., reliability and validity, contribute mainly the generalisability of the results which come from the measurement addressed (Scherman, 2007).

2.6.1 VALIDITY CONSIDERATIONS IN TIMSS

To ensure the quality of the data to be collected in survey research, there are two characteristics of importance: reliability and validity. Validity refers to the inferences about "the adequacy of a scale as a measure of a specific variable" (DeVellis, 1991, p.43). As far as validity is concerned in quantitative research, it is suggested that careful sampling, appropriate instrumentation, and appropriate statistical treatments of the data can improve data validity (Cohen, Manion & Morrison, 2007).

There are several types of validity typically assessed in survey research, including content validity, criterion-related validity, and construct validity. Content validity indicates how well items measure what is intended to be covered, and in order to ensure this, items should be sampled carefully (Cohen et al., 2007). Therefore, it is assessed by experts in some aspect of the subject. Criterion-related validity involves predictive validity and concurrent validity (Gay & Airasian, 2003). Predictive validity is concerned with another instrument being administered in the future, while concurrent validity can be measured by collecting data at the same time but in different ways, such as observations, interviews, and surveys (Cohen et al., 2007). It can be said that TIMSS attempted to partly achieve concurrent validity by administering triangulation questionnaires shown in student, classroom, and school levels. Construct validity indicates theoretically how meaningful a survey instrument is, and tends to be determined after years of experience by numerous investigators (Litwin, 1995). Therefore, ensuring or building construct validity is regarded as gathering a variety of evidence to support validity, but this is not a simple process (Gay & Airasian, 2003). Specifically, 'discriminant validity', involved in researching different constructs, can be investigated by factor analysis (Cohen et al., 2007).

In particular, TIMSS placed emphasis on content validity in the process of the development of the instrument. To ensure content validity of the assessment instrument, TIMSS 2003 made a tremendous effort in developing items. To begin with, the international item pool was developed and aligned with the assessment framework. Participants from more than 30 countries and each national research centre conducted this work. In the case of science, each draft item was classified according to whether or not it was intended to measure knowledge or skills associated with the scientific inquiry strand. Finally, an initial item pool covering a broad range of science topics was developed. The initial item pool was examined, complemented, and screened in subsequent review by the mathematics and science task forces. The next review was carried out by the item review committee, along with a group of experts, then reviewed once more by the item review committee. Field-tests were also administered to representative samples of students in each country. The NRCs were involved and contributed to the development at every stage. The final forms of the test, endorsed by the NRCs, had an opportunity to be assessed by test-curriculum matching analysis to investigate the appropriateness of the TIMSS 2003 test for students in the participating countries. The results have shown that, generally, the proportion of the items judged appropriate was high (Martin, Mullis, Gonzalez & Chrostowski, 2004).

2.6.2 RELIABILITY CONSIDERATIONS IN TIMSS

Opposed to validity, that concerns the judgements about how adequate a scale is to measure a specific variable, reliability indicates how stable measurement is over time and over similar samples. In particular, in quantitative research such as used for this study, it is argued that reliability is correspondent to dependability, consistency, and replicability over time, over instruments, and over groups of respondents (Cohen et al., 2007). There are several kinds of reliability to be used in research: test-retest reliability, split-half reliability, and internal consistency reliability. In TIMSS, they also included items that had been used in the 1995 and 1999 assessments in order to ensure reliable measurement of trends over time. As a result, 74 in science, including both multiple-choice and constructed-response items, are trend items addressed in 1995 and 1999 at Grade 8.

As another way to enforce test reliability, TIMSS developed many items (383) for the assessment to be more reliable, and designed the survey using a matrix-sampling technique. Here, each item was assigned to one of a set of item blocks to ensure broad subject-matter coverage preventing overburdening of students which could decrease reliability. Since sampled students did not take the same items, TIMSS estimated student achievement using the IRT scaling method, where students' scores do not depend on using the same set of items. To improve reliability of the scaling method, TIMSS used an approach known as 'conditioning', where reliable scores are produced even though individual students respond to relatively small subsets of the total item pool.

Furthermore, TIMSS was concerned with 'inter-rater reliability' in relation to scoring the constructed responses. A back-reading process was conducted to monitor scoring reliability and a random sample of more than 100 booklets, scored independently, was compared to establish the reliability of the scoring within each country. In 2003, some student-constructed responses from 1999 were rescored to provide scoring reliability over time (Martin, Mullis & Chrostowski, 2004). As a result, Cronbach's alpha scoring reliability coefficient was as high as 0.84 in the science test overall. In particular, it was 0.87 for Korea and 0.84 for South Africa (Martin, Mullis, Gonzalez & Chrostowski, 2004).

2.7 CONCLUSION

Since 1995, the IEA has conducted global studies in science and mathematics every 4 years. The studies also include surveys to collect information about the educational system in terms of that subject. For the study, the IEA developed its own conceptual framework and instruments. The study consists of two parts including assessment and questionnaires. The assessment is with respect to science subjects and questionnaires survey for the educational background information for the students tested. TIMSS has focused on student achievement at two populations, viz., Grade 4, and 8. Data is collected at the end of the school year in each country.

Since TIMSS aims at broadly covering the science curriculum and measuring trends over years, IRT was involved in assessment design and as such a matrix-sampling booklet was issued to each student to eliminate concern about examinees’ difference in terms of achievement. Data collected was finally scored and processed within the requirements of validity and reliability. TIMSS intends to get a picture of education in the subject in question and find out the strengths and weaknesses, and ultimately inform policy changes in curriculum or instructional practice.

CHAPTER 3

RESEARCH ON FACTORS INFLUENCING STUDENT PERFORMANCE

3.1 INTRODUCTION

In this chapter, factors influencing student performance, in particular science achievement and school effectiveness research (SER) are examined. Policymakers around the world need to be able to measure the effectiveness of the education on offer in their countries, and this can be appraised by measuring outcomes gained by students. Therefore, it is not surprising that outcomes of education have been the focus of education research over the past decades. Many factors influencing student outcomes were identified at a similar time as the formulation of the SER field. By identifying effective factors, along with effective schools, researchers have developed school effectiveness models based on findings and evidence, and applied these to school improvement projects. These will be explored in this chapter.

SER is inextricably linked with teacher effectiveness research (TER) as the two areas both aim to improve student achievement. Nonetheless, SER conducted thus far has taken place mainly in developed countries, using mathematics or language achievement as a dependent variable. To address this weakness, research should be undertaken in developing countries, also investigating achievement in learning areas of particular importance to their development, for instance science. This chapter provides some background information on SER, as a conceptual framework for the study based on school effectiveness models, and reflects on effective factors related to science achievement of students. In

Section 3.2, the literature on SER is reviewed, followed in Section 3.3, by factors related to science achievement. Conclusions are drawn in Section 3.4.

3.2 SCHOOL EFFECTIVENESS RESEARCH

In this section, the historical background of SER is explored in relation to models on the development of evidence-based school effectiveness. The contribution made by SER in school improvement is examined and teacher effectiveness reviewed in the light of SER. Finally, an argument is made for SER in developing countries, particularly in science subjects.

3.2.1 THE HISTORY OF SCHOOL EFFECTIVENESS RESEARCH

SER has formed a considerable part of education research since it started in the USA in the mid-1960s (Teddlie & Reynolds, 2000). Early school effectiveness research, such as that conducted by Coleman et al. (1966), showed that school made little difference in terms of student achievement when compared to family factors. The studies conducted under the auspices of the IEA between 1966 and 1973 supported Coleman et al.'s argument (1966), resulting in a similar finding that schools had little bearing on student achievement (Walker, 1976).

However, in reaction to such a diminished view of school effectiveness, many studies were conducted which reported that schools do in fact have an impact on student achievement. Comber and Keeves (1973), examining the Second International Science Study (SISS) data, found that opportunities to learn, mostly determined by schooling, had a strong impact on student achievement in science. They contended that it is not possible to detect weak but consistent and cumulative effects of schooling at any single point in time, whereas strong family effects are more easily identifiable (Comber & Keeves, 1973). Coleman (1975) who earlier initiated SER, later reported in the secondary analysis of the IEA studies that school effectiveness varies across countries and subjects, and it does mean that schools matter and have an influence on student achievement.

In another response to the results of Coleman et al.'s report (1966), effective schools were investigated in an attempt to identify the common characteristics that make some schools more effective than others (Scheerens, 1992). The findings identified five-factors within effective schools, including strong educational leadership, emphasis on acquiring basic skills, an orderly and secure environment, high expectations of pupil attainment, and frequent assessment of pupil progress (Edmonds, 1979). A meta-analysis of the previous literature undertaken by Walberg (1990) identified nine factors which influence educational productivity from a comprehensive psychological perspective. These factors were the ability or prior achievement of students, biological development, motivation, quantity of instruction, quality of instruction, home environment, classroom or school environment, peer group environment, and mass media environment. He excluded such organizational factors of schools as size, and individual characteristics such as gender, as these factors are less alterable. More comprehensively, Scheerens and Bosker (1997), drawing on school effectiveness studies conducted mainly in 1990s, listed the most commonly mentioned factors as:

- Achievement orientation, high expectations, teacher expectations,
- Educational leadership,
- Consensus and cohesion among staff
- Curriculum quality, opportunity to learn
- School climate
- Evaluative potential,
- Parental involvement,
- Classroom climate

- Effective learning time (classroom management),
- Structured instruction,
- Independent learning,
- Differentiation, adaptive instruction
- Feedback and reinforcement

The findings from SER explored above could be applied to other areas such as school improvement programmes (Clark & McCarthy, 1983; McCormack-Larkin, 1985). Findings emerging from SER have thus been used in two ways: to identify and measure the indicators of school monitoring (Barr & Dreeben, 1983; Shavelson, McDonnell & Oakes, 1989; Mulford, 1988; Zuzovsky & Aitkin, 1990; Suter, 1995; Fitz-Gibbon, 1996; Mayer, Mullens, Moore & Ralph, 2000), and to develop an understanding of factors within SER which may contribute to the building of a conceptual framework (Scheerens, 1990; Stringfield & Slavin, 1992; Creemers, 1994). An economic-driven input and output paradigm tends to involve such school resources as expenditure per pupil and student characteristics such as socio-economic status (SES), but it does not include classroom or school processes. In contrast, taking into consideration the process factors leads to another framework, namely instructional effectiveness theory.

The most adopted theory of instructional effectiveness is Carroll's school learning theory, which consists of five factors all linked to the use of time (Carroll, 1963). Together with considering instructional effectiveness, the economic input-output paradigm was translated into an organizational paradigm, concerned with the hierarchical and multivariate nature of the school system (Zuzovsky & Aitkin, 1990). In addition, statistical progress (or computer development), such as multilevel analysis technique which assesses more accurately the effects of all levels, made this evolution possible. Along with the development of multilevel modelling, the early 1990s saw the development of integrated and multilevel educational effectiveness models based on literature (Scheerens, 1990; Stringfield & Slavin, 1992; Creemers, 1994). Such comprehensive models of school effectiveness as Creemers', Scheerens', and Stringfield and Slavin's include contextual, organizational, instructional conditions or factors presumed to enhance educational performance (Scheerens, 1992). All these aspects work towards developing the theoretical underpinning of SER.

Some research tested the conceptual models discussed above to offer empirical evidence (Reezigt, Guldemond & Creemers, 1999; Kyriakides, Campbell & Gagatsis, 2000a; 2000b; De Jong, Westerhof & Kruiter, 2004). Creemers' model has been tested against integrated and multilevel educational effectiveness models (discussed below). However, findings from research do not always support Creemers' model, including those of Reezigt et al. (1999), who tested its main assumptions on the expected effects on student achievement of individual classroom and school level factors in language and mathematics in primary school in the Netherlands. The results showed inconsistency across the subjects, and that time for learning and opportunity to learn, which are essential factors in Creemers' model, had negative effects attributable to the mismatch of the language and mathematics tested and the actual content taught by the teachers. The study implies that the possibility of different effective factors not presented in Creemers' model should be considered (Creemers, Scheerens & Reynolds, 2000).

Kyriakides et al. (2000b), using Creemers' model, reported on mathematics in a Cypriot primary school. This study revealed less disappointing results, although time on task and the quality of instruction showed little correlation with student achievement. However, the results did show multilevel influences on achievement and that the effect of the classroom was greater than that of the school, thus arguing for the importance of learning contexts. On the other hand, attention should be given to the finding of inconsistency across subjects in primary school, as in Reezigt et al. (1999), and educational effectiveness should be studied according to systems or subjects, just as effective teacher behaviour should be qualified in different grades or contexts (Brophy & Good, 1986).

De Jong et al. (2004) added to the validity of the main concepts in Creemers' model in conducting a study of mathematics in the first year of lower general education in the Netherlands. Their findings were more improved than previously seen, and revealed that time spent, opportunity to learn, and quality of instruction were strong predictors of achievement. Kyriakides (2005) tested the validity of Creemers' model in different criteria such as mathematics, Greek language, and affective aims, assuming the considerable unexplained variance at student level might be attributed to some variables that should have been included in Creemers' model. The results of Kyriakides' study, adding psychological factors such as personality and styles of thinking to the student level, showed a decrease in the unexplained variation from 24.3% to 17.6%.

The three studies examined above, viz., Reezigt et al. (1999), Kyriakides et al. (2000b), and De Jong et al. (2004), revealed that selection and collection of data related to factors in the model were important, however all reveal some shortcomings and weaknesses. Reezigt et al. admit data of the key factors, for instance, time for learning or opportunity to learn, were collected imperfectly. Kyriakides et al. depended only on questionnaires and De Jong et al. used only ethnicity and gender as social context variables for reasons of privacy, which are not considered adequate. Reflecting on this weakness, Kyriakides (2005) used 11 well-trained observers to measure factors related to quality of teaching, and the results showed factors related to teachers were more likely to influence student achievement.

As explored above, studies to test the school effectiveness models are still rare, therefore further studies, such as this secondary research, need to be undertaken in order to gain evidence-based support and give wider and deeper insight into the school effectiveness models, for instance the current study on the teaching of science in developing countries.

3.2.2 SCHOOL EFFECTIVENESS RESEARCH AND SCHOOL IMPROVEMENT

School effectiveness models based on identified effective factors and newly developed multilevel modelling, in turn, motivated some school improvement research (Teddlie & Stringfield, 1993; van der Werf, Creemers & Guldemond, 2001). The main aim of SER is to identify malleable factors to influence student achievement so that policymakers may manipulate the factors by appropriate reform projects. Therefore, the approach and knowledge base of school effectiveness could be used for school improvement and development of education systems (Scheerens, 2001). The empirical evidence of school effectiveness based on the recently developed conceptual models is still controversial and under development, however could be covered by evaluative, monitoring programmes, and reform projects, aimed at educational improvement (Muijs & Reynolds, 2000; Van der Werf, Creemers, de Jong & Klaver, 2000; Peng, Thomas, Yang & Li, 2006). It has been proposed by Reezigt and Creemers (2005) that there is a link between two areas, namely SER and school improvement, and they attempted to formulate a theoretical framework of school improvement based on a school effectiveness model. In contrast to the focus on classroom level in school effectiveness, they pointed out that the school level process tends to occupy a central position in the framework, based on effectiveness and improvement theories. This integration could result in enforcement of experiment-based evidence (Creemers, 2002; Creemers & Reezigt, 2005).

One can see more powerful results from the improvement project based on the conceptual frameworks of SER in the following examples. Teddlie and Stringfield (1993) suspected generalization of the five factors, identified in light of equity issue in 1970s, and studied effective schools across different contexts, such as low, middle and high SES, primary and secondary schools, and rural and urban areas in the Louisiana School Effectiveness study. Their findings, gathered from classroom observation, gave some insight into school improvement efforts related to teacher evaluation. Houtveen, van de Grift and Creemers (2004) conducted action research to find out if the Mathematics Improvement Program (MIP), developed from the perspective of constructivist teaching and Creemers' ideas about school effectiveness, was effective in Grade 3 of the Netherlands. The results of adaptive instruction of mathematics supported the overall positive effect of the programme, resulting in a considerable decrease of students struggling with the subject. In addition, their multilevel analysis showed that 15% of the variance in student results could be explained at the school level. These findings imply that SER can contribute to a school improvement programme.

3.2.3 SCHOOL EFFECTIVENESS RESEARCH AND TEACHER EFFECTIVENESS

SER tends to merge with instructional, or teacher effectiveness, depending more on classroom level and especially teachers' behaviour within classroom (Scheerens & Bosker, 1997; Scheerens, Bosker & Creemers, 2000; Reynolds et al., 2003). The merging of SER and TER has occurred across countries (Reynolds et al., 2002; Ellett & Teddlie, 2003; Lee, Lam & Li, 2003; Reynolds et al., 2003). The two areas are similar in that the aim of the two research areas is to identify effective factors and to improve student achievement.

Muijs and Reynolds (2000) concentrated specifically on effective teaching behaviour of teachers in mathematics classes in the UK, examining nine effective teachers together with classroom organization, and reflecting the cumulative impact of various forms of effective teaching behaviour (Sweeney, 2003). It is of interest that whole-class interactive teaching, predominant in mathematics classes in Eastern Asian countries, was introduced in the study. Multilevel analyses showed that between 60% and 100% of pupil progress on the numeracy tests was accounted for by teacher behaviour, and confirmed the relation of teaching factors with student achievement. The study concluded that whole-class interactive teaching contributes indirectly to student progress in the way that effective teaching behaviour depends on both time on task and classroom organization, and time on task, in turn is influenced by classroom organization related to whole class interactive teaching.

Traditionally, teacher effectiveness has been studied with respect to student cognitive outcomes (Brophy & Good, 1986). Recently, the need for multiple criteria for measuring SER has been raised in reaction to achievement having been the only outcome variable focused on thus far (Teddlie & Reynolds, 2000; Konu, Lintonen & Autio, 2002), and a multi-faceted teacher role has been explored reflecting the function of the school in the globalising world (Kyriakides, Campbell & Christofidou, 2002; Muijs, Campbell, Kyriakides & Robinson, 2005). Opdenakker and Van Damme (2000) researched coherence and consistency among teachers and teacher instruction, including staff co-operation, and found the relative influence of classes and schools on achievement was much higher than the influence on wellbeing.

Campbell, Kyriakides, Muijs and Robinson (2004) illustrated that teacher effectiveness, incorporating moral values, demanded independent learning and a classroom climate associated with teacher effectiveness. By the same token, Muijs et al. (2005) pointed out in their study into differentiated teacher effectiveness across different domains, such as cognitive or affective area, that teacher factors should encompass affective aspects as well as cognitive ones related to student learning. For example, teachers' high expectation towards students can facilitate and raise students' self-concepts. Kyriakides, Charalambous, Philippou and Campbell (2006) explored teachers' attitudes toward mathematics reform introduced in Cypriot primary schools recently, and reported that teachers with high efficacy beliefs held more positive attitudes towards reform and are more likely to implement it. Considering that teacher behaviour is based on their attitudes or belief, relationships between teacher behaviours and attitudes should not be ignored. The most recent study conducted by Hattingh et al. (2007) in South Africa, showed that teachers' perceptions of their learners influence their use of practical work in science classes. As shown in the many studies above, it cannot be overstated that teacher effectiveness is a vital factor in influencing student learning and achievement.

For that reason, policymakers need to improve the quality of teachers through training or evaluation programmes that include changes in approach to the curriculum, as many studies show that the identification of effective teacher behaviour or attitude is linked to teacher training or evaluation (Teddlie & Stringfield, 1993; Kyriakides et al., 2002; Lee et al., 2003; Teddlie, Stringfield & Burdett, 2003; Kyriakides, Demetriou & Charalambous, 2006). In terms of TER, effective teaching isolated from the effect the school has on student performance, can be avoided when teacher evaluation is based on the theoretical models (Kyriakides, Demetriou & Charalambous, 2006). Kyriakides et al. (2002) proposed school-based self-evaluation of teachers to overcome the traditionally limited conceptions of teaching and disconnection from teachers' professional development. At that stage, the criteria of effective teacher or teaching generated by researchers had not been linked to professional development. They argue that teachers' involvement in formulating the criteria for an effective teacher or teaching can induce teachers' commitment to professional development and eventually improve teaching and learning. The criteria identified in their study are in line with the previous research findings.

3.2.4 SCHOOL EFFECTIVENESS RESEARCH IN DEVELOPING COUNTRIES

Most SER was conducted in developed countries such as the USA, the Netherlands, the UK, and Australia, in mathematics or language, although a few studies were undertaken in developing countries (Scheerens & Bosker, 1997). Research shows that schools and teachers have a more significant effect on student learning in developing than developed countries (Heyneman & Loxley, 1983; Fuller, 1987; Fuller & Clarke, 1994). A study of van der Werf et al. (2001) conducted in Indonesia confirmed that factors at the classroom level are also relevant in developing countries, particularly the importance of quality of instruction to improve the quality of education. In a study conducted in China (Peng et al., 2006), the findings were that factors other than competitive educational aspiration or educational policy should be considered, as pointed out by Scheerens (2001). Inconsistency across subjects was also shown up by this study, and it was suggested that developing countries, where differences in educational conditions or outcomes are more numerous than in industrialized countries, should proactively focus planned changes and retroactively select indicators for the purpose of evaluation and monitoring.

Scheerens (2001) states that there are considerable differences between schools in developing countries, whereas the effect of school is minimal in developed countries. Material and human resource factors have strong effects in the developing countries but are negligible in industrialized countries, as shown in the “Heyneman-Loxley effect” (Baker et al., 2002). It was evident that there were great differences between advantaged and disadvantaged schools in South Africa (Howie, 2002; Reddy, 2005b), but in Australia there was no significant difference between rural and urban areas in terms of resource availability (Webster & Fisher, 2000), and in Korea the availability of school resources for mathematics did not have a convincing effect on achievement across schools (O’Dwyer, 2005). Scheerens (2001) points out that the effect of instructional factors receiving empirical support in developed countries is not clear in developing countries, suggesting that cultural factors are most likely to influence the effectiveness of specific educational systems in international comparative studies. This is more likely the case in comparison to East Asian countries, with its Confucian heritage.

The points above are supported in international comparative studies. Secondary analyses on TIMSS have found explanations for the variance of achievements from a perspective of culture or environment along with instructional factors. For example, House (2002) assessed the relationship between instructional practices and mathematics achievement in Chinese Taipei, and reported that cooperative learning, which had been proved as an effective instruction strategy to improve student self-confidence and achievement in Western countries, seemed not to hold for Asian students. Papanastasiou (2002) using the TIMSS data, compared attitudinal and instructional variables which differentiated 4th-grade students in Cyprus, Hong Kong and the USA. The results indicated that the same results in different contexts could be as a result of different reasons. Leung (2001) contrasted Eastern Asian mathematics compared to Western mathematics by six dichotomies:

- content versus process
- rote learning versus meaningful learning
- studying hard versus pleasurable learning
- extrinsic versus intrinsic motivations
- whole class teaching versus individualized learning
- subject versus pedagogy with respect to competence of teachers.

In spite of higher performance shown in TIMSS, Asian students’ low confidence in subjects can be attributed to Confucian culture that emphasizes modesty (Leung, 2002). Shen (2005), conducting a comparison of the US middle school system with the five high-performing Asian school systems in TIMSS, found that American schools were less valued than Asian schools by parents and students and had a relatively shorter school year, higher student body mobility, more absenteeism, and frequent class interruptions.

Such differences between developed and developing countries appeared in tracking or grouping issues as well as cultural aspects. O’Dwyer (2005) explored the relationships between the learning environments in mathematics in 23 countries from the TIMSS data. Where education systems were not being tracked, variance of achievement occurred within classrooms, unlike schools where education systems were tracked. Specifically, students in Korea were shown to be taught in the most heterogeneous classrooms, which means no tracking. For South Africa, the most homogeneous classrooms were seen in 1995, but in 1999 classrooms had more heterogeneous groups, reflecting the large shifts in the education system since 1994. Based on this finding, it could be expected that achievement in Korea was accounted for by student-level factors, whereas South African students could be more influenced by school-level factors that the current study attempts to answer.

Taking into consideration Scheerens' arguments on cultural factors and the findings from TIMSS, the factors do not necessarily have the same influence on students in different contexts (Fuller & Clarke, 1994). Even though the outcomes or phenomena are similar in different contexts, factors underlying them could vary across countries (Bos & Kuiper, 1999; Papanastasiou, 2002; House, 2006). Furthermore, the comparison of educational systems or the evaluation of effectiveness of educational systems in developing countries should make allowances for contextual factors (Fuller & Clarke, 1994; Scheerens, 1997; 2001; Harber & Muthukrishna, 2000; Reddy, 2005a). It is argued that contextual relevance and the ideological context should be taken into account when the effectiveness of schools is evaluated (Harber & Muthukrishna, 2000). In the case of South Africa, elements of peace and democracy, such as non-violence and non-racism can be related to effectiveness from a South African point of view. As proven by Howie (2002) who examined the relationship between language and mathematics achievement, language is an issue specific to South Africa. Tracking resulting from SES and race is another issue to be considered in South Africa (O'Dwyer, 2005; Reddy, 2005b). As for Korea, an examination-driven competitive education system and Confucian culture should be considered, as in other Eastern Asian countries. As shown in Reynolds et al.'s (2002) comparative study concerning nine countries, the distinctions in school effectiveness vary across the cultures or SES, as well as across the countries. Therefore, it is plausible that schools with different contexts work differently to be effective in terms of outcomes (Teddlie & Stringfield, 1993; Reynolds et al., 2002) and educational effectiveness should be evaluated by multiple criteria, not by a single achievement test (Reynolds & Teddlie, 2000).

3.2.5 SCHOOL EFFECTIVENESS RESEARCH BASED ON SCIENCE

As mentioned above, SER focused on mathematics and language as independent variables (Scheerens et al., 2000) and consequently the findings are limited to the specific subjects. The notion that school effectiveness is subject-specific has been noted (Comber & Keeves, 1973; Coleman, 1975; Brophy & Good, 1986; Fuller & Clarke, 1994), while it was pointed out by Comber and Keeves (1973) that the effects in science could be different from other subjects such as reading, since science is more likely to be dependent on school instruction. Coleman (1975), who motivated SER, confirmed that schools had a larger impact on science rather than reading achievement of students in the secondary analysis of the IEA studies. As shown in the two consecutive studies of Kupermintz, Ennis, Hamilton, Talbert and Snow (1995), and Hamilton, Nussbaum, Kupermintz, Kerkhoven and Snow (1995), science was different from mathematics, as well as being very different from language or reading (Fuller, 1987). Their studies showed that mathematics, with its sequential-hierarchical structure of courses, was strongly affected by tracking and consequently only a few factors were shown to have an impact on achievement. In contrast, science with more likely heterogeneous content was less influenced by tracking and the effective factors vary across the content domains. From the comparison of TIMSS across participating countries, Grønmo, Kjærnsli and Lie (2004) found correlations in mathematics were much higher than in science, which means the patterns of science education across countries might be more heterogeneous, as in science content. Therefore, differential effectiveness across different subjects, or across different components, needs to be studied (Muijs et al., 2005).

Scheerens et al. (2000) also pointed out that empirical evidence needs to be supported across teachers, subjects, students, and schools. Leung, Yung and

Tso (2005) reported in the secondary analysis of Hong Kong science results in TIMSS 1999 that effective teaching methods varied between able and less able students. Besides, classroom conditions and climates influenced subjects differently. The study showed that the classroom conditions and climates influenced science achievement to a lesser extent than mathematics. Furthermore, it was found that value-added school effect was larger in science than in mathematics or language, and in developing countries than in developed countries (Scheerens & Bosker, 1997). Nonetheless, studies of school effectiveness have been rarely conducted when related to science or within developing countries.

3.3 FACTORS RELATED TO SCIENCE ACHIEVEMENT

In this section, many factors such as extrinsic and intrinsic factors, which tend to influence student achievement in science, are explored through presenting evidence from previous studies. Extrinsic factors operate from outside and can be manipulated by policy or intervention, whereas intrinsic factors are inherent in nature and cannot be changed by intervention. Although they can be discussed separately, as shown in the section below, they are interlinked. Firstly, two main extrinsic factors, time on task and opportunity to learn, are identified in the literature review. Considering that both are fundamental in each educational level, as represented in the conceptual framework, the two factors are reviewed in particular across these educational levels (3.3.1 and 3.3.2). Following the cross-level review on time and opportunity to learn, effective factors at the student level are explored more specifically, including intrinsic factors such as aptitudes, attitudes, and social context (3.3.3). Next, the classroom/teacher-level factors are investigated (3.3.4) and the factors of the school level are finally defined (3.3.5). All of the factors reviewed in these sections constitute the conceptual framework built in Chapter 4.

3.3.1 TIME ON TASK

‘Time on task’ is time spent on the learning task by students and is also called ‘effective learning time’ (Scheerens & Bosker, 1997, p.125) or ‘academic learning time’ (Creemers, 1994, p.28). It should be distinguished from ‘opportunity to learn’, which Carroll (1963) formulated as ‘time allowed for learning’ in his model of school learning in terms of time dimension. Time on task can operate according to each education level, viz., student, classroom, and school level.

At the student level, time on task contains the *time spent on doing homework*, *private tutoring*, or *outside-school activities*. Research shows that time spent on homework influences student science achievement in secondary school (Fraser, 1989; Reynolds & Walberg, 1991; 1992; Cooper, Lindsay, Nye & Greathouse, 1998). It was found that whereas there was a positive relationship between time spent on homework or daily out-of-school study time and high science achievement from the results of TIMSS and IAEP (International Assessment of Educational Progress) in higher achieving countries like Korea, this was not the case for lower achieving countries like Slovenia (Šetinc, 1999). The results of TIMSS 2003 also showed that the time spent on doing science homework was not associated with higher achievement, suggesting that the lower-performing students might be assigned more homework to keep up academically (Martin, Mullis, Gonzalez & Chrostowski, 2004). It was even reported that frequent homework was associated with lower attainment in core school subjects like mathematics, English, and science in the primary school (Farrow, Tymms & Henderson, 1999). It is apparent that teachers use homework differently, depending on the grade, and thereby the relationship between homework and achievement varies across subjects and grades (Van Voorhis, 2003), as was the case in Fraser’s study (1989) where the effects of homework were found to be negative in primary schools and positive in secondary schools, increasing with grade. For homework to be an effective means to extending the curriculum beyond school, it is evident that homework should be offered to students with consideration of appropriateness, their grade, and aims. For example, Van Voorhis (2003) found that interactive homework led to family involvement in homework and improving student science achievements and attitudes in a secondary school. In contrast, out-of-school time, namely leisure time, was found to have a negative effect on student science achievement (Fraser, 1989). This implies that there is more time spent watching television and less on learning tasks at home.

At the classroom/teacher level, the determination of time for learning can be made by the *time spent on teaching* by teachers in classrooms. In the studies conducted by IEA, FISS and SISS, time given to science teaching was proved to be related to the average achievement level of a country (Comber & Keeves, 1973; Postlethwaite & Wiley, 1992). At the classroom level, *instructional time* is important to achievement. Fraser (1989) reported that instructional time indexed by the total number of semesters of different science courses was a significant predictor of science achievement in the analysis of NAEP (National Assessment of Educational Progress) science assessment. Baker and Jones (2005) found in the secondary analyses of TIMSS and PISA that there is no consistent relationship between time spent on teaching science and science achievement internationally if one considers time allocated to science teaching; however, the frequency of interruptions to class showed a relationship with science achievement. This implies that actual time spent on teaching science influences student achievement. It was documented in TIMSS that in high performing countries, students tend to spend more time in their school and have more instructional time than in lower-performing countries (Martin, Mullis, Gonzalez, Smith & Kelly, 1999).

At the school level, time for learning involves the *time scheduled for science class,* such as *duration of class, school day per week or year* and the *frequency of field trips* as allocated by school policy. Rice (1999, p.223) reported that longer science classes in high school allow teachers more time to work with small groups of students using innovative instructional practices, and more time to discuss material as a group. Time on task, as engaged time in each learning process, is considered to be influenced by the perseverance of the student, the quality of the pedagogy, and opportunity to learn (Tate, 2001). It is evident that more time guarantees more student engagement with the learning task, though it does not mean more learning content, therefore, opportunity to learn should be taken into account along with time on task.

3.3.2 OPPORTUNITY TO LEARN

'Opportunity to learn' as opposed to learning time is mostly defined as content covered or curriculum alignment, and is measured in terms of the correspondence between learning tasks and the desired outcomes (Scheerens & Bosker, 1997). Opportunity to learn is concerned with what content is taught. The decision about what is taught, however, is made first at the context level by policy, which is called the intended curriculum in IEA studies (PIRLS and TIMSS). Following the decision made at this level, schools choose what should be taught, and teachers decide on the content to be implemented in the classroom. The review of opportunity to learn made here is explored from the higher educational level to the lower.

The intended curriculum is translated into rules or agreements about science instruction, such as *selecting a specific science textbook* and *arranging science courses* at each school. In the two studies conducted by IEA, FISS and SISS, opportunity to learn provided in the curriculum was shown to be related to the average achievement level of a country (Comber & Keeves, 1973; Postlethwaite & Wiley, 1992). In the re-analysis of the IEA studies involving reading, civics, and science, Coleman (1975) reported that science and civics are less influenced than reading by home background, which means science and civics could be more influenced by school factors. It is plausible that science knowledge differs from knowledge in such areas as reading and literature, and is more likely to be dependent on school instruction than on family factors (Comber & Keeves, 1973). This is the case especially in conditions of poverty, since schools could be the only resource that offers learners the opportunity to learn science (Reddy, 2005b). *Curriculum differentiation like tracking* or *course-taking opportunities* resulting in intra-school segregation, and thus producing differential learning opportunities, is another form of opportunity to learn (Hoffer, 1992; Spade, Columba & Vanfossen, 1997; Tate, 2001). It was also evidenced that, at the school level, course taking or course requirements can make a difference to the opportunity to learn (Hamilton et al., 1995), and the pattern of course offering and requirements showed a strong relationship with science achievement (Postlethwaite & Wiley, 1992).

Once science content to be taught is assigned at the school level, teachers make a final decision by implementation in the classroom, which is referred to as the implemented curriculum named in the IEA study. At the classroom level, the teacher can *emphasize specific content* that might be related to his/her major contribution to variance in opportunity to learn. Wang (1998b) found that *content exposure*, that is opportunity to learn, was the most significant predictor of student test scores, especially written test scores in Grade 8 science. Students make use of different opportunities to learn whether attending class or not. *Extracurricular activities like field trips* run by the school or out-of-school activities such as museum visits with parents also offer students opportunity to learn (Hamilton et al., 1995; Tate, 2001).

Opportunity to learn at the student level is mainly concerned with outside-of-school activities such as *private tutoring* and *doing homework*. *Visiting a zoo or museum, or participating in a science club* can also offer opportunities to learn (Griffin & Symington, 1997; Lindemann-Matthies & Kamer, 2006; Tran, 2007). It was found such activities as science museum visits can improve spatial-mechanical ability, which is seen to be instrumental in the variance within learning science (Hamilton et al., 1995). Additionally, *absenteeism* can negatively influence the opportunity to learn at the student level, considering that the school is the place where the main exposure to science knowledge occurs, especially in developing countries (de Feiter et al., 1995).

Opportunity to learn can be considered in terms of societal equity as well as education, because of a matter of access to content to learn. Researchers argued that opportunity to learn science is likely to be dependent on social contexts of the pupils such as SES, gender, and ethnicity as reviewed in Section 3.3.3.3 (Finn, Reis & Dulberg, 1980; Hamilton et al., 1995; Tate, 2001).

3.3.3 STUDENT BACKGROUND

This section explores student factors that are intrinsic in nature, and thus cannot be manipulated by policy as is the case with time on task or opportunity to learn. These intrinsic factors include student aptitude, attitude towards science, and SES. It should however be noted that attitude towards science is controversial in terms of manipulation's point of view as reviewed in Section 3.3.3.2.

3.3.3.1 Aptitude

'Aptitude' is described in different ways by different authors. Sometimes known as prior knowledge (hereafter both terms are interchangeable), aptitude is what the student already knows, and has been identified as the single most important factor influencing achievement (Fraser, 1989; Lindemann-Matthies & Kamer, 2006).The ability to understand instruction depends on student aptitude (Creemers, 1994). It was proposed by Walberg (1990) that aptitude consists of three elements, viz., prior achievement, biological development, and motivation or self concept. Taken as a whole, these aptitudes are defined as prior knowledge measured by tests in the early learning stages of teaching and learning. Research has found that prior achievement has a greater impact on science achievement in secondary school (Reynolds, 1991; Reynolds & Walberg, 1991; 1992). In his study of the effects of the classroom assessment environment on mathematics and science achievement, Brookhart (1997) found that prior science achievement and general reading ability had the greatest impact on science achievement. Howie (2002) found a strong relationship between mathematics achievement and English proficiency in South African students and suggested that language factors could be a substitute for student aptitudes in this context. This relationship could also hold for science, given that the results of the later study with respect to science in South Africa did not differ much from those of mathematics (Howie et al., 2008).

3.3.3.2 Attitude

Research shows a relationship between attitude and achievement. The concept of attitude can be defined as a tendency or propensity to react to things and ideas (Simpson, Koballa, Oliver & Crawley, 1994), and favourable or unfavourable feelings toward a specific object (Papanastasiou, 2000). Since attitude contains the components of affect, cognition, and behaviour, it covers values, beliefs, and motivation. Attitude, either positive or negative, is proposed as one of the outcomes to be gained (Carey & Shavelson, 1989; Reynolds & Walberg, 1992) and therefore attitude towards science can be operationalized in many different ways among researchers, including *science self-concept, the degree of enjoying science,* and *perception of the value of science* as in TIMSS.

Since Bloom (1976) reported that 25% of the variance in school achievement could be accounted for by attitudes, including affective characteristics and subject-related self-concept (p.104), research has consistently - if not as much as Bloom predicted - shown that in science education, students’ attitudes influence achievement (Freedman, 1997; Papanastasiou & Papanastasiou, 2004; Park & Park, 2006), or achievement influences attitudes (Reynolds & Walberg, 1992). More recently, relationships between attitudes and achievement in science have shown a reciprocal effect overall, although the examination by gender indicated a slightly different trend (Mattern & Schau, 2002). It was confirmed this reciprocal effect exists between attitudes and reading achievement (Williams, Williams, Kastberg & Jocelyn, 2005).

Regardless of the causal relationship between science attitudes and achievement, certain research found a correlation between attitudes and achievement (Kahle, Meece & Scantlebury, 2000; Shen & Pedulla, 2000; Papanastasiou & Zembylas, 2004; Chang & Cheng, 2008; Howie et al., 2008; Shen & Tam, 2008). Shen and Tam's (2008) cross-national examination of the TIMSS data, collected in 1995, 1999, and 2003 respectively, found that for within-country data there is a positive correlation between student achievement scores in science and mathematics. However, in a between-country analysis, the relationship is negative and these findings are consistent for both mathematics and science across the data for all three administrations, a finding which Wilkins (2004) confirmed. Papanastasiou and Zembylas (2004) examined a cross-cultural context using data from TIMSS and discussed differences in the attitude-achievement relationship in science in Cyprus, Australia, and the USA. The findings show that relationships between attitudes and achievement and the direction of the relationships or the impacts vary across the countries. For instance, high achievement generally was a good predictor of attitudes towards science in Australia. This works in reverse in that positive attitudes towards science were a good predictor of achievement in Cyprus. In the USA, high achievement had a relationship with poor attitudes, unlike in Australia. Suffice it to say that there is not an absolute or permanent relationship between science achievement and attitudes, but rather it can vary across countries, in what Papanastasiou and Zembylas called "a spatial and temporal locality of the relationship" (2004, p.259).

Generally, there are some explanations for students' attitudes, both positive and negative, towards science. Lyons (2006) found that the transmissive pedagogy, decontextualized content, and unnecessary difficulty of school science cause students' negative attitudes and lead to an aversion of careers in the field of science. Assuming that students' motivational characteristics are strongly related to the preferred kinds of learning activities and styles of teaching, when they experience less preferable instructional approaches, such experience is likely to de-motivate them (Stark & Gray, 1999).

In particular, students' negative attitudes towards science, even with high achievement, have been attributed to burn-out from examination-driven hard work (Papanastasiou & Zembylas, 2004; Murphy, Ambusaidi & Beggs, 2006), or cultural aspects such as modesty, shown in East-Asian countries (Leung, 2002). It was argued by Shen and Tam (2008) that the low confidence of the high-achiever might be due to high academic standards and expectations at the context level. By the same token, the high confidence of the lower-achiever might result from low academic standards and the expectation of society. As pointed out by Papanastasiou (2002), even though Cyprian students showed positive attitudes towards science, their achievement in TIMSS was poor, perhaps attributable to teachers' lower expectations of them.

Another feature of attitudes towards science is the decline of positive attitudes towards science as the grades progress (Greenfield, 1996; Stark & Gray, 1999; Wilkins, 2004; Murphy et al., 2006). The higher the grade, the more difficult the content (Lyons, 2006), and such decline in attitude seems unavoidable. Student achievement and attitudes are influenced jointly by a number of factors rather than by a single dominant one (Henderson, Fisher & Fraser, 2000), and these attitudes are difficult to change (Reynolds & Walberg, 1992; Papanastasiou & Papanastasiou, 2004). However, the decline of positive attitudes might not be the case globally, as proved by an example of the Singapore TIMSS results, where students retain positive attitudes towards science while maintaining high academic standards and expectations (Aun, Riley, Atputhasamy & Subramaniam, 2006; Shen & Tam, 2008).

Research shows that the quality or the nature of science instruction strongly influences student attitudes toward science (Freedman, 1997; Lyons, 2006). For example, investigating the attitudes towards science amongst 8th-grade students in Australia, Canada, Cyprus, and Korea, using the TIMSS data, Papanastasiou and Papanastasiou (2004) found that the strongest direct influences on attitudes toward science are teaching factors. In particular, instructional strategies concerned with regular practical work, laboratory instruction, and hands-on activities have been found to positively improve the student attitudes toward science, and in turn their achievement (Dechsri, Jones & Heikkinen, 1997; Freedman, 1997). George and Kaplan (1998) found that hands-on learning in the classroom or extracurricular science activities outside the school have the strongest direct effect on science attitudes. Odom, Stoddard and LaNasa (2007) concluded that attitudes and achievement among students can be improved through frequent use of student-centred teaching methods and degraded through frequent use of teacher-centred methods, indicating that attitudes towards science depend on how it is taught.

3.3.3.3 The social context of the students

Students' social contexts, which may have an influence on both attitude and achievement, refer to the *socio-economic status (SES), ethnicity, language,* and *gender* of the student. These aspects are inextricably interwoven and thus were discussed individually, as well as together. In addition, *'peer environment'* can influence student achievement and was discussed lastly.

The *SES* of students is determined by their parents' occupation and educational level, and the factor can operate in many ways, such as *parent education level, parent occupation, family size, books in the home, parent involvement,* and *mother tongue*. The home background of the student related to SES is the strongest factor influencing student achievement (see Section 3.2), and many studies show that, all being equal, students from families with a high SES outperform in science those with a low SES (O'Brien, Martinez-Pons & Kopala, 1999; Von Secker, 2004; Howie et al., 2008). The IEA studies consistently show that students from homes with extensive educational resources and/or well-educated parents have higher achievement in science than those from less advantaged backgrounds (Comber & Keeves, 1973; Postlethwaite & Wiley, 1992; Beaton, Martin, Mullis, Gonzalez, Smith & Kelly, 1996; Martin et al., 2000; 2004). It was found that home computers and visits to science museums, considered as general SES advantages, were significantly related to spatial-mechanical reasoning, which is essential in science learning (Hamilton et al., 1995). In addition, Von Secker (2004) found that students' home environment, including parents' education and literacy levels, were more strongly related to their science achievement as they progressed through school. In particular, books in the home, as an indicator of the domestic academic environment, were found to be an important factor related to student achievement in mathematics, science, and reading in the PISA study (Marks, Cresswell & Ainley, 2006). Goldhaber and Brewer (2000) reported that family background variables explain a considerable amount of the variance in Grade 12 mathematics and science test scores, in particular a statistically significant positive impact on tests by the father's level of education.

As a socio-economic indicator, parental involvement has been considered another reflection of SES (Bracey, 1996). *Parents* or *family* can influence children's education, and in turn achievement, in various ways, such as encouraging them to work hard, providing materials needed for learning, taking them to museums, or involving them in a school programme (Papanastasiou & Papanastasiou, 2004). This parental effect based on SES has been reported as being greater in science and mathematics than in reading and writing domains (Ma, 2000). Reynolds and Walberg (1991) found that the home environment factor has shown a positive, although moderate, effect on Grade 8 students' science achievement. Their study reported that student attitudes were strongly influenced by the home environment indirectly in Grade 10 science (1992).

Family support, including families' expectations of school performance, verbal encouragement or interactions regarding schoolwork was found to have a positive effect on science achievement (Cornelius-White, Garza & Hoey, 2004). In particular, parental aspirations or their expectations for their children's education achievement have been found to have the strongest relationship with students' academic achievement, including that in science (Trivette & Anderson, 1995; Fan & Chen, 2001). In addition to effects on achievement, parental involvement was proved to have a strong and direct, as well as indirect, influence on science attitudes in the way of mediation through science activities and library or museum visits (George & Kaplan, 1998). The family can also improve student achievement through helping with homework (Van Voorhis, 2003; Xu & Corno, 2003). In the USA a positive association was reported between ethnic minority students whose parents encouraged study of advanced science and their science achievement (Smith & Hausafus, 1998).

Ethnicity gaps have been found in science achievement (Greenfield, 1996; Adigwe, 1997; O'Brien et al., 1999) and generally students from ethnic majority groups record higher achievement levels in science than those students from minor ethnicity groups (Hamilton et al, 1995; Adigwe, 1997; Klein, Jovanovic, Stecher, McCaffrey, Shavelson, Haertel, Solano-Flores & Comfort, 1997). In South Africa, there are different Black ethnic groups in schools, and the language of instruction is usually English[7]. These students are faced with being taught in a *language* different from the one spoken in the home, and this contributeds to underachievement (Rollnick, 2000; Dempster, 2006; Howie et al., 2008).

From a social constructivist perspective, *language* in the science class plays an important role because scientific meaning is constructed through the social practices of teachers and learners (Fox, 2001). Rollnick (2000) contended that because of the difference between everyday language and science terminology, learning science seems to necessitate the learning of a new language, even for

---

[7] The use of English as the language of instruction is the advantage of the White English speaking minority, thus perpetuating racial inequalites.

first language speakers. Therefore, the second language learners face two challenges simultaneously, namely to study the language of teaching and learning itself, and to learn science in their classroom. In South Africa, where most students learn science in schools in a second or a third language, language proficiency is a strong factor influencing student science achievement (Howie et al., 2008).

Gender issues are not new in education with boys performing better than girls in science (Comber & Keeves, 1973; Husen, Fagerlind & Liljefors, 1974; Postlethwaite & Wiley, 1992; McCrum, 1994; Hedges & Nowell, 1995; Beller & Gafni, 1996; Lee & Burkam, 1996; Adigwe, 1997; Burkam, Lee & Smerdon, 1997; Wang & Staver, 1997; Erinosho, 1999; Martin, Mullis, Gonzalez & Chrostowski, 2004; Van Langen, Bosker & Dekkers, 2006). Girls have been found to lag behind in mathematics and science while outperforming boys in language (Hedges & Nowell, 1995; Mau, 1995; Van Langen et al., 2006). Gender gaps in science achievement are a concern and have been researched, since the gaps are substantially greater than for other school subjects (Hedges & Nowell, 1995; Beller & Gafni, 1996). In 29 out of 45 participating countries in TIMSS 2003, boys significantly outperformed girls in science (Martin, Mullis, Gonzalez & Chrostowski, 2004). Generally, the variances in boys have been found to be greater than those in girls in science (Hedges & Nowell, 1995).

Research has shown that gender differences in science achievement vary across content domains and girls fare better than boys in biology, while boys outperform girls in physics (Husen et al., 1974; Postlethwaite & Wiley, 1992; McCrum, 1994; Beller & Gafni, 1996; Lee & Burkam, 1996; Burkam et al., 1997; Erinosho, 1999; Martin, Mullis, Gonzalez & Chrostowski, 2004). Gender differences were found in specific cognitive domains, such as spatial-mechanical ability, where boys perform better than girls (Hamilton et al., 1995). As TIMSS 1995, 1999, and 2003 show, males consistently outperformed girls in physics and earth science (Beaton et al., 1996; Martin et al., 2000; 2004), evidence that the gender variance in science achievement tends to persist and increase as the student progresses through school, regardless of whether girls have positive or negative attitudes towards science (Husen et al., 1974; Postlethwaite & Wiley, 1992; Burkam et al., 1997; Klein et al., 1997; Von Secker, 2004). Ultimately, this variance leads to a reduction of women’s participation in science-related careers (Burkam et al., 1997; Erinosho, 1999; Gillibrand, 1999; Van Langen et al., 2006).

Gender gaps in science achievement were commonly explained in connection with differences in opportunity to learn, which results from differentiated educational systems, course-selection, and out-of-school science experiences. Gender gaps in science achievement appeared even among students within the same curriculum (Beller & Gafni, 1996). Students' prior science-related experiences and differential opportunity to learn, compounded by participation, cultural and social expectations, could increase gender gaps in science performance (Burkam et al., 1997). For example, it has been found that parental separation brings on an earlier-than-usual beginning of the female disadvantage in science achievement (Smith, 1992). There is a finding that the attitudes and expectations of male and female teachers are, like those of parents, greatly influenced by the traditional sex stereotyping of roles (Haussler & Hoffmann, 2002). Girls in single-sex classrooms or schools had more favourable attitudes towards science than those in mixed classrooms or schools (Dhindsa & Chung, 2003). Gillibrand (1999) found that girls who elected to study physics in a single sex class gained confidence in physics, and this was associated with better achievement. By contrast, Haussler and Hoffmann (2002) found that dividing classes according to gender has no effect on achievement, apart from improved interest in physics.

Not limited to educational factors, gender gaps in opportunity to learn science may emerge in various ways. The differences in socialization according to different social status, ethnicity, and SES may cause differentiated experiences of and interest in science (Klein et al., 1997; Jayaratne, Thomas & Trautmann, 2003). It therefore was pointed out achievement gaps in SES and ethnicity tend to be paralleled by gender gaps (Von Secker, 2004). For instance, Hamilton et al. (1995) found that the Black and Hispanic students in the USA had similar trends as girls fared better than boys in reading, but worse in spatial-mechanical reasoning. Adigwe (1997) also reported that there were significant differences in science test performance between ethnic groups as well as gender in Nigeria. Kahle et al. (2000) found in the analysis of urban African-American students that girls with more home support tended to have friends with science-oriented activities.

Some research attributed gender differences in science achievement to test format. The multiple-choice format has been found to favour males who are more willing to take this risk (Hamilton et al., 1995), while the open-ended format contributes to relatively higher performance among females (Bolger & Kellaghan, 1990). This could be attributable to the open-ended format being subject to language proficiency, in which girls tend to be stronger than boys (Van Langen et al., 2006). Hamilton (1998) found that boys outperformed girls on test items with visualization requirements and those which involved experience beyond school. Klein et al. (1997) found that girls scored slightly higher than boys on the performance assessments. These findings led to using performance assessment along with multiple-choice items, as tried in TIMSS (Kind, 1999).

Many interventions have been introduced to improve girls' attitudes towards science as gender differences in achievement tend to be mediated by parallel differences in attitudes, interests, perceived values, and self-concept (Williams et al., 2005). Instructional changes, including the adoption of regular hands-on activities, have improved girls' interest in science and reduced the gender gap (Lee & Burkam, 1996; Burkam et al., 1997). It was proposed by Van Langen, Bosker and Dekkers (2006) that integrated and comprehensive curricula and educational systems can reduce the gender gap, assuming that self-confidence for girls in differentiated versus integrated educational systems is associated with some sort of self-fulfilling prophecy mechanism and their achievement. As seen in the series of IEA studies, there have been declines in gender differences consistent with shifting educational opportunities, social roles, and the demands of the workplace (Linn & Hyde, 1989).

Lastly, *peer group* can influence student achievement. Walberg regarded peer environment as one of the important factors influencing student educational productivity (1990). The peer group was shown to influence student science achievement indirectly, mediated by instructional quality and instructional time (Reynolds & Walberg, 1991). It was documented that there was a positive correlation between peer support and academic achievement (Ashwin, 2003).

3.3.4 CLASSROOM-LEVEL FACTORS

Factors at classroom level also influence student outcomes, particularly in developing countries, where teacher and school factors prove to have a deeper effect on student science achievement than in developed countries (Heyneman & Loxley, 1983). The classroom level involves the science curriculum, the science teacher, the classroom climate, as well as the physical resources. In order for teaching and learning to take place in classrooms in practice, a science curriculum for teaching and learning should be in place with materials to support that teaching and learning. As they work together, such compositions induce unique climates in classrooms. As Creemers (1994) stated, factors identified here are important in any attempt to create an optimal composition and to enhance effectiveness, particularly if the classroom effect is higher than that of the individual factors.

3.3.4.1 Science curriculum

TIMSS conceptualizes the intended curriculum at the national level, the implemented curriculum at the teacher level, and the attained curriculum at the student level (Mullis et al., 2003). The science curriculum is mostly defined at the context level in the form of ministerial directives, instructional guides, school inspections, and recommended textbooks. At the school level, the science curriculum is considered in terms of curriculum management, as shown in Section 3.3.5.1. At the classroom level, the science curriculum is translated into science content, which is then taught using the recommended textbooks and workbooks in classroom. Therefore, the science curriculum reviewed here can be regarded as implemented curriculum at the classroom level.

When science teaching and learning take place in a classroom, a science teacher and his/her students have a science textbook or workbook as recommended at the country level according to the intended curriculum. A textbook not only represents an educational standard but also reflects comparative focuses of each educational system depending on distribution of space to different content and skills (Valverde & Schmidt, 2000). In science and mathematics textbook comparison in the USA and 21 high-achieving countries in TIMSS, it was found that coherence, focus, and level of curriculum were deficient in the USA, unlike the higher-achieving countries (Valverde & Schmidt, 2000). Most teachers use a textbook as the primary basis or a supplementary resource for their lessons (Martin, Mullis, Gonzalez & Chrostowski, 2004), implying that it helps them make decisions on the implemented curriculum, viz., opportunity to learn at the classroom. Therefore, the science textbook used in the classroom can be an important factor influencing student learning.

3.3.4.2 Teacher background

Science teacher quality examined here is divided into two aspects, including ‘teacher background’ and ‘teaching practice’. The role of the teacher in teaching and learning is important in implementing the intended curriculum. Teacher quality, depending on background and teaching practice, might be vital, given that many concepts in science are counterintuitive and difficult to understand even for adults, and under-qualified teachers may teach incorrect content or fail to correct their students' distorted understandings (Brophy & Good, 1986). Freedman (1997) argues that the quality of science education is correlated with the quality of instruction, which in turn is determined by teacher quality, and so affects student achievement (Darling-Hammond & Hudson, 1989). It is argued that the quality of teaching is also an important determinant of students' attitude towards science (Osborne, Simon, and Collins, 2003).

Mayer et al. (2000) identified four teacher characteristics as one part of school quality indicators, including *teacher academic skills, teacher experience, teaching assignment,* and *professional development,* and this is reinforced by Greenwald, Hedges and Laine (1996), who argued in meta-analysis that teacher quality, including *teacher ability, teacher education,* and *teacher experience,* was very strongly associated with student achievement. These aspects are inter-related but need to be discussed individually, as well as together.

Academic skill refers to teacher competence in terms of academic learning and is vital since it can influence subject matter knowledge and pedagogical skill. As seen in the majority of countries participating in the studies by IEA, FISS and SISS, students of teachers who were *experienced and competent* in science performed better (Comber & Keeves, 1973; Postlethwaite & Wiley, 1992). Similarly, Ehrenberg and Brewer (1994) found that the higher the quality of the institution a teacher attended, the more his or her students tended to learn.

As far as the teaching assignment is concerned, when teachers who lack subject matter knowledge teach the subject, they not only convey inaccurate content, but also fail to identify and remedy their students' misconceptions (Brophy & Good, 1986). Jita (1998) found that many science teachers in South Africa were deployed in other subjects as well as in science, and argues that teaching two or three different areas, including science, that demands more professional knowledge, might lead to teachers not being able to devote sufficient time to prepare adequately for effective teaching practice. According to Ingersoll (1999), this out-of-field teaching is likely to result in substandard teaching, and when conducted by a teacher without a strong background it might contribute to low science achievement at the Grade 12 level. Regardless of student achievement, it leads to boring teaching practice relying on textbooks, and failing to promote students' interest in the subject or development their critical-thinking ability. In addition, Ruby (2000) found that many teachers certified as K-6 teachers in the USA were often compelled to teach in middle schools, resulting in a lack of confidence in teaching science and a reduction in the intended content, especially in physical science which is considered difficult.

Teacher experience is significant in the light of teacher pedagogical content knowledge, related by Shulman (1986) to the teaching of subject matter knowledge and to be gained by means of the teaching practice as well as research. It seems practical that experienced teachers can represent topics to make their students understand better than novice teachers. Nye, Konstantopoulos and Hedges (2004) found that students learn more from experienced teachers than they do from inexperienced ones. A lack of science-teaching experience was pointed out as one of the challenges to reform of science education, particularly as more than 45 percent of general science teachers had fewer than two years' teaching experience in South Africa (Howie, 1999). In addition, TIMSS 2003 showed that the percentage (75%) of science teachers under 39 teaching Grade 8 in South Africa was higher than international average (50%) (Martin, Mullis, Gonzalez & Chrostowski, 2004). It was documented that the effects of teaching experience are curvilinear and teachers with five-to-ten years of experience have a more positive impact on achievement (Darling-Hammond, 2000; Nye, Konstantopoulos & Hedges, 2004).

Professional development, or “the process whereby teachers' professionality and/or professionalism may be considered to be enhanced” (Evans, 2002, p.131), is planned and offered by policymakers and educational reformers respectively to improve and develop teacher knowledge, skills, and practice, and thus improve student achievement. It is considered the best way to improve teaching practice, although teachers consider this an unfavourable learning source (Supovitz & Turner, 2000). In contrast to the initial intention, professional development programmes ultimately fail to change teachers' attitudes or teaching practices (Roehrig, Kruse & Kern, 2007) and short-term and event-like programmes might be regarded as contributing to such failure. In contrast, evidence shows that high quality professional development, consistently provided, improves science teachers' instruction (Kahle et al., 2000; Supovitz & Turner, 2000; Desimone, Porter, Garet, Yoon & Birman, 2002). Highly intensive, inquiry-based professional development in science and mathematics might change teachers' attitudes towards reform, their preparation, and teaching practices (Supovitz, Mayer & Kahle, 2000). Therefore, professional development which effects changes in teaching practice and classroom culture can in turn improve student achievement.

Teacher education is vital for developing subject matter and pedagogical knowledge as well as methodology prior to beginning a career. Based on a premise that the implemented curriculum may vary depending upon teachers' subject-matter knowledge and pedagogical knowledge, *teacher preparation of content* was argued to have a significant impact on teaching practice and classroom culture (Turner-Bisset, 1999; Supovitz & Turner, 2000; Darling-Hammond, 2007). Such subject-matter and pedagogical knowledge can be acquired through *pre-service education*, namely *major in undergraduate school including degree* and *certification*, and *in-service education*, namely professional development. However, the type of pre-service education is important in determining the quality of teacher training.

With respect to the relationship between teachers' formal qualifications and student achievement, it was found that the relationship between the formal education of the teacher and student results is generally weak in the West, yet this is stronger for science and mathematics than for other subjects (Brophy &

Good, 1986). Research shows that students taught by teachers holding a science degree or certification in science teaching, outperformed those with teachers who were not science-trained (Druva & Anderson, 1983; Monk, 1994). Monk (1994) found that high school students’ science test scores have a bearing on the subject-matter preparation of their teachers, although to a lesser extent in mathematics. Goldhaber and Brewer’s study (2000) contradict this, as they found no significant effect on student achievement in mathematics and science in terms of teacher certification and degree. However, this evidence should be interpreted with care, considering that most US college students selecting education majors tend to be drawn from the lower part of the ability quotient. Nonetheless, the studies also reported that subject matter preparation by means of a higher degree and certification has an effect on student achievement even after controlling for variables such as ethnicity and SES in science, albeit to a lesser extent than in mathematics (Goldhaber & Brewer, 2000).

In addition to these factors, teacher background includes *gender*. Although it was believed that there is no impact from teachers’ gender difference on student science achievement (Brophy, 1985), it was found that 15% of the variation in students' science achievement scores was due to teacher differences, and one of the two teacher factors was gender (Kahle et al., 2000). There was a higher level of science achievement in female teachers' classes in their study, with female teachers more likely to take responsibility for their students' learning than male teachers (Curtis, 1999).

3.3.4.3 Teaching practice

Effective teaching practice is a core of instructional quality along with teacher background in science, given that it can directly influence student achievement (Brophy & Good, 1986; Johnson, Kahle & Fargo, 2007). The main effects of instruction on mean science achievement of a school was analyzed by Von Secker and Lissitz (1999) and they found that instructional practices affect individual science achievement interacting with gender, minority status, and SES. Some factors with respect to instructional quality were identified in SER. Scheerens and Bosker (1997) proposed structured instruction, including structure and preparation of lessons, direct instruction, and monitoring. Creemers (1994) reported the more detailed factors under three components, teacher behaviour, grouping, and curriculum, to be explored further in Chapter 4. Wise and Okey (1983) examined the effects of various categories of teaching strategies on achievement in science in primary through high schools, and identified 12 categories of teaching techniques: “Audio-visual, Focusing, Grading, Inquiry, Manipulative, Modified, Presentation approach, Questioning, Teacher direction, Testing, Wait-time, and Miscellaneous” (p. 420).

Thereafter, Wise (1996) reported the results of a secondary meta-analysis of 140 studies comparing the effects of traditional science teaching strategies with those of alternative strategies on student science achievement at middle and secondary schools. Consequently, the 12 alternative science teaching strategies identified previously were reduced into eight categories considering usefulness: “Questioning, Focusing, Manipulation, Enhanced Materials, Testing, Inquiry, Enhanced Context, and Instructional Media” (p.337).

Recently, Schroeder, Scott, Tolson, Huang, and Lee (2007), examining the extant body of recent studies in science teaching to provide research-based evidence of effective teaching strategies, suggested ten strategies modified and employed on the basis of the Wise’s eight teaching strategy categories: “Questioning, Focusing, Manipulation, Enhanced material, Assessment, Inquiry, Enhanced context, Instructional technology, Direct instruction strategy, and Collaborative learning strategy” (pp.1445-1446). Two strategies, namely direct instruction and collaborative learning were added to the original set to reflect more recent emphasis. Two other strategies, viz., assessment and instructional technology strategy, were renamed to broadly cover the related elements. In what follows, additional research evidence is presented in an expository way along with the definitions made in the study above.

Questioning strategies are concerned with the timing and positioning of questions used by teachers and include the use of wait-time or pause at a key point. This strategy was found to have the strongest effect on student achievement in Wise's study (1996). It should be borne in mind that questioning strategies are inextricably linked with 'focusing' and 'assessment' strategies explored below. 'Teacher questioning' has evolved to interaction and the discourse taking place in the science classroom as constructivist approaches become prevalent. From a perspective of social constructivism, questioning in the science class can be adopted to clarify meanings, examine a variety of views, and finally construct scientific knowledge by means of using language. Van Zee and Minstrell (1997a) proposed the so-called ''reflective toss'' strategy, which includes a student statement, teacher question, and additional student statements to promote the responsibility for thinking in the discourse. The authors (1997b) state that the more open questions the teachers ask, and the more they acknowledge student contributions, the more students tend to engage in taking responsibility for thinking in the classroom discourse. In analyzing classroom talk and interaction in the science class, Chin (2007) stated that discourse based on questioning can help students scaffold their thinking and construct scientific conceptions.

Focusing strategies provide or reinforce objectives or use advanced organizers during the middle sections or at the closing of a class, to strong effect. As indicated above, focusing (or 'emphasizing') strategies can be examined in terms of the interactive context between a teacher and his or her students. In the examination of classroom interaction and discourse, Chin (2007) found that focusing strategies encourage students to develop productive-thinking abilities and thereby promote multi-faceted views. Her other finding showed that when teachers offer students a question-based summary, it helps them strengthen the key points of the lesson. She argued that using such strategies appropriately can serve to reinforce basic skills that students should learn, so that they can apply the basic knowledge to solve more complex problems later. It is therefore evident that focusing strategies assist in the learning of basic and existing knowledge in science.

Manipulation strategies involve students in physical activities such as operating apparatus through practical work, and permeate most laboratory activities. There is evidence that students who had regular laboratory instruction scored significantly higher achievement ratings in science knowledge than those who had no laboratory experience (Freedman, 1997). It was confirmed in documentation by Von Secker and Lissitz (1999) that instruction emphasizing laboratory inquiry was invariably associated with higher achievement. Odom et al. (2007) reported that near-daily implementation of group experiments, giving reasons for answers, solving problems, providing information to support answers, and learning from classmates, have a positive association with student achievement. Practical work was also proven to increase students' positive attitudes towards science (George & Kaplan, 1998), as well as their achievement This positive effect might be attributable to the fact that practical work makes learning science meaningful (Hattingh et al., 2007).

In particular, Burkam, Lee and Smerdon, (1997) found that practical work favoured girls and students from minorities or of low SES. Despite such positive effects of practical work, there is a reverse finding as well. For example, in science classes the time spent on laboratory and equipment per se was not related to learning. This suggests that students' active involvement in laboratory work is more important than the quantity of lab work or quality of the equipment. Different aims for practical work depending on different contexts may lead to such inconsistent results (Swain, Monk & Johnson, 1999). In addition, strict rather than helpful teacher behaviour was found to correlate negatively with practical test performance (Henderson et al., 2000). Some social constructivists argue that practical work in school science should be used as open-ended investigation intended to develop problem-solving rather than a pedagogical means of science learning (Kind, 1999).

Enhanced material strategies are those in which the teacher modifies instructional materials to make them more suitable to student needs or status. Leung et al. (2005) found that effective teaching methods on less able students were different from those used with able students, contending that teachers should adjust their instructional methods according to student need. SER lends support to this point, and Muijs et al. (2005) argue that teaching strategies should be different according to students' ability and SES. There is evidence that effective teachers adjust their teaching to fit the needs of different students and the demands of different instructional goals, topics, and methods (Darling-Hammond, 2000).

Assessment strategies include diagnostic and formative testing, immediate or explanatory feedback, and testing to mastery. Bloom (1974) named the whole procedure of the original teaching practice, the feedback, and the correctives as the quality of instruction under the mastery of learning conditions. Where the quality of instruction is high, student achievement and time on task in the classroom improve, and vice versa, with formative assessment improving student learning (Black & Wiliam, 1998). In a study into the effects of the classroom assessment environment on mathematics and science achievement, Brookhart (1997) found that the frequency of oral reports, written reports, and science projects were more important to science achievement than to mathematics achievement. Oral reports, which may be time-consuming, had negative effects, while science projects had positive effects, and written reports showed mixed effects. Black and Wiliam (1998), in examining classroom formative assessment, provided evidence that well-designed questioning, tests, and feedback in science classroom improve student learning. Chin (2006) studied classroom interaction in science and identified the various forms of feedback presented by science teachers. The feedback classified in the study was categorized into four forms:

"Affirmation-Direct Instruction, Focusing and Zooming, Explicit Correction–Direct Instruction, and Constructive Challenge" (p.1326). The author found that, in particular, 'Focusing and Zooming' and 'Constructive Challenge' feedback types prompted students' responses, encouraged generative thinking, and improved the conceptual knowledge of students.

Inquiry strategies are student-centred and relate to discovery instruction. Inquiry-based instruction covers facilitated inquiry, guided discoveries, inductive laboratories, and indirect instruction. Whereas teacher-centred strategy involves whole-class instruction, recitation, and limited independent practice, student-centred strategy has to do with active student engagement, interactive scientific inquiry, and lifelong learning. In particular, emphasis on laboratory inquiry at the school level has shown a positive relationship with science achievement (Von Secker & Lissitz, 1999). It was found that emphasis on problem-solving and understanding among the instructional factors was associated with basic knowledge and reasoning in science (Hamilton et al., 1995). Active involvement in the science classroom has shown that the gender achievement gap can decrease due to improving gender equity (Burkam et al., 1997).

Chang (1999) reported that an instructional model based on problem-solving significantly improved the achievement of students in a Taiwanese ninth grade earth science class. Kahle et al. (2000) studied the influence of standards-based teaching practices, including inquiry, problem-solving, and open-ended questioning and detected a positive effect on science achievement in urban African-American students. Similarly, Gaigher, Rogan and Braun (2006) found that a structured problem-solving strategy in physics improved South African student achievement in this area. There is therefore significant evidence that collaborative laboratory work based on student-centred and active learning in the high school classroom can lead to enhanced content knowledge and process learning for their students (Taraban, Box, Myers, Pollard & Bowen, 2007).

Enhanced context strategies are related to field trips, group discussions, self-paced learning, problem-based learning, games, and simulations. Teachers can use organizational schemes or contexts differing from the ordinary to draw students' interest and engage them in learning. It was documented that student participation in extracurricular science activities such as science clubs and fairs have significant influence on their attitudes toward science (George & Kaplan, 1998). Griffin and Symington (1997) contended from the observation of a school excursion visit to a museum in Australia that field trips should be used as informal learning and are a valuable teaching strategy. The finding showed that students who have worked on a topic at school before visiting a museum, and who have prepared for their visit, learn most from their experience. Outside-school activities such as field trips were reported to offer students physical engagement experiences to foster learning (Lindemann-Matthies & Kamer, 2006). Many outside-school activities tend to be related to biology or earth science domains, in contrast to physics learning. However, Anderson and Nashon (2007) show the possibility of physics learning based on meta-cognition in organized school visits to informal contexts.

Instructional technology strategies include instruction based on audio and video materials, media, and such technology as computers. The effect of computer use in a science class was shown to be positive, but negative in mathematics in Korea (Park & Park, 2006). There is evidence that teacher-directed computer-assisted instruction can be an alternative in teaching basic science concepts in the secondary classroom. Chang's (2003) research of the comparative efficacy of computer-assisted instruction and traditional instruction on student science learning in a Taiwanese secondary school found that students experiencing teacher-directed computer assisted instruction had significantly higher score gains than those engaged in student-controlled computer-assisted instruction in earth science. It was documented that interventions, such as the use of computer-supported learning environments, strengthen the performance of able students, whereas less able students tend to show a poorer performance.

Information and communication technology (ICT) was shown to be an educational medium for a variety of learning tasks focusing on strengthening the knowledge base and thinking skills (Taconis, Ferguson-Hessler & Broekkamp, 2001). There is, however, a reverse finding that technology use has a negative effect on science achievement (Aypay, Erdoğan & Sözer, 2007). This finding was confirmed by Waight and Abd-El-Khalick (2007), wherein the use of computer technology hampered 'inquiry' in the sixth grade science classroom, contrary to expectation. They went on to contend that this result could be attributed to less time dedicated to group discourse, which is seen to lead to critical, meaning-making conversations. This could however be because computers in science are employed for the wrong reasons, such as a substitute for solid instruction and active investigation (Burkam et al., 1997).

Direct instruction newly added by Schroeder et al. (2007), involves teachers' verbal delivery of information or explicit guides for students, for example in designing experiments, using a microscope and making measurements. Direct instruction is more likely to meet teacher-centred traditional strategies, while teacher-led direct instruction was proved to be more effective than individualized instruction (Brophy & Good, 1986). Examination of classroom interactions related to difference in students' science achievement by Zady, Porters and Dan Ochs (2003), it was confirmed by Walberg (1991) that direct teacher instruction was more prevalent with high achievers than low achievers. The many children who learned about experimental design from direct instruction learned more and performed as well as those few children who found their own way in the third and fourth grade (Klahr & Nigam, 2004). This, however, was not confirmed in the longer term framework, as Dean Jr. and Kuhn (2007) found that only when direct instruction was coalesced with regular practice, was the effect strong. Finally, Fradd and Lee (1999) contended that learners with more authoritarian cultures may benefit from a more directly explicit approach regarded as traditionally teacher-centred.

Collaborative learning strategies, reflecting the recent emphasis on grouping learning in science, arrange students in flexible groups to work on various tasks. In reality, laboratory activities, inquiry projects, or discussions, are mostly practiced in groups. Harskamp and Ding (2006) studied the effects of structured collaborative learning and individual learning in the physics class of a secondary school in Shanghai, concluding that students who learnt to solve problems in collaboration, and those who learnt to solve problems individually with information or hints, were more likely to improve their problem-solving skills than those who learnt to solve the problems individually, without hints. Group-working students tend to solve problems in a less organized way, and as Odom et al., (2007) found, groups working with student-centred strategies learn from peer interaction, and thus improve their achievement.

It should be borne in mind that each alternative strategy examined above does not run alone, but becomes integrated as effective teaching is a product of various mixed strategies employed by the teacher (Muijs & Reynolds, 2000). For instance, practical work, based on an inquiry strategy, may take place in groups, and there is evidence that successful teachers are more likely to use various teaching strategies than a single approach, considering objectives to be taught and student needs (Hanushek, 1971; Doyle, 1985). This point is supported by Taconis, Ferguson-Hessler and Broekkamp (2001), who found that while problem-solving strategies provide the learners with guidelines, criteria, and immediate feedback that improved problem-solving skills, group work without such variables did not lead to positive effects. After enumerating all these effective teaching strategies, Wise (1996) reinforced an inquiry-oriented strategy as a common feature underlying all these alternative strategies relative to traditional strategies, and suggested that teachers should take inquiry strategies as the principal approach in science instruction.

Whatever strategy is used, there is an emphasis on the importance of students' active engagement and connection with everyday life, reflecting a constant emphasis on engagement in science along with constructivism (Floden, 2001). Such findings were confirmed in a project titled *School Innovation in Science in Australia*, which identified effective teaching practices in a science classroom from a perspective of teaching and learning (Tytler, 2003; Tytler et al., 2004). Drawing from the interviews with teachers, they identified eight effective components, summarized as students' active engagement with class, monitoring of and reflecting on students' needs and learning, and emphasis on linkage with daily life and the community. Students' active engagement and the emphasis on linkage with daily life represent constructivist strategies that have been proved to be influential in science teaching (Brophy, 1992). Odom et al. (2007) support this point by stating that when the more engaged students are actively generating and testing hypotheses, there is greater understanding and a better attitude towards science.

3.3.4.4 Classroom climate

Classroom climate is the atmosphere developed in a dynamic relationship by teachers and students within their learning environments during the school year (Fraser, 1994). Such psychological environments as morale or climate of the classroom formed by a social group were considered important factors that influence student outcomes in a theory of educational productivity (Walberg, 1990). The empirical evidence was documented, as Haertel, Walberg and Haertel (1981) studied the secondary analysis to find correlations between student perceptions of the social-psychological environments of their classes and learning outcomes in eight subject areas, including science. Their results indicated that student learning achievement had a positive association with *cohesiveness, satisfaction, task difficulty, formality, goal direction, democracy,* and *the material environment* and a negative relationship to *friction, cliquishness, apathy*, and *disorganization*.

The classroom climate for middle-grade students in secondary schools seems more important than for other grade students. The findings of Fraser (1989) in the study of analysis of NAEP science assessment reveal that classroom climate during science lessons was shown to have a stronger impact on the science achievement of 13 year-old students than on that of 17- and 9-year old students. There is evidence that classroom climate has influenced student science achievement indirectly, mediated by instructional quality and instructional time (Reynolds & Walberg, 1991). The learning atmosphere, resulting from the interactions between a teacher and students, was found to persist beyond their classrooms, such as in visiting museums (Tran, 2007). Therefore, it is evident that a favourable climate works not only within the classroom but also outside it, for student learning.

Desirable student outcomes could be expected to emerge from a stable climate in the classroom, but in order to create this the management behaviour of the teacher must come into play (Creemers, 1994). *Teacher attitudes* may be one factor to indirectly contribute to this classroom climate, since *teachers' beliefs, perceptions* or *interests towards science, teaching science,* or *their students* influence teaching practice or strategies (Jita, 2004; Hattingh et al., 2007; Roehrig et al., 2007), and in turn teaching practice influences students' attitudes towards science as examined above. SER also identified that classroom climate is enhanced by orderly-management (Creemers, 1994; Scheerens & Bosker, 1997). It was found that teacher's strong leadership and provision of a degree of student responsibility are more likely to promote achievement, whereas a greater degree of strict behaviour by the teacher and emphasis on rules (regulation on acting in laboratory) and clarity in science laboratories are negatively related to student achievement (Henderson et al., 2000). This occurs because the former results in a well-organized and responsible involvement of the students, whereas the latter makes them withdraw and not get into trouble.

Teachers' attitudes, *students' attitudes and behaviour,* based on their social contexts, can contribute to classroom climate. The finding that the classes of high performance schools showed fewer *intrusions* and *disruptions*, which leads to more instructional or learning time, is well documented in research (Creemers, 1994). When Dumay and Dupriez (2007) examined the TIMSS 2003 data, they found a significant part of the between-class variance in mathematics could be explained by class climate, particularly the joint effect of students' composition and such class processes as teaching practice. In the comparison of the USA and five top-performing Asian countries in TIMSS 1999, Shen (2005) found that there was more absenteeism and frequent class interruption in American schools than in the Asian schools, and American parents and students valued schooling less than their Asian counterparts. Therefore, students as well as teachers play a role in generating a favourable atmosphere to learning, and thus at the classroom level students can contribute to their own achievement.

3.3.4.5 Physical resources at a classroom level

Science depends on physical resources that assist in understanding scientific knowledge and developing skills through hands-on activities (Rogan, 2000). In addition, physical resources are important, given that enhanced material strategies and instructional technology strategies are regarded as effective science teaching practice as explored above. Science-specific physical resources include *laboratory equipment* and *materials for science experiments, science instructional materials, audio-visual facilities, computer software, availability of computers,* and *internet access for science teaching*. It was documented that science equipment had a positive effect on science achievement in eight countries participating in SISS (Postlethwaite & Wiley, 1992). Physical resources such as technologies or devices may help students objectify the observed world and appropriate learning tools can improve science instruction (Tate, 2001). Essentially, teachers can improve their instructional quality when provided with the appropriate classroom resources combined with professional learning opportunities and support (Tate, 2001). Available resources in schools, including instructional materials, time for teachers to plan and prepare lessons, and availability of relevant science supplies, were reported as having a statistically significant impact, in particular on teachers' investigative practices (Supovitz &Turner, 2000). Therefore, when implementing science curriculum reform, physical resources were regarded as an important factor, together with factors of teacher and student, school ethos and management (Rogan & Grayson, 2003; Rogan & Aldous, 2005).

In particular, the availability of computers for teachers and students is becoming a vital resource in schools, reflecting the importance of preparation for a highly IT-centred society around the globe. It was documented that using such technology as computers fostered and encouraged students to engage in learning (Tal, Krajcik & Blumenfeld, 2006). However, there is a controversial issue about the effects of computer technology. As reviewed above, the way that technology is used in the classroom depends on teaching practice, hence, the availability of computers and access to the Internet should be considered from a perspective of educational resources in a different way from the one discussed in teaching strategies.

On the other hand, students from minorities, or of low SES, can benefit from practical work using instructional materials as mentioned above (Burkam et al., 1997). This could be attributable to limited access to various informal experiences and the material offered in the classroom being the only opportunity for them to experience science activities (de Feiter et al., 1995). It was found by Hattingh et al. (2007) that the less proficient the learners are in the instruction language, the higher the need for practical work. In particular, in countries such as South Africa, where many students study science in a language different from their mother tongue, teachers need to use practical work to compensate for poor verbal communication.

In contrast to the aforementioned benefits resulting from the presence of physical resources, it is argued that lack of such resources as science teaching facilities, laboratories and equipment, together with large class size, leads to students' view of science as memorization rather than problem-solving (Black, Atwaru-Okello, Kiwanuka, Serwadda, Birabi, Malinga, Biumigishu & Rodd, 1998). It was universally reported in TIMSS that shortages of resource and material had an adverse effect on science instruction (Mayer et al., 2000).

Class size was reported to having a significant impact on student learning in the classroom, although how many students should be in one classroom is manipulated by policy at the higher context level (Mayer et al., 2000). Considerable research has provided evidence that class size influences student achievement (Greenwald, Hedges & Laine, 1996; Blatchford, Russell, Bassett, Brown & Martin, 2007). Research shows that, in particular, younger, disadvantaged, and minority students learn better in smaller classes (Mosteller, 1995; Rice, 1999). From the secondary analysis of the National Education Longitudinal Study (NELS) of the US Department of Education data of Grade 8 students, Akerhielm (1995) reported that small class size had a positive influence on student achievement in certain subjects, including science.

It was reported that both students and teachers benefit from small class size (Blatchford & Mortimore, 1994; Blatchford et al., 2007), while Rice (1999) confirmed the above findings in a study examining the impact of class size on instructional practices, and the use of time in high school mathematics and science. From the perspective of students, it is easy to focus on and spend more time on the learning task, as more attention and teaching from the teacher encourages them to develop good attitudes towards their learning. As a consequence, small classes tend to lead to higher levels of engagement, which in turn results in higher student achievement (Finn & Achilles, 1990; Blatchford et al., 2007).

Unlike small class size, large class size can cause non-instructional use of time, such as conducting administrative tasks and maintaining order in the classroom. Therefore, more time resulting from less interruption allows the teachers more opportunities to use teaching materials, leading to broader and deeper curriculum cover and improved student confidence, knowledge, and skills in science. However, one thing should be borne in mind in terms of the benefits of a small class, that only when accompanying a change of teaching practice and support of qualified teachers will the effect of small classes have a positive impact on student achievement (Mayer et al., 2000).

As opposed to the positive contribution of resources to teaching practices, there are some negative findings about the use of resources. The presence of resources does not guarantee use of them, as shown in the Stark and Gray study (1999) where the low use of computers in secondary science was reported by pupils despite the highest number of computers per school in the TIMSS report. Hattingh et al. (2007) examined practical work in the teaching of natural science in the light of curriculum implementation in South Africa, where an outcomes-based curriculum was being taught. In a related study, Rogan and Aldous (2005) found no relationship between availability of resources and the level of practical work. Nonetheless, ironically, the most commonly reported problems in the conduct of laboratory work were related to poor conditions, insufficient equipment and an extended preparation time (Wilkinson & Ward, 1997). It is a general belief that availability of science facilities has a significant and direct effect on science experiments and thus on student achievement (George & Kaplan, 1998).

3.3.5 SCHOOL-LEVEL FACTORS INFLUENCING SCIENCE EDUCATION

More attention has been given to factors at the school level which influence student achievement than to classroom-level factors, because school-level factors are only alterable by policy or financial investment, although various factors at the school level tend to be inter-related and difficult to quantify.

Accordingly, they are likely to indirectly influence student learning, and to mediate through teachers and classrooms (Mayer et al., 2000), whereas teacher or classroom level attributes influence student learning directly. Factors at the school level reviewed here are curriculum management, professional teaching force, school climate, and resources. Since the school unit encompasses other subjects as well as science, the review is likely to be general rather than science-specific.

3.3.5.1 Curriculum management

Curriculum management involves the way schools work on *curriculum-related tasks or decisions,* such as *choosing textbooks, determining course content, course offerings, student grading policies, assigning teachers to science classes,* and *instructional days or hours per year*. The curriculum taught in the school may vary depending on which kind of textbook is chosen and used, although the decision of the intended curriculum is made at the context level. The number of instructional days in the school year was reported as having a positive correlation with the national mean achievement in science as well as mathematics in TIMSS (Martin et al., 1999). Instructional days are inextricably linked to ‘time on task’ or ‘opportunity to learn’, considering that more instructional days a year may offer students more time in their school and thereby more instructional time. Therefore, the policy of the number of hours per year devoted to science directly influence the instructional time for science.

The number of hours per year allocated to science education influence the implemented curriculum, particularly if science is taught as integrated or separate units. It was found that students who were being taught science as separated disciplines had more instructional time than those who are taught science as an integrated subject (Martin, Mullis, Gonzalez & Chrostowski, 2004). The content taught in each grade can also influence student learning, therefore the decisions on course content and offerings are important. Two important reasons for US students' poor performance in international comparative studies like TIMSS have emerged: one is a 'cafeteria-style' and diffuse science and mathematics curriculum, which means a lack of content focus; the other is a variation in topic coverage across classrooms (Mayer et al., 2000; Valverde & Schmidt, 2000). This is especially evident in countries with a decentralized curriculum, as in the USA. At school level, it is important to appropriately and consistently choose and arrange science courses or content to ensure that teachers do follow a standards-based curriculum.

3.3.5.2 Professional teaching force

The professional teaching force involves *educational leadership, consensus or cohesion among school staff including teachers,* and a *stable body of teachers*. Educational leadership by principals was consistently reported to be an effective factor of achievement (Edmonds, 1979; Mulford, 1988; Scheerens & Bosker, 1997; Tate, 2001). Although the core role in the professional teaching force is thought to be played by a principal, in reality principals, according to TIMSS findings, tend to manage administrative duties rather than instructional leadership activities, such as overseeing curriculum planning, training teachers, and working with teachers to develop educational objectives (Martin et al., 1999).

However, this is not always the case for all schools, public or private. There is evidence that public schools are different from private schools in terms of the structure of their governance. In the school district administration common to public systems, teachers tend to regard their principals as lower-level managers, while in private schools the principals tend to take more responsibilities and play the role of a leader (Mayer et al., 2000).

In addition, principals influence teaching and learning in schools differently across countries. Reynolds et al. (2002) studied SER across nine countries in an attempt to determine which school and teacher factors were effective in different countries, which were universal, and which specific to certain countries. Their findings indicated that in English-speaking countries, including the USA, the UK, Ireland (Republic of Ireland), Australia, and Canada, school effectiveness depends more on the leadership of a principal, whereas non-English-speaking societies including Hong Kong, Taiwan, the Netherlands, and Norway have, according to Reynolds et al., such a well-ordered and well-engineered educational system that individual leadership and the relationships among the staff members are less important than system variables. A similar finding, reported above, indicates that the leadership factor shows a positive effect on student achievement in the USA but this is not the case in the Netherlands (Creemers, 1994). It is worthy of attention that Singapore, the highest-performing country in three sequential TIMSS administrations, showing no gender difference and no expense of affect in their science achievement, places emphasis on the CEO-like systemic commitment towards a good school organization through the special leaders-in-education programme for potential school principals (Aun et al., 2006).

Apart from principals, school staff and teachers mould a professional teaching force as well. For instance, *regular meeting of teachers* may be effective in improving cohesion and collaboration among teachers. Teachers, staff, and a principal working collectively within a school can have a positive effect on student learning. It was found that teachers valued collegial support and team planning, and the support was most effective when coordinated by a science administrator through frequent meetings focused on student learning (Roehrig et al., 2007). The professional teaching force is likely to establish common goals, to focus cohesively on student learning, be willing to collaborate and be open to new ideas, all directed toward high student achievement. Cohesion among staff and teachers in a school can be translated into consistency, and in turn develop a more favourable atmosphere, yet it should be noted that without appropriate professional development and supporting resources, a shared vision and cohesion alone does not guarantee the successful implementation of the intended curriculum (Singh & Manser, 2000).

It was suggested that an *experienced* and *stable community* of teachers is more likely to be professional (Hanushek, Kain & Rivkin, 1998). Jita (1998) found in the study of the context of science education in a South African rural area that 84.4% of respondents were under the age of 39, reflecting a lack of veteran and experienced teachers in the secondary science classroom. Unstable employment contributed to the unstable teacher community in this context and, in addition, the high rate of teacher attrition was reported to decrease teacher morale (Howie, 1999).

3.3.5.3 School climate

Research has shown that an *orderly school atmosphere* and a *positive disciplinary climate*, coupled with other attributes of school, teacher, and classroom, are conducive to student learning (Good & Brophy, 1986; Mulford, 1988). In addition, culture of school that is acceptable seems to support effective schooling, resulting in school improvement (Creemers, 2002). A study by Scherman (2005) into school climate in secondary schools of South Africa identified five factors which could distinguish the sampled schools in terms of school climate, viz., *Interaction, Cohesion, Learning environment, Resources, and Violence*. Certainly, students benefit from a school climate that minimizes discipline problems and clearly encourages academic excellence. School discipline related to school climate includes *student disrespect for teachers, absenteeism, tardiness, bullying, fighting,* and *theft*.

The TIMSS data also shows that the less absenteeism the more stable the student body, and the fewer problems the higher the achievement (Martin et al., 1999). It was reported in the USA that offences such as student tardiness, fighting, suspensions, and arrests had a negative effect on student achievement in science, as well as mathematics, reading, and social studies in secondary schools (Mayer et al., 2000). In a comparison of the US and five Asian top-performing countries in TIMSS 1999, Shen (2005) identified the following differences: A *relatively shorter school year*, *a higher student body mobility*, more *absenteeism* and *frequent class interruptions*, *students spending more time watching TV, playing sports, and working on paid jobs*, *a higher percentage of students from single-parent families, on average, parents having a relatively higher educational background, a higher percentage of students with computers at home,* and *a lower percentage having their own desks*. American parents and students' undervaluing of schooling was attributed to all these variances, and thereby the lower achievement.

Problems that preclude an ethos or atmosphere conducive to academic achievement have been shown to be associated with students from lower SES backgrounds. Therefore the *type of community* in which schools reside has been shown to influence school climate and thereby science achievement (Howie et al., 2008). Teddlie and Stringfield (1993) contrasted low-SES schools with middle-SES schools and suggested creating boundaries to buffer the school from negative influences from the low SES community by increasing contact with a middle-SES community and encouraging parents with high educational expectations to exert pressure for school achievement. In contrast, high expectations from the school, community, and home were found to have a bearing on student achievement (Phillips, 1997). With the assumption that rural and urban schools do not share equitable resource availability, which may account for the variance of academic achievement between the two areas, Webster and Fisher (2000) examined the TIMSS of Australia. Their multilevel analysis failed to show a relationship between availability of resources and achievement in science and mathematics, but found a strong and negative effect of rural location on student science and mathematics achievement. In the Korean TIMSS results, the location of school was proved to be the most important factor behind the variance in science and mathematics between schools (Park & Park, 2006). In South Africa, from the results of TIMSS 2003, Reddy (2006) also compared rural areas with urban areas, finding the differences to be substantial, especially in terms of school resources. South African performance in science has been shown to be stratified, especially by race, despite the abolition of the racial division of education departments in 1994. Such regional variances appear around the world, e.g. in Latvia (Bagata, Geske & Kiselova, 2004), thus, the effects of *school location* should be considered in the study of educational effectiveness.

An achievement-oriented school can improve student learning, as shown in SER previously (Scheerens, 1992; Scheerens & Bosker, 1997), just as parents' high expectations contribute to high achievement. In particular, academic pressure emerging from high expectation was found to improve student achievement (Phillips, 1997).

3.3.5.4 Resources

Resources at the school level involve *building, grounds, gymnasia, library, heating/cooling and lighting, budget for science supplies, general instructional material,* and *budget-related resources like teacher salary* and *student-teacher ratio*. Fraser (1989) found that the science teaching budget per pupil was a significant predictor of science achievement in secondary schools rather than in primary schools in the USA. Although student-teacher ratio within a school does not translate into class size, it is thought to reflect the extent of supporting a school system and indirectly teaching and learning. The largest school-level influence on teachers' practices and classroom culture in the USA was reported to be school poverty (Supovitz & Turner, 2000).

Hanushek (1986) reviewed quantitative studies from a perspective of economics and reported that *school expenditures* including *teacher salary, expenditures per pupil, administrative inputs*, and *facilities* had no strong or systematic relationship with student performance in the USA. However, Hedges, Laine and Greenwald (1994) pointed out that Hanusheck's study used inappropriate statistical methods and poor data, and found the reverse, that is that budget spent on education had a positive bearing on student outcomes. This finding was confirmed by the replication of the previous study (Greenwald, Hedges & Laine, 1996), suggesting that the size of the effect was large enough to show a significant increase in achievement through financial investment.

3.4 CONCLUSION

In this chapter, the literature has been reviewed from two perspectives, viz., SER and science education. School effectiveness research (SER) has identified many effective factors that influence student achievement and explain the achievement variances between educational systems. In the process of the research field development, SER attempted to develop comprehensive education models that can explain educational system in terms of achievement. Researchers apply these models to school improvement projects. In addition, SER is inextricably linked with TER with a common goal to improve student achievement based on the process in the classroom.

As one of the models developed in SER, the Creemers' model offers in particular a view of the teaching-learning perspective. It was recommended to serve as a framework for an international comparative study to view the results of countries which differ from each other in terms of geography, culture, and the socio-economic situation (Kyriakides & Charalambous, 2005). However, most of the attempts have been made to explain school effectiveness using language or mathematics thus far (Kyriades et al., 2000; De Jong et al., 2004; Houtveen et al, 2004) in European countries, but few are in effectiveness of science education particularly in African or Asian countries. Considering these points mentioned above, the current research needs to adapt the Creemers model to reflect the context of developing countries and science education.

On the other hand, research has documented many factors influencing student achievement in science. Research shows such effective factors at the student level as aptitude, attitude, and the social context, such as ethnicity, gender, SES and language. At the classroom level, science curriculum, teacher background, teaching practice, classroom climate, and physical resources-related factors were identified from the literature. At the school level, curriculum management, professional teaching force, school climate, and resources-related factors were distinguished. In the following chapter, the model designed for the study is constructed, based on the factors reviewed in this chapter and some SER models.