UNDERSTANDING OF THE NATURE OF SCIENCE: A COMPARATIVE STUDY OF CANADIAN AND KOREAN STUDENTS

By

Hyeran Park

A Dissertation submitted in conformity with the requirements for the Degree of Doctor of Philosophy
Department of Curriculum, Teaching, and Learning
Ontario Institute for Studies in Education
University of Toronto

© Copyright by Hyeran Park 2012
UNDERSTANDING OF THE NATURE OF SCIENCE:
A COMPARATIVE STUDY OF CANADIAN AND KOREAN STUDENTS

Doctor of Philosophy 2012
Hyeran Park
Department of Curriculum, Teaching, and Learning
University of Toronto

ABSTRACT

This study was designed to identify students’ perceptions of learning activities, assessment formats, and content on their understanding of the nature of science (NOS) by comparing and examining constructs created by Canadian and Korean students. Participants were 217 Canadian and 319 Korean Grade 8 students that filled out questionnaires; additionally, 9 students volunteered for semi-structured interviews.

Descriptive statistics, multivariate analysis of variance and partial least squares were used to examine the quantitative data. A conceptually clustered matrix was used for the qualitative analyses. Results indicated that students from both countries perceived 1) their learning activities were teacher-directed, 2) class presentations and discussions occurred least frequently, 3) paper-and-pencil tests determined science scores, 4) science tests relied heavily on knowledge of science while knowledge about science was least likely to be assessed, and 5) generally students held relativistic views on science.
The effect for country on NOS concepts was statistically significant across all of their perceptions except for the concepts of culturally embedded science and the perceptions of short-answer test formats. Specifically, Canadian students perceived that they had relatively more student-directed activities while Korean students perceived that they had more teacher-directed science lab activities. Further, Canadian students were inclined to hold more relativistic views across the NOS concepts. It was also noted that Korean students provided more political examples while Canadian students provided stem cell research or environmental issues.

An examination of associations revealed that students’ learning activities, assessment formats, and content are good predictors of NOS understanding since these constructs explain variances from 19.7% for Empirical NOS to 63% for Scientific Methods. Results from students’ open-ended responses to the NOS concepts and the semi-structured interviews were consistent with the quantitative analyses. Most interviewees agreed that what, and how, they learned science-- and how their learning was assessed--affected their views of science since school science education was the important factor in developing their scientific knowledge.

These results imply that diverse learning activities and assessments could prove to be a better approach to enhancing students’ understanding of NOS than teacher-directed learning activities and test formats requiring a single correct answer.
ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Professor Earl Woodruff for accepting me as his doctoral student, and for his patient guidance over the years from beginning research, comprehensive examination, and proposal through to this dissertation. Without his help, this long journey could not be possible.

I would also like to express my appreciation to other members of my dissertation committee. Professor John Wallace has provided constructive feedback since the comprehensive examination. Professor Jim Hewitt encouraged me to move forward. Their guidance in preparing this dissertation and their support enabled me to reach my final oral defense. Professor Richard Wolfe helped me solve problems in quantitative data analysis. Professor Dr. Anthony Clarke served as an external member of my dissertation committee and provided a grateful appraisal. Dr. Douglas McDougall and Dr. Clare Brett supported my final oral examination and shared their intelligence as internal examiners. My thanks to them are beyond my words.

My gratitude must be expressed to the science teachers, the school principals and all the students who shared their experience, knowledge and ideas as they participated in the survey and interviews. I am very much grateful for their kind support. Their knowledge and experiences enriched this study.

I am thankful for so many others who helped my studies. Rev. John Bodner proofread my work and shared his wisdom. Dr. Zhang provided his insightful comments and guidance. Dr. Stagg-Peterson, Miss Karen, Ms. Ursino, Ms. Woo Hyunmi, Ms. Kim Okjoo, Ms. Park Kyeongae and Ms. Choi helped me to recruit the participants, and their supports were tremendous. Pei-Ying Lin and Youngsin provided warm support.
My deepest appreciation goes to my son, Joohyun Kim, and my daughter, Hyejoon Kim. They were my source of energy and ideas. Though young, their endless love encouraged me to complete my study and accepted patiently their mother’s necessary absences. My thanks go to my sister, Oksoon Park, and my mother-in-law, Heejah Park. They pray for me everyday. I also give thanks for my parents, Namsun Park and Wooboon Yoon, and for my husband Wonkook Kim, all now in heaven, whose love for me and pride in me have sustained me.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... ii
ACKNOWLEDGEMENTS ....................................................................................................................... iv
TABLE OF CONTENTS .......................................................................................................................... vi
LIST OF TABLES ........................................................................................................................................ xii
LIST OF FIGURES ...................................................................................................................................... xiv
LIST OF APPENDICES ............................................................................................................................ xvi

CHAPTER 1. INTRODUCTION ............................................................................................................... 1
  1.1. Overview ....................................................................................................................................... 1
  1.2 Problem Statements ....................................................................................................................... 3
    1.2.1 Incongruent Research Results .................................................................................................. 4
    1.2.2 Rare Research on the Relationships between Achievement Assessment and Students’ Understanding of NOS ............................................................................................................. 5
    1.2.3 Inadequately Assessed NOS ..................................................................................................... 7
    1.2.4 Psychological Resistance to Changing Concepts ........................................................................ 8
    1.2.5 Lack of cross-cultural studies .................................................................................................. 9
  1.3 The purpose of this research .......................................................................................................... 10
  1.4 Research Questions ...................................................................................................................... 11
  1.5 Significance of the Study ............................................................................................................... 12
  1.6 Outline of the Thesis ..................................................................................................................... 14

CHAPTER 2. LITERATURE REVIEW ON NOS ....................................................................................... 17
  2.1. Overview ....................................................................................................................................... 17
  2.2. Definition of NOS ....................................................................................................................... 17
    2.2.1 Rational for selection of the five concepts of NOS .................................................................. 20
    2.2.1 The Tentative Nature of Science ............................................................................................ 22
    2.2.2 The Empirical Evidence Based Scientific Knowledge ........................................................... 26
    2.2.3 Theory-laden scientific observation (Subjectivity) ................................................................. 28
    2.2.4 Social and Cultural Embeddedness ....................................................................................... 30
2.2.5. Diverse Scientific Research Methods .......................................................... 31
2.2.6. Section summary ......................................................................................... 40
2.3 Why is Students’ Understanding of NOS Important? ...................................... 40
2.4 Why NOS Should Not Be in Science Education? .............................................. 46
2.5 What Has Been Done So Far? .......................................................................... 48
  2.5.1 Measuring the concepts of NOS ................................................................. 49
  2.5.2 Science Textbooks and Students’ Understanding of NOS ......................... 53
  2.5.3 Teachers’ understanding of NOS and students’ understanding of NOS ....... 55
  2.5.5 Pedagogical approaches to NOS ................................................................. 56
2.6 Cross-country Studies ...................................................................................... 59
  2.6.1. NOS Studies in Canada and Korea ............................................................ 60
2.7 Summary ........................................................................................................... 63

CHAPTER 3 CLASSROOM LEARNING ENVIRONMENT, ACHIEVEMENT ASSESSMENT AND CONCEPTUAL FRAMEWORK .......................................................... 65
  3.1. Overview ....................................................................................................... 65
  3.2. Definition of Learning Environment ............................................................. 66
  3.3 Reviews of Learning Environment .................................................................. 67
  3.4 Instruments for Assessing Learning Environment .......................................... 70
  3.5 Learning Environment and NOS Studies ....................................................... 72
    3.5.1 Behaviouristic science education and research on NOS ......................... 72
    3.5.2 Constructivist Science Education and Research on NOS ....................... 74
  3.6 Review of Achievement Assessment .............................................................. 77
    3.6.1 Achievement Assessment Formats, Content and Learning Strategies ....... 77
    3.6.2 Large Scale and High-Stake Assessment in Teaching and Learning ........... 80
  3.7 Instruments of Measurement for Student Perceptions of Assessment ............. 82
  3.8 A Theoretical Model for This Study ............................................................... 85
  3.9. Chapter Summary ......................................................................................... 89

CHAPTER 4 METHODOLOGY .................................................................................... 90
  4.1 Introduction .................................................................................................... 90
4.2 Rationale for a Mixed Method and Research Design ................................................................. 91
4.3 Participants ................................................................................................................................. 95
  4.3.1 Canadian Participants ......................................................................................................... 96
  4.3.2 Korean Participants ......................................................................................................... 96
  4.3.3 Interviewees .................................................................................................................. 97
4.4 Research Context ..................................................................................................................... 98
  4.4.1 Canadian Context .......................................................................................................... 99
  4.4.2 Korean Context .............................................................................................................. 102
  4.4.3 Summary of Research Context ...................................................................................... 105
4.5 Data Collection ....................................................................................................................... 105
  4.5.1 Categories of the Instrument ......................................................................................... 105
  4.5.2 Emerged concerns .......................................................................................................... 106
  4.5.3 Measuring instrument .................................................................................................... 108
  4.5.4 Semi-structured interviews ............................................................................................ 115
  4.5.5 Other data resources ...................................................................................................... 116
4.6 Reliability and Validity ............................................................................................................. 117
  4.6.1 Assessing the NOS measurement models: Reflective models .................................... 117
  4.6.2 Assessing PLAT, PAFORMAT and PACONTENT: ....................................................... 119
  4.6.3 Structural model ............................................................................................................ 121
  4.6.4 Comparison between quantitative and qualitative data ................................................ 122
4.7 Data Analysis ............................................................................................................................ 122
  4.7.1 Quantitative data ............................................................................................................ 122
  4.7.2 Qualitative data .............................................................................................................. 130
  4.7.3 Integration of Quantitative and Qualitative Data .......................................................... 136
4.8 Ethical Consideration .............................................................................................................. 137
4.9 Summary ................................................................................................................................. 138

CHAPTER 5 RESULTS FOR THE SURVEY QUESTIONNAIRE .................................................. 140
  5.1 Introduction and Overview ..................................................................................................... 140
  5.2 Variable Codes ...................................................................................................................... 141
  5.3 Data Rearrangement and Exclusion ...................................................................................... 142
5.4 Descriptive Statistics........................................................................................................144
  5.4.1 Perceptions of Learning Activities and Tasks (PLAT)........................................144
  5.4.2 Perceptions on Assessment..................................................................................149
  5.4.3 Understanding of NOS......................................................................................157
  5.4.4 Summary of Descriptive Statistics....................................................................163
5.5. Validity and Reliability..............................................................................................164
  5.5.1 Reflective Measurement Model: Understanding of NOS concepts..............166
  5.5.2 Formative Measurement Model: PLAT, PAFORMAT and PACONT..............172
  5.5.3. Formative measurement model: Combined Sample......................................175
  5.5.4 Summary of Reliability and Validity of Measuring Instrument....................176
5.6 Differences..................................................................................................................178
  5.6.1 Multivariate Analysis of Variance (MANOVA): PLAT.................................178
  5.6.2 MANOVA: Differences between Canadian and Korean Students’ Perceptions on
  Science Assessment.................................................................................................182
  5.6.3 MANOVA: Differences in Students’ NOS Understanding.............................188
5.7 Associations................................................................................................................196
  5.7.1 PLS Analysis Overview....................................................................................196
  5.7.2 Canadian Sample...............................................................................................198
  5.7.3 Korean Sample..................................................................................................209
  5.7.4 Combined Sample..............................................................................................216
  5.7.7 Section Summary...............................................................................................221
5.8 Chapter Summary........................................................................................................222

CHAPTER 6 RESULTS OF THE QUALITATIVE DATA ANALYSIS..........................225
  6.1 Introduction and Overview......................................................................................225
  6.2 Responses to the Open-Ended NOS Questions.....................................................225
    6.2.1 Tentativeness....................................................................................................227
    6.2.2 Subjectivity (Theory-laden observations).......................................................238
    6.2.3 Empirical Evidence Based Scientific Knowledge..........................................248
    6.2.4 Social and Cultural Influences on Science......................................................256
    6.2.5 Scientific Methods............................................................................................268
6.2.6. Section Summary ..............................................................................................................275
6.3 Semi-Structured Interview Analyses ..................................................................................276
   6.3.1 Overview .......................................................................................................................276
   6.3.2 Learning Activities and Tasks .......................................................................................277
   6.3.3 Interviews on Test Formats and Test Content ..............................................................282
   6.3.4 Strategies to get high scores in science tests ...............................................................289
   6.3.5 Knowledge of Science and Knowledge about Science .................................................292
   6.3.6 PLAT, PAFORMAT and PACONT and NOS .................................................................295
   6.3.7 NOS ..............................................................................................................................299
6.4 Section summary ................................................................................................................304

CHAPTER 7 DISCUSSION AND CONCLUSIONS.................................................................306

   7.1 Overview of This Study .....................................................................................................306
   7.2 Answers to the Three Research Questions ........................................................................308
      7.2.1 Identification of Students’ PLAT, PAFORMAT, PACONT and Understanding of NOS ........................................................................................................308
      7.2.2 Differences and Similarities .......................................................................................312
      7.2.3 Associations .............................................................................................................315
   7.3 Discussion of Major Findings of This Study ....................................................................317
      7.3.1. Valid Instrument Development .................................................................................317
      7.3.2 Gaps between real and preferred learning activities ..................................................320
      7.3.3. Cognitive aspects of students’ perceptions on lab experiments ...............................321
      7.3.4 Korean students’ perceptions on lab activity assessment .........................................322
      7.3.5 Lack of time to ruminate ...........................................................................................323
      7.3.6 Homework related to the lack of time to ruminate ....................................................323
      7.3.7 PACONTphs in science tests .....................................................................................325
      7.3.8 Same test formats among different schools vs. different formats within the same school .....................................................................................................................327
      7.3.9 Informed or naïve views on Tentativeness .................................................................328
      7.3.10 The unique response from the religion based school about Subjectivity .................330
      7.3.11 Examples of historical anecdotes .............................................................................330
7.3.12 Discussions about School differences based on MANOVA results .................. 332
7.3.13 Teacher Effects on Students’ Understanding of NOS ................................. 335
7.3.14 Partly Mediated Effects of Assessment on Learning Activities .................... 335
7.4 Conclusions ............................................................................................................. 339
7.5 Recommendations for Future Research ............................................................... 340
7.6 Limitations of the Study ........................................................................................ 344

References ................................................................................................................... 347
LIST OF TABLES

Table 2-1 NOS Instruments .......................................................... 51
Table 5-1 Variable Code and Variable Description .................................. 141
Table 5-2 Frequencies, Means and Standard Deviations for PLAT Frequencies, Means and Standard Deviations ......................................................... 148
Table 5-3 Frequencies, Means and Standard Deviations for PAFORMAT & PACONT ...... 155
Table 5-4 Descriptive Statistics for NOS Concepts .................................................... 162
Table 5-5 Cross Loadings, Composite Reliability, AVE, Correlations (Canadian) .......... 168
Table 5-6 Cross Loadings, Composite Reliability, AVE and Correlations (Korean) .......... 170
Table 5-7 Cross Loadings, Composite Reliability, AVE, Correlations (Combined Sample) . 171
Table 5-8 Formative Measurement Models: Multicollinearity (Canadian Sample) .......... 174
Table 5-9 Formative Measurement Models: Multicollinearity (Korean Sample) .......... 175
Table 5-10 Formative Measurement Models: Multicollinearity (Combined Sample) ........ 176
Table 5-11 MANOVA Results for PLAT .......................................................... 180
Table 5-12 Tests of between Subject Effects (PLAT) .................................................... 182
Table 5-13 Tests of Between Subject Effects on PAFORMAT ............................................. 186
Table 5-14 Tests of Between Subjects Effects on Subjectivity .............................. 191
Table 5-15 Tests of Between Subjects Effects on Empirical Evidence Based NOS ........ 193
Table 5-16 Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Canadian) .................................................................................................................. 199
Table 5-17 Variable Notation ............................................................................. 201
Table 5-18 Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Korean) .................................................................................................. 210
Table 5-19 Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Combined Sample) ................................................................. 217
Table 5-20  The Effects of PLAT/PAFORMAT/PACONT on Understanding of NOS............222
Table 6-1  Summary of Common Themes on Tentativeness ........................................229
Table 6-2  Scientific Theory Changes and Error Inclusion and Approach to Absolute Truths
....................................................................................................................................237
Table 6-3  Summary of Common Themes on Subjectivity (Theory-Laden Scientific Research)
........................................................................................................................................240
Table 6-4  Summary of Common Themes on Evidence Based Scientific Knowledge ........250
Table 6-5  Summary of Common Themes on Empirical Evidence Based Scientific Knowledge
(Atomic Model) ................................................................................................................252
Table 6-6  Summary of Common Themes on Cultural, Political and Societal Embeddedness of
Scientific Research...........................................................................................................258
Table 6-7  Summary of Common Themes on Diverse Methods of Scientific Research .......271
Table 7-1  Results of Standard Multiple Regression to Predict PLAT (Y) from PAFORMAT
(X1) and PACONT (X2) ..................................................................................................337
LIST OF FIGURES

Figure 2-1. A Process for Knowledge Generation .......................................................... 39

Figure 3-1. Theoretical Model .......................................................................................... 88

Figure 4-1. Research Design and Process ................................................................. 93

Figure 4-2. Research Questions, Data and Analysis .................................................. 94

Figure 4-3. Items for PLAT ......................................................................................... 110

Figure 4-4. Items for PAFORMAT ............................................................................ 111

Figure 4-5. Items for PACONT ................................................................................. 112

Figure 4-6. Items for Tentativeness ......................................................................... 113

Figure 4-7. Formative Measurement Model .............................................................. 114

Figure 4-8. Reflective Measurement Model (Tentativeness) .................................... 115

Figure 4-9. Conceptually Clustered Matrix ................................................................. 136

Figure 5-1. Distribution of PLATEecture ................................................................. 145

Figure 5-2. Distribution of PLATquestion ................................................................. 145

Figure 5-3. Distribution of PLATlab ......................................................................... 146

Figure 5-4. Distribution of PLATmemo ................................................................. 147

Figure 5-5. Distribution of PLATpresentation ............................................................ 148

Figure 5-6. Distribution of PAFORMATclose .............................................................. 150

Figure 5-7. Distribution of PAFORMATshort ............................................................ 151

Figure 5-8. Distribution of PAFORMATalternative .................................................. 152

Figure 5-9. Distribution of PACONTknowledge ..................................................... 153

Figure 5-10. Distribution of PACONTapplication .................................................. 154

Figure 5-11. Distribution of PACONTLab ................................................................. 154

Figure 5-12. Distribution of PACONTphs ................................................................. 155
Figure 5-13. Distributions of NOS Concepts ................................................................. 161

Figure 5-14. Theoretical Model ................................................................................... 166

Figure 5-15. Means of Individual Schools in PLAT ......................................................... 181

Figure 5-16. Means of Schools in PAFORMAT ............................................................... 186

Figure 5-17. Means of Schools for PACONT ................................................................. 187

Figure 5-18. Means of Schools in Tentativeness ........................................................... 190

Figure 5-19. Means of Schools in Subjectivity .............................................................. 192

Figure 5-20. Means of Schools in Empirical Evidence Based Science ......................... 193

Figure 5-21. Means of Schools in Diverse Scientific Methods ...................................... 195

Figure 5-22. Theoretical Model (a) and Modified Model (b) (Canadian) ...................... 200

Figure 5-23. An Example of Student's Response .......................................................... 202

Figure 5-24. Means of Observed Variables (EMP1, EMP2, EMP3) (Canadian) .......... 207

Figure 5-25. Theoretical (a) and Modified Model (b) (Korean) .................................... 211

Figure 5-26. Theoretical Model (Combined Sample) ....................................................... 217

Figure 5-27. Means of EMP ......................................................................................... 219

Figure 6-1 A Model of Scientific Method ..................................................................... 269

Figure 7-1. Partly Mediated Causation Effect Between PACONT & PLAT ................. 336
LIST OF APPENDICES

Appendix A Survey...........................................................................................................................................375
Appendix B: Interview Protocol for Students ..................................................................................................386
Appendix C: A Letter to Principals ...................................................................................................................388
Appendix D: A Letter to Science Teachers ......................................................................................................391
Appendix E: A Letter to Parents and Students .................................................................................................394
Appendix F: Lab Report ......................................................................................................................................397
1. INTRODUCTION

1.1. Overview

Nature Of Science (NOS) covers the aim, development, criticism and explanation of science (Abd-El-Khalick, 2001; Hickey, 2005; Lederman, 1992; Matthews, 1994; McComas, 2004). Understanding the nature of science maintains the demarcation between science and pseudoscience (Popper, 1963), as it addresses science learning (Hodson, 1998, 2008; McComas, 2004), the cultural aspects of science (Chalmers, 1999; Kuhn, 1970), the enhancement public scientific literacy (Driver et al, 1996; Jeon, 2006) and pluralistic views of science (Siegel, 2002; Southerland, 2000). NOS, therefore, has become the central means to enhance the public's scientific literacy and to forward science curriculum reform in North America (America Association for the Advancement of Science (AAAS), 1989, 1993; Council of Ministers of Education Canada, 1998; National Academy of Science (NAS), 1996; Ontario Ministry of Education (Ontario MoE), 2007; Rudolph, 2003), Korea (Ministry of Education, 2001; Park & Cho, 1999) and around the world (Lederman, 2007; Turner, 2008).

The application of NOS to science education has raised a number of important issues for consideration. For instance, the typical science curriculum for K-12 students focuses too much on gaining knowledge of science; thus, such a curriculum fails to address knowledge about science (Duschl, 1990; Hodson, 1998; Lederman, 1992, 2006; Matthews, 1994; Park & Cho, 1999; Roth & Roychoudhury, 1994; Rudolph, 2003; Siegel, 2002). Also, Lederman’s comprehensive reviews on NOS studies (2007) reported that students’ and teachers’ understandings of NOS were unsatisfactory. Studies on Korean science teachers (Han, 1995; Han & Jeong, 1997) and students (Han & Jeong, 1997; Seoh, 1996) showed results similar to Lederman’s (2007). Consequent studies have yielded inconsistent interpretations of the factors,
which can affect students’ understanding of NOS. In addition, learning science and learning about science should take into account the differing cultural contexts of science education (Aikenhead, 2008; Liu & Lederman, 2007); however, not many have done such studies cross-culturally (Cobern, 1989; Griffiths & Barman, 1995; Liang et al., 2009). In consequence, only limited cross-cultural studies are available; they have identified some significant differences and argued the value of further studies (Cobern, 1989; Griffiths & Barman, 1995; Halai & McNicholl, 2004; Liang et al., 2009). These studies require more empirical evidence. Furthermore, very rarely have studies evaluated the influence of achievement assessment on students’ understanding of NOS although assessment forms a major part of education. Thus, no clear relationships have been established between science assessment and students’ understanding of NOS.

To address these issues, this study examines junior middle school Grade 8 students in Canada and in Korea for their understanding of NOS. Canada is a multicultural society, where the science curriculum is rather flexible while Korea is a highly mono-cultural society where a nationwide curriculum and standardized tests control all learning environments. Key NOS concepts addressed in the study are the tentativeness of scientific knowledge, scientific methods, the subjective nature of theory-laden observations, empirical evidence based scientific knowledge, and the social, cultural embeddedness of science. The primary aim is to identify students’ understanding of these specific concepts of NOS, noting significant differences or similarities between the two countries. Secondarily, the study investigates possible reasons for these students’ conceptions of NOS. This research should shed light on factors missed in previous research, such as students’ perceptions of their achievement assessments and of learning activities and tasks. A mixed method for data collection and data
analysis was employed. Following a survey, semi-structured interviews provided further in-depth understanding of results. This study should contribute to compensating the weakness of previous research by elucidating better forms of science achievement assessment, and by clarifying the effect of different educational environments for the student understanding of NOS.

1.2 Problem Statements

Since the last major science curriculum reform in the United States (AAAS, 1989, 1993, NRC, 1996), many countries including Canada (The Council of Minister of Education of Canada, 1998; Ontario MoE, 2007; Turner, 2008) and Korea (MoE of Korea, 2001; Jeon, 2006; Paik et al., 2005) have included NOS concepts in their K-12 science curricula. Efforts to identify the concepts of NOS and enlighten teachers’ and students’ understanding of NOS have been made by researchers, teacher educators and ministries of education. Still, many studies have revealed that students and teachers persistently maintain naïve views of NOS (Jung et al., 2005; Lederman, 1992, 2007; Lederman & O’Malley, 1990; Meichtry, 1992; Paik et al., 2005; Ryder, Leach, & Driver, 1999; Ryan & Aikenhead, 1992). Research reports have noted several possible impediments to enhancing the understanding of NOS: the science curriculum itself, teachers’ lack of NOS knowledge, and the lack of time and supporting resources. Also, studies have produced contrasting results and interpretations on issues such as the causal /non-causal relationship between students’ and teachers’ understanding of NOS (Han & Jeong, 1997; Lederman, 1986, 1992; Lederman & Zeidler, 1987), the role of NOS in an authentic science curriculum (Bell et al., 2003; Meichtry, 1992), explicit/implicit instructional approaches (Khishfe & Abd-El-Khalick, 2002), and the effects of studying the history of science (Abd-El-Khalick & Lederman, 2000).
Admitting that individual research contexts differ and that NOS concepts are complex, these differences in the basic assumptions of science education (teacher-student relationship, curriculum and students’ understanding of the subject domain) call for further reconsideration; what factors have the studies missed? What methodology did a given study adopt? How were relevant NOS concepts defined and measured? Therefore, this section states and discusses the following problematic points: 1) incongruent research results on factors influencing student understanding of NOS; 2) rarely conducted research on the influence of student achievement assessment on understanding NOS; 3) the measurement of NOS concepts; 4) psychological resistance to change established concepts, and 5) the lack of cross-cultural studies which reflect the effects of differing educational environments on students’ understanding of NOS.

1.2.1 Incongruent Research Results

Empirical studies report several possible factors affecting students’ understanding of NOS. Among the most probable and commonly mentioned are teachers’ understanding of NOS, the nature of the science curriculum and textbooks, and the approach to instruction. However, the different studies produce differing results, which unfortunately do not converge at all.

Many studies focus on teacher influence on learning of NOS (Choi, 2005; Han & Jeong, 1997; Lederman, 1986, 1999; Olafson & Schraw, 2006; Soh, 1998). The assumptions are that a teacher cannot teach what she/he does not know, and that a teacher’s understanding of a subject domain will directly affect her/his instructional behaviors and students’ understanding. Lederman (2007) concluded these assumptions are invalid based on his own (1986, 1999) and others’ research (Abd-El-Khalick et al., 1998; Bell et al., 2000; Hodson, 1993). Yet, a considerable number of studies (Mackay, 1971; Cho, 2001; Gallagher, 1991; Palmquist &
Finley, 1997) assert that teachers’ understandings of NOS do affect their classroom practices. Therefore, the assumptions of teacher effects on students’ understanding of NOS are inconsistent among previous studies.

Many blame conservative science textbooks as a major cause for students’ and teachers’ lack of NOS understanding (Johnson & Stewart, 1991; Kuhn, 1962, 1970; Latour, 1987). Conventional science textbooks largely describe scientific knowledge from a realist viewpoint and explain scientific methods with empiricist and inductive approaches (Cho & Park, 1994; Kuhn, 1970; Latour, 1987). Thus, trial science curricula have given more focus to NOS, such as Chem-Study in chemistry (Tamir, 1972), PSSC (Physical Science Study Committee) in physics (Tamir, 1972) and BSCS (Biological Science Curriculum Study) in biology (Jungwirth, 1971; Miechtry, 1992). Follow-up research on these curricula has reported that some aspects of NOS showed small but as yet inadequate improvement.

The alternatives of explicit or implicit instruction in teaching NOS concepts also generate controversy. Basically many studies agree that explicit instruction in NOS is more effective than an implicit approach (Abd-El-Khalick, 1998; Abd-El-Khalick & Lederman, 2000; Khishfe & Abd-El-Khalick, 2002; Schwartz & Lederman, 2002). Some, however, view this sort of direct instruction as another form of dispensing the knowledge of science rather than giving knowledge about science (Matkins et al., 2002; Moss, Abrams, & Robb, 2001).

1.2.2 Rare Research on the Relationships between Achievement Assessment and Students’ Understanding of NOS

These incongruent research results prompt a reconsideration of other factors crucial for students’ understanding of NOS. Educational goals, curriculum, teachers’ pedagogical philosophy or instructional approaches are too abstract for students to deal with in their
everyday tasks. Most students’ concerns are what content will be in the test, and how their learning will be evaluated. It is not an overstatement to say that achievement assessments (format, content, time, methods) are decisive for what, when and how students study. These assessments have a serious impact on their self-esteem and self-confidence (Dorman & Knightley, 2005; Scouller, 1996; Scouller & Prosser, 1994). Most NOS researchers have overlooked the effect of assessment on students’ understanding of NOS.

Reynolds, Doran, Allers and Agruso (1995) argue that effective learning occurs under the congruence between conditions of instruction, assessment and outcome. It is not hard to postulate a situation in which a science class may be taught the tentative nature of science and the subjective nature of scientific observation, yet at achievement assessments, the class is given a test measuring scientific laws and theories as enduring realities by demanding the memorization of a body of scientific knowledge. Such a test may not allow for more than two potentially correct answers. How can this approach possibly emancipate a student’s mind from naïve realistic views? How can such a test enforce pluralistic views of scientific knowledge? What is the point of teaching that scientific knowledge is tentative, when the test seems to demand hard-and-fast knowledge?

Unfortunately, very rarely have studies paid attention to students’ perception of their assessments, or to the formats of assessment in relation to students’ understanding of NOS. Lederman’s comprehensive reflection on NOS research (1992, 2007), Bell et al.’s (2001), The Nature of Science and Science Education: Bibliography, and academic journals like Science & Education, Science Education do not provide any studies of students’ achievement assessment in relation to their understanding of NOS.
1.2.3 Inadequately Assessed NOS

Measuring students' understanding of NOS remains an issue in NOS research (Aikenhead, 1987; Aikenhead, Fleming & Ryan, 1987; Miller et al., 2010; Lederman & O'Malley, 1990; Lederman, Wade, & Bell, 1998; Lederman, Abd-El-Khalick, Bell, Schwartz, 2002). One reason may be that NOS covers a broad range of disciplines from science to sociology, philosophy, and history of science. Lederman et al. (1998) pointed out that most evaluative instruments for measuring students’ understanding of NOS work with very narrowly defined concepts, or often address inappropriate areas such as science process skills and attitudes toward science, which in fact are beyond the scope of NOS. Since individual researchers have measured different aspects of NOS using a variety of instruments, comparison among studies becomes difficult, if not altogether meaningless (Lederman et al., 1998; Paik et al., 2005; Soh, 1998).

Another reason may lie with the formats of measuring instruments. Aikenhead (1987), Lederman et al. (1998), Abd-El-Khalick, Bell & Lederman (1998) and Lederman et al. (2002) raised validity issues on the standardized, forced-choice instruments assessing NOS because students misunderstood survey questions, and respondents had divergent reasons behind their answers. Also, the standardized instruments strongly reflected the bias of their inventors and failed to keep sufficient neutrality to be valid. Such invalid instruments could partially account for inconsistent results in NOS studies. Meanwhile, Miller et al. (2010) discuss the difficulties of analysis in open-ended responses in terms of unclear responses, heavy dependence on respondents’ writing skills and many unclassifiable responses. Details on this will follow in the literature review section.
1.2.4 Psychological Resistance to Changing Concepts

From the perspective of conceptual change in learning (Posner, Strike, Hewson, & Gertzog, 1982), students’ existing conceptions are very resistant to change and tend to influence their learning of new concepts. It is certainly difficult to conceive to any approach to teaching without taking into account students’ existing conceptions. This conceptual change could also be applied to the development of an understanding on the concepts of NOS. Before implementing any new curriculum/instruction intended to development students’ understanding of the NOS, science educators must understand what factors affect students in forming their concepts of NOS and explain the alternative frameworks involved with NOS. Some studies were hasty in reaching conclusions (Carey et al., 1989; Chun & Oliver, 2000; Khishfe & Lederman, 2006; Liu & Lederman, 2002; Sandoval & Millwood, 2005). They implemented a method for only one or two weeks, a couple of workshops or a term and then measured for changes in their participants’ NOS concepts. They did not reckon on students’ and teachers’ psychological resistance to changing their conceptions. As Chi and Roscoe (2002) pointed out, juveniles’ conceptions on science are vulnerable, but that does not mean conceptions formed through their whole formal educational life are easily altered within a short period of time. Miechtry (1992) and Miller et al., (2010) provide studies showing how difficult it really is to change one’s conceptions. There is no rule of thumb how long an implementation should last, but a short period of time giving explicit instruction which treats NOS concepts like factual knowledge is quite insufficient to discern whether, how or why students’ understanding of NOS may change.
1.2.5 Lack of cross-cultural studies

Not many cross-cultural and international studies have been conducted in relation to NOS in science education. Multiculturalists and relativists refuse to accept the traditional assumptions that science is universal. Instead, they argue that science has “contextual values” (Longino, 1990, p.4) which are partially the outcome of culture (Curd & Cover, 1998; Kuhn, 1970). Likewise science, both in learning science and in learning about science differing educational contexts in different cultures should be considered (Aikenhead, 2006, 2008; Aikenhead & Otsuji, 2000; Longino, 1990).

Before introducing the cultural aspects of science broadly, cross-cultural studies of NOS can reduce the risk of trial-and-error. However, not many studies verifying the cultural aspects of science education have examined the understanding of NOS among students from non-Western countries (Aikenhead & Otsuji, 2000; Dogan & Abd-El-Khalick, 2008; Kang, Scharmann, & Noh, 2004). Even fewer studies have been conducted across countries comparing science education in different cultures and students’ NOS (Cobern, 1989; Liang et al., 2009). In rare cases, studies have identified intriguing differences, suggesting the need for more work in this area (Cobern, 1989; Griffiths & Baraman, 1995; Halai & McNicholl, 2004; Liang et al., 2009).

These problem statements form the basic parameters for this study. No single effort can answer all these questions or solve the difficulties from previous studies. However, this study intends to shed light on some aspects of these issues by providing a baseline reference to the effects of differences in educational culture and the effects of different achievement assessments on students’ understanding of NOS concepts.
1.3 The purpose of this research

The aforementioned problems in previous NOS studies with the inconsistent results and interpretations have confused both teachers and reformers of science curricula. The purpose of this research is to clarify the effects of students’ perceptions of their learning activities and tasks in science classes and their perceptions of achievement assessment formats and content on their understanding of NOS, which previous studies have not fully examined. Also this study intends to compare the international and cross-cultural differences in understanding NOS concepts, which might distinguish a multicultural society like Canada from a mono-cultural society like Korea. The results and findings of this study can be used as a reference to reforming junior middle school science curricula, learning activities and tasks during science classes and to preparing formats and content for science assessments.

The participants were from Canadian and Korean Grade 8 students. Rather than implementing a special method that is not natural in typical science education, this research adopted natural settings; thereby, the comparison of these students across countries could reflect their realities and identify potentially effective factors for students’ understanding of NOS. With the help of diverse data sources, the study sought to show both differences and similarities between a Western country, Canada, and an Asian country, Korea; between multicultural and monocultural class settings; between less competitive learning and high-stake test-driven learning; between constructive-dominant and cognitive-dominant approaches to knowledge.

NOS studies have rarely examined the influence of achievement assessments on students’ understanding of NOS. This is most regrettable. Achievement assessment is crucial in classroom education; it pragmatically determines how, when, and what is taught and studied.
If a particular achievement assessment is a high-stakes test, students will adjust both their study strategies and their grasp of science to meet the demands of that assessment (Biggs, 2003; Scouller, 1998). Samples of achievement tests and their rubrics were collected for this purpose. As well, the survey and interviews examined the formats (open-ended/ close-ended, paper-and-pencil/ alternative), the contents, and student perceptions of achievement assessments. Acknowledging achievement assessment as an effective factor in science education, the study sought to the effects of assessment on the understanding of NOS.

1.4 Research Questions

Question 1 aims to establish participants’ perceptions of their learning activities and tasks, and of assessments and their understanding of NOS. Question 2 addresses differences and similarities in the three domains between two countries. Question 3 clarifies the relationships between the perceptions and the understandings and possible factors. These questions were answered by a survey questionnaire, students’ semi-structured interviews, samples of science tests and rubrics, and open sources from school board websites and Trends in International Mathematics and Science Study (TIMSS). The research questions are followings:

1. What are the junior middle school students’ perceptions on their learning activities and tasks, achievement assessment and the nature of science?

2. To what extent are the students' perceptions similar or different in the two contexts?

3. What are the relationships, if any, among students’ perceptions on their learning activities and tasks, the perceptions on their assessment formats and content and their understanding of NOS?
1.5 Significance of the Study

What contributions to the field of NOS can be expected from the study? It can provide a measuring instrument that is conceptually consistent among different domains in education; students’ perceptions on the learning activities and tasks, students’ perceptions on the assessment and the concepts of NOS. Certainly the research findings should expand on previous studies in measuring NOS concepts, and students’ perceptions on learning activities and tasks and assessments. They should also add to current international and cross-cultural studies, offer guidance to the practice of NOS in K-12 classrooms, and provide resources for planning policy on achievement assessment and curriculum design. A few comments on these matters are in order.

First, this research attempts to provide an instrument for three different domains of learning and teaching of science: 1) students’ perceptions on the learning activities and tasks, 2) students’ perceptions on the assessment formats and content and 3) the understanding of NOS concepts. No instrument has been developed to examine across these domains within one instrument. Particularly, NOS concepts are complex and are not easy to measure accurately (Lederman, 1992; Lederman et al., 2002; McComas et al., 1998). History shows many instruments have been invented, only to be buried under more criticism than compliment. All had drawbacks from which to learn. Some reflect heavily the author’s philosophy of science; some include other domains of NOS (e.g. students’ attitudes toward science), while yet others have constrained ‘true/false’ formats, which cannot provide reasons for the answers given. This research adopted a mixed format. Quantitative questions reduce blank answers and provide standardized answers. However, they can raise issues of misinterpretation (Aikenhead, Ryan, & Fleming, 1989; Lederman et al., 2002). VOSTS (Views On Science-Technology-
Society) (Aikenhead, Ryan & Desautels, 1989) employed a mixed format to get the reasons for student responses, providing both forced answering and alternative questions. The weakness of VOSTS was that it posed too many questions and its questions were particularly biased to Canadian culture. Its usefulness was therefore limited. Since this research is conducted across two countries, the instrument developed must be validated to function more or less neutrally in both Western and Oriental cultures. In this way the study’s results will help build sound research both in cross-cultural and NOS studies.

Second, this research considers students’ perceptions on their learning environment (classroom activities and tasks) and assessment as an important factor in their learning. Most NOS studies have focused on the actual components of the learning environment (e.g., explicit vs. implicit or curriculum content) rather than how students perceive them. Studies have revealed that students’ perceptions of their learning and learning environment are more important to their learning outcomes than the actual learning environments (Fraser, 1998; Dorman, Adams, & Ferguson, 2003). Since understanding of NOS is assumed as one of the outcomes of science education, clarifying the relationships of students’ perceptions of their learning science and their perceptions of assessment and the understanding is a meaningful inquiry.

Lastly, this study examines the effects of achievement assessment on students’ understanding of NOS, a question rarely addressed by previous research. Curriculum reform efforts have not produced fruitful results. Many science educators and scholars report that students and teachers do not reach a productive level of NOS understanding despite curriculum reforms and pre-and in-service teacher education (Lederman, 1992; Lederman, 2006; McComas et al., 1998). If results from this research show the relationships between students’
understanding of NOS and achievement assessment are significant, or that different countries produce different results and relationships, its findings will provide guidelines for the policy planning and classroom practice of assessment. That is, the study may demonstrate how large-scale assessment (assessment of learning) guides the classroom practice of achievement assessment (assessment for and assessment as learning) (Earl & Katz, 2006; Black & Williams, 1998).

Warner (2008) says that a study in the area of social studies cannot produce definitive evidence, but the accumulation of evidence makes for a sound, solid argument. Not many studies have sought evidence for the effects of classroom factors on students’ understanding of NOS, except in studying the influence of their teachers’ NOS understanding (Lederman, 1986, 1992; Martin-Dunlop, 2004). Martin-Dunlop (2004) pointed out that the studies of Lederman and Druger (1985) and Lederman and Zeidler (1987) are the only published research that investigated relationships between classroom variables and understanding of NOS. This study would furnish positive or negative evidence as to how learning activities, assessment and students’ understanding of NOS relate.

1.6 Outline of the Thesis

This thesis consists of seven chapters.

The current Chapter 1 provides an overview of the field, and outlines the purposes, research questions and significance of the study.

Chapter 2 is a review of the definition of NOS concepts and literature relating to NOS studies. It starts with a general overview of NOS concepts, noting both disagreements and common features among the differing schools of thought. Then, the NOS concepts crucial to this study are reviewed philosophically and historically. Next, a review of previous NOS studies
pays attention to details of measuring instruments, instructional approaches based on differing on educational theories, focusing factors, and international studies on NOS. The last section of Chapter 2 discusses and compares NOS studies relevant to Canada and Korea.

Chapter 3 examines previous research done on learning environments and their impact on students’ outcomes as well as on how prevalent learning perspectives have affected the measurement of NOS concepts. In addition, a review on the achievement assessments and students’ perceptions was followed. This review provides the rationale for the study’s focus on these two effects upon students’ understanding of NOS.

Chapter 4 explains the research methodology adopted for this research. A rationale for the use of mixed methods in research design opens the chapter. There follows a discussion describing the participants, the research contexts in typical Canadian and Korean junior middle schools and the development of the measuring instruments. Further sections are devoted to data collection and data analysis procedures. Account is given of how the validity and reliability of the instruments in this research were ensured. There follows a discussion of the conceptually clustered matrix for NOS concepts and of the guidelines for semi-structured interviews.

Chapter 5 includes the results of the quantitative data analyses. It examines how validity and reliability was established for the formative measurement models for the perception of science class learning tasks and activities, as well as for the perception of science assessment formats and content. A similar examination is made of the reflective measurement models of the five concepts of NOS over the Canadian, Korean and combined samples. Descriptive statistics then outline features of participants’ responses. Inferential statistics of
differences between two countries and among schools and of associations among variables
provide a profile of features for the sample population.

Chapter 6 illustrates the responses to open-ended questions and semi-structured
interviews concerning NOS concepts. In analyzing responses to open-ended questions,
attention was paid to common themes, clarification of close-ended responses and consistencies
between open-and close-ended responses and comparisons between Canadian and Korean
students. The analysis of semi-structured interviews was categorized by themes clarifying
connections between learning, assessment and NOS concepts.
Chapter 7 summarizes the thesis and discusses the notable findings, contributions and
implications of this study. Constraints and limitations to the study are noted, followed by
recommendations for further research.
2. LITERATURE REVIEW ON NOS

2.1. Overview

The complexity of science accounts for the difficulties of defining NOS (Lederman, 2007; Lederman et al., 2002; Matthews, 1994; McComas & Olson, 1998). Just as sociologists, philosophers of science, and scientists view the ontology and epistemology of science and scientific knowledge differently, so too with the concepts of NOS. Section 2.2 attempts to define the concepts of NOS in light of these discordant views. The common elements from all different views are suggested as a guideline for science educators in order to provide their students with appropriate and unbiased views of NOS. The following two sections presented both the case for teaching the concepts of NOS and the counter opinions. Section 2.5 categorized studies on students’ understanding of NOS according to the factors on which they focused. The last section reviewed international cross-cultural, Canadian and Korean NOS studies.

2.2. Definition of NOS

What is science and scientific knowledge? Are scientific laws and theories discovered from nature? Or, are they invented by scientists and their community? The discipline of Nature of Science (NOS) seeks to answer these questions. It deals with the epistemology of science, science as a way of knowing, or the values and beliefs inherent to the development of scientific knowledge (Lederman, 1992). McComs, Clough and Almazroa (1998) describe NOS as:

[A] fertile hybrid arena which blends aspects of various social studies of science including the history, sociology, and philosophy of science combined with research
from the cognitive sciences such as psychology into a rich description of what science is, how it works, how scientists operate as a social group and how society itself both directs and reacts to scientific endeavors. (p. 4)

These definitions show that NOS covers broad academic domains. The complexities of knowledge formation in science and its related ontological issues engender controversies among sociologists and philosophers in science (Kuhn, 1962; Lederman et al., 2002; Loving, 1991; McComas et al., 1998; Popper, 1963). The complexity and contradicting views were presented in Loving’s (1991) Scientific Theory Profile; four quadrants divided by an ontological continuum and an epistemological continuum. Many significant sociologists and philosophers of science are scattered through these four different quadrants; for instance, Thomas Kuhn as an antirealist-naturalist, Karl Popper as a realist-rationalist, Hempel as a realist-rationalist and Feyerabend as an antirealist-naturalist.

The National Science Education Standards (1996) or Benchmarks (AAAS, 1993) describe different views of NOS and provide confusing guidelines for science teachers. For instance, the Standards (NRC, 1996) describe science from Anti-realist and Realist viewpoints at the same time: “Scientific ideas are tentative and open to change” (p. 171). Yet, “Most scientific ideas are not likely to change greatly in the future” (p.171) “All scientific knowledge is subject to change.” (p.201). However, “The core ideas of science are unlikely to change” (p.201). The Benchmarks of AAAS (1993) show similar contradictions in defining NOS: “Results of similar scientific investigations seldom turn out exactly the same…Science investigations generally work the same in different places.” (p. 6) (bold type added).
Despite these disagreements, McComas and Olson (1998, pp. 6-7), the Delphi Study of the Expert Community (Osborne et al., 2003, p. 713), and Lederman (2007, pp. 833-834) suggest that the following common components of NOS should be included in science instruction:

- Scientific knowledge is durable and contains a tentative character.
- Scientific knowledge relies heavily, but not entirely, on observation, experimental evidence, rational arguments, and scepticism.
- There is no one way to do science.
- Laws and theories serve different roles in science.
- People from all cultures contribute to science with corporation and collaboration in the development of scientific knowledge.
- Observations are theory-laden.
- Scientists are to be creative questioning individuals.
- The history of science reveals both evolutionary and revolutionary phases in scientific inquiry.
- Science is a part of larger social and cultural traditions.

Based on these fundamental concepts, this research addressed five crucial aspects of NOS: 1) the tentativeness of scientific theories and laws, 2) the theory-laden subjectivity of science, 3) the empirical nature of scientific knowledge, 4) the social and cultural embeddedness of scientific knowledge, 5) and diverse scientific research methods.

The following sections were devoted to providing the rational of selecting the NOS concepts and identifying the meaning of these focused concepts including different schools of philosophers and historians. The explanation of individual characteristics of NOS included the
meanings, the origins of the concepts, the strength and weakness in explaining science and scientific knowledge, and the introduction of some salient scholars.

### 2.2.1 Rational for selection of the five concepts of NOS

This research has selected the above five NOS concepts for several reasons. The study tries to examine both ontological and epistemological perspectives of science. Additionally, because the research contexts involve two different cultures, particular attention is paid to student understandings of the socio-cultural aspects of science. Further, these particular concepts were felt to be more easily understood and less confusing at the participants’ age level. And these concepts have figured largely in both educational reform documents and NOS studies.

Multifaceted science and scientific knowledge have been classified into two broad constructs: 1) ontological realist or relativist perspectives and 2) epistemological perspectives such as the views of traditional empiricists, rationalists versus constructivist (Hickey, 2005; Godfrey-Smith; 2003; Loving, 1991; Matthews, 1994; Reichenbach, 1951). Although no clear borderline exists between ontological and epistemological facets of science, some NOS concepts more or less incline to one aspect or another. For instance, Paik et al. (2005) classified tentativeness and empirical testability as ontological, while theory-laden observation, social, cultural and political embedded scientific enterprises and diverse ways of scientific methods were classed as epistemological aspects of science and scientific knowledge. This research aimed to cover both ontological and epistemological aspects of science rather than one aspect.

Furthermore, this study deals with a comparison between a multicultural society, Canada, and a mono cultural society, Korea. Cultural aspects of science were involved, since
many scholars reviewed agreed cultural aspect of science (People from all cultures contribute to science with the corporation and collaboration in the development of scientific knowledge, and science is a part of larger social and cultural traditions). This study, therefore, takes an interest in students’ views on the cultural aspects of science and scientific knowledge and on the diversity of methods that approach the study of nature.

Previous studies revealed that some particular concepts were more controversial than others. This study selected less controversial concepts. For instance, for this reason the distinction between scientific laws and theories are excluded. The idea that scientific observations are theory-laden has been accepted among philosophers of science, while the distinction between scientific laws and theories has become obscure (Hickey, 2005; Trefil & Hazen, 2010). Trefil and Hazen (2010) define scientific theory as “a description of the world that covers a relatively large number of phenomena and has met many observational and experimental tests” (p.8, italic added) whereas Lederman et al. (2002) differentiate laws and theories as “Laws describe relationships, observed or perceived, of phenomena in nature. Theories are inferred explanations for natural phenomena and mechanisms for relationships among natural phenomena” (p.613, italic added). Also, almost all studies report that the concepts of scientific laws and scientific theories were confusing to students. They responded hierarchical relationships (Griffiths & Balman, 1995; Hong, 2001; Lederman et al., 2002).

Another controversial concept discussed in NOS is the assertion that scientific works are creative. Although most scholars agreed to the concept in general, some divide scientific works into a discovery context and a judgment context (Godfrey-Smith, 2003; Hickey, 2005; Popper, 1963). These two contexts of science are quite beyond Grade 8 students’ understanding of science. For these reasons, such confusing concepts were excluded.
The tentative nature of science and empirical evidence based scientific knowledge are included in this research because they are discussed as major themes in the *Standards* (NRC, 1996) and almost all studies of NOS (Abd-El-Khalick et al., 1998; Akerson, Morrison, & McDuffie, 2002; Hodson, 2008; Lederman et al., 2002; Matthews, 1994; Rubba & Anderson, 1978). Throughout the history of science, some long successful scientific theories (e.g., Geocentrism was a quite successful theory to describe heavenly objects) have been falsified by new evidence or by re-interpretation. Scientific knowledge is essentially reliable but changeable. Another essential entity of science is that its knowledge based on reproducible observations and experiments (Popper, 1963; Trefil & Hazen, 2010). It is one of the criteria that distinguish science from pseudoscience.

For these reasons, these five concepts of NOS were selected and examined. The following sections were devoted to identifying the meaning of these concepts including different schools of philosophers and historians. The explanation of these NOS concepts covered their meanings, origins, strength and weakness in explaining science and scientific knowledge, and introduced some of the most significant scholars related to them.

2.2.1 The Tentative Nature of Science

As the *Standards* (NRC, 1996) describe, scientific knowledge is durable and reliable, but it can never be proven as an absolute and final sense. It changes over time and can be appraised differently when more empirical data emerge, or when scientists reinterpret the data differently due to detecting and observing technology advances, cultural and social changes or dominate paradigm changes. This changeability is not limited to scientific hypotheses and theories but also encompasses facts and laws, which are assumed to be non-changeable. McComas et al. (1998) and Hodson (1998) point out a potent myth of science in science
education: people assume that scientific laws and ideas are absolute; they rarely appreciate that all knowledge in science is tentative. Ontological realists and anti-realists view this changeability in scientific knowledge differently.

Two major streams in the philosophy of science are from opposing ontological views of scientific knowledge. In one school of thought, scientific theories uncover knowledge that directly mirrors external reality; factual statements can be definitively evaluated for their truth status. In the other school, science creates theoretical terms and interpretations as instruments for understanding and explaining natural phenomena in the world (Boyd, 2002; Hickey, 2005; Loving, 1991; Matthews, 1994; Psillos, 1999; van Fraassen, 1980). In the history of science, especially during the eighteenth and nineteenth centuries, scientific theories posited a number of unobservable entities and processes, such as light corpuscles, lightwaves, the aluminiferous ether, molecules, atoms, various forces, etc. And using these entities, science was able to explain and predict many natural phenomena. Scientific realism has appealed to this empirical success of scientific theories to regard them as realistic interpretations. According to the realist conception, these postulates are attempts to characterize or represent nature’s reality (Boyd, 2002; Hickey, 2005; Psillas, 1999). If they are not exactly identical to reality, they are at least approximate truths supported by their empirical predictability (Chalmers, 1999; Hickey, 2005; Kasser, 2006; Godfrey-Smith, 2003; Psillas, 1999). According to Hickey’s explanation (2005), realists accept tentativeness in scientific theories because, as new and empirically superior theories are developed, their realistic interpretations produce a new ontology with new ideas and beliefs about what is real, including ideas of the nature of causality. When a new theory is suggested by a scientist, the community of scientists doubts the theory at the beginning; then,
as the theory accumulates supporting empirical evidence, it gradually gains acceptance and becomes an authorized, established reality.

Meanwhile, Popper (1963) has rejected the concept of the confirmation of scientific theory even though a theory may be successfully supported by empirical evidence. He argues that a scientific theory can never be confirmed; instead, if an experiment does not falsify a theory, then the theory becomes corroborated; it has not been falsified but stood up to severe testing. To those who side with Popper, scientific theories are vulnerable rather than durable. If a single item of counter evidence is detected, a theory should then be judged to be wrong. Even though science does not actually work in the way that Popper argues, his emphasis on falsification has strongly affected real science communities and philosophy of science in light of the value of bold conjectures and predictability in science rather than the value of science in explaining our nature.

Anti-realists largely base their views of the tentative nature of scientific knowledge such as the paradigm shift theory (Kuhn, 1962, 1970). Paradigm shift view is grounded in the history of science. History has revealed continuous changes in scientific theories. Kuhn’s studies in the history of science reveal many cases of discrepancy between how science should be done and how science has actually been done. He finds that the development of scientific knowledge follows neither inductivists’ nor of rationalists’ ways. According to Kuhn, paradigms form when significant numbers of scientists begin to work on related problems, based on the same theoretical assumptions. Scientists mostly work under a dominant paradigm, but when abnormal observations and results accumulate and threaten a paradigm of normal science, a new critical experiment brings scientific revolution and establishes a new paradigm. This new paradigm is not built on the existing paradigm but replaces it. For Kuhn, scientific
knowledge is not absolute but changeable, and its development is not continuous, but rather disconnected or incommensurable. A classic case of this is the shift from Newtonian to Relativistic perspectives.

Lakatos (1970) suggests a model to explain the development of scientific knowledge. His Scientific Research Programmes consist of two parts: Hard Core and Protective Belt. The Hard Core includes well-developed theories, principles and the domain-related laws, which are generally durable, and do not change easily. The Protective Belt surrounding the Hard Core is likely to be changeable consisting of auxiliary hypotheses, assumptions and possible inquiries. Scientific Research Programmes mark the mid-point in scientific knowledge between changeability and durability.

Bas Van Fraassen’s pragmatic explanation of scientific theories (1977) suggests that explanations are highly contextualized rather than describe an objective feature that one can have good reasons to believe that the theory really explains the observable natural phenomena. His assertion on scientific theories was that although a theory, T, explains an explanation on a fact E, it does not mean the theory T is true or empirically adequate. He concludes no absolute adequate explanation exists but context determines what factors count as salient and can serve to rule out the alternative possibilities in the contrast class. His argument was scientific theories are to answer a why question on a natural phenomenon, and explaining and answering the question is a matter of salient factors that are highly context sensitive not absolute.

As reviewed above, scientific knowledge is tentative not deterministic. For realists, scientific knowledge represents the best description and explanation of the physical world. However, they do not insist that it be seen as unchangeable or absolutely true. They admit to the tentative character of scientific knowledge in ways different from the views of anti-realists.
Realists respect the predictability of a theory, and view its postulated entities as a possible reality. Anti-realists regard scientific theories as an invention or consensus among scientists and their communities. For anti-realists, empirical evidence itself is a theory-driven observation. Changing a theory means, for realists, finding more developed ways to describe reality, while for anti-realists a new theory is merely another consensus on how to explain natural phenomena. Anti-realists do not think that the postulated theoretical terms of a theory are necessarily real. Therefore, it cannot be problematic to advance more than one theory for a phenomenon. Anti-realists thus find the tentative nature of scientific knowledge to be very natural.

**2.2.2 The Empirical Evidence Based Scientific Knowledge**

Scientific knowledge is held to be capable of being tested empirically and to require the support of empirical evidence (AAAS, 1990; Lederman, 1992; Lederman et al., 2002; Rubba & Anderson, 1978; Ruse, 1982). This is one of the major criteria that demarcate science from pseudoscience (Popper, 1963; Ruse, 1982). AAAS (1990) also supports the empirical testability of scientific theories. It emphasizes observation and empirical evidence in the natural world. It insists on empirically adequate evidence as the most important criteria in evaluating scientific theory. According to Lederman et al. (2002) and Chalmers (1999), however, while repeated testing against accepted observations establishes a scientific theory as valid, empirical evidence is insufficient to establish the theory as truth, because evidence itself is under the control of the related theories.

Empirical evidence is the standards by which a theory is confirmed or disconfirmed (Curd & Cover, 1998). If there are two or more rival theories on a phenomenon which are not compatible, the judgment and acceptance of any among them must rest solely upon empirical
evidence, not upon the fame of a scientist nor yet upon society’s values (Hickey, 2005; Curd & Cover, 1998). Kuhn (1970) also says that in the disciplinary matrix, a theory should be in a good agreement with the results of observations and experiments. Empirical accuracy is the first criterion; such accuracy should be based on objective and rational judgments. For instance, when Albert Einstein suggested the general theory of relativity in 1915, he was not a famous physicist at all. The general theory of relativity was originally advanced to account for the anomalous precession of the perihelion of Mercury (Holton, 1978). After the observation of solar eclipse by Eddington in 1919, the theory was accepted and then, Einstein became famous. The reverse does not occur, though meta-physicists argue that scientific theories are subject to the “Matthew effect” (the rich get richer, the poor get poorer) (Hickey, 2005; Merton, 1988; Strevens, 2006). Popper (1963) says a theory that is a bold conjecture, with high possibility to be falsified is a good scientific theory. The capacity of a theory for accurate prediction has value for its acceptance of a scientific theory upon objective, rational judgment. Although the measures of accuracy, objectiveness and rational judgment are under the control of a theory, neither realists nor anti-realists can accept a theory as scientifically valid without the procedures of being proven by empirical evidence.

Empirical evidence and accuracy are critical not only for scientists but also for metaphysicists. Carnap (1956), a logical empiricist, states that observable (physically sensible) entities are true, and if theoretical terms in a theory are able to be translated into observable terms, the theory can be accepted as truth, that is, as a confirmed sensible reality. While Carnap’s criticism (judgment) of scientific theory inclines to a phenomenal description, Duhem (cited in Hickey, 2005), self-identified as a positivist, maintains that an ability of a valid theory to make correct predictions with a sufficient degree of approximation is the range
of indeterminacy produced by a degree of measurement error. Duhem expands the ranges of empirical acceptability for a scientific theory beyond sensible categories. Duhem rejects the atomic theory because no mechanical atomic theory has been found to be sufficiently accurate in his time. Due to his emphasis on empirical evidence for a physical theory, he denies the explanatory function of a theory in science and regards such explanatory functions as metaphysics.

Since empirical evidence is taken as a major criterion for the criticism of science, many debates on theoretical terms have continued to treat them as real entities or instruments to explain natural phenomena. When one regards the theoretical terms as instruments, then they do not have to be true; they are classed only as convenient or inconvenient (Duhem, cited in Hickey, 2005; Psillos, 1996). However, if an instrument is not convenient, it cannot properly predict what will happen in the future. The empirical nature of science is essential to understand how science works.

**2.2.3 Theory-laden scientific observation (Subjectivity)**

To speak of theory-laden observation is to recognize that there is no theory-neutral observation; “What scientists observe depends on the theories they accept” (Curd & Cover, 1998, p. 220). The empiricist school of epistemology claims that science is objective and independent of the observer’s previous knowledge, psychological and social milieu (Chalmers, 1999; Curd & Cover, 1998; Hickey, 2005; Kasser, 2006). For logical empiricists, the accumulation of unbiased observation lays the foundation for a general description (a scientific theory) and leads the way to an approach toward absolute truth. Popper (1963) partially endorses the notion of theory-free observation when a radical change of theory occurs because past experiences or theories cannot guide scientists to modify the anomalies; rather, objectivity,
rationality and elimination of subjectivity lead to a new theory. However, relativists argue that a theory-free observation is simply not possible; if it were actually possible, it could not construct a productive scientific theory (Chalmers, 1999; Kuhn, 1977; Curd & Cover, 1998). Scientific theories are not conjectured in a vacuum but from existing background theories.

In the late 19th and the early 20th centuries, an important distinction was made between observational and theoretical entities in the epistemic development was in order to verify the existence of theoretical entities such as atom, electrons, magnetic electronic fields, etc. The distinction, however, is neither absolute nor sharp. In addition, it is largely based on pragmatic convenience and lacks epistemic significance (Hickey, 2005; Psillos, 1995). Duhem maintains, “an experiment in physics is not simply an observation of a phenomenon. It is besides, the theoretical interpretation of the phenomenon” (1906, p. 144, cited in Hickey, 2005). In other words, observation in science is not merely reporting a phenomenon; instead, it is interpreting a phenomenon on the basis of related theories. In discussing this theory-ladenness of observation, Kuhn (1970), admitting the argument of theory-laden observations, insists on semantic incommensurability among different paradigms. He claims that two paradigms are incommensurable because there is lack of common language, and of background theories. As far as scientists are under the control of language, and languages include semantics, observation cannot be free from theories.

While admitting to some changes of meaning along with changes in a relevant paradigm, it must be said in critique of Kuhn’s position that the variance of reference involved is not so huge that different paradigms cannot be correlated, insofar as they share a common reference point in nature itself (Curd & Cover, 1998; Psillos, 1995). Some aspects in the meaning of a term may be changed, but not all theoretical changes create entire changes of
meaning and reference. To illustrate, Copernicus and Ptolemy both dealt with the sun in their respective accounts of the universe. There are not two different suns in the sky! Conceptions of the sun differ between the two systems, yet proponents of either system can communicate with each other. If, as Kuhn’s view of incommensurability insisted, theory totally controlled the meaning of every observational term, then the meaning of all terms would change altogether with every change of theory. This simply does not occur in the history of science.

Semantic holism (Quine, 1953) advocates that the meaning of a language is not and cannot be isolated, but should rather be understood within a network of laws and theories; though, all observation is theory-laden, it can lead to the mistaken conclusion that the entire meaning of observational terms is also specified in a holistic way.

It is true that what we see is dependent on what we know. Also, we have been given no reason to think that what we perceive must always agree with what we believe (Curd & Cover, 1998, p. 222). Einstein (1925-1926), Heisenberg (1971), Feynman (1985), outstanding physicists, argue that neither 100% theory-independent observation, nor 100% theory-dependent observation really exists.

2.2.4 Social and Cultural Embeddedness

Science is a human enterprise; it cannot avoid its societal, political power and cultural historical milieu. Kuhn (1962, 1970, 1977) and Longino (1990) stress that science like scientists is one of the products of culture. Philosophical multiculturalism (Haack, 1998; Matthews, 1994; Siegel, 2002), which is ontologically on the same side as relativism, maintains that knowledge is constructed, valued and accepted and rejected in particular social, cultural and temporal contexts, and often depends on individuals’ psychological makeup. Multiculturalists accept that people have many ways of constructing reality. They disagree
with the ideas of realists that truth is “out-there” to be discovered; instead, reality is apprehensible to society and individuals through multiple, intangible mental constructions.

In contrast, universalism’s basic thesis is that scientific knowledge is trans-historical and trans-culture. Universalists, empirical positivists (Curd & Cover, 1998; Hickey, 2005; Matthews, 1994; Siegel, 2002), insist that there is permanent framework in science, which determines the nature of rationality, knowledge, and truth. For them, scientific knowledge is the result of objective outcomes free from the bias of the scientists or social and cultural influences, so that this knowledge is universally valid for all time and places and for all people. Because the nature exists completely independent human beings, valid scientific knowledge should be the discovery of reality.

2.2.5. Diverse Scientific Research Methods

Science is a study of the physical properties of nature. What differentiates science as a way of knowing from other subject domains, like literature, art, or theology? Many believe that the ways by which science learns to describe nature correctly and comprehensively are different from others. One of the standards by which to judge whether a theory to is actually science or merely pseudoscience is the process of study by which it is developed and pursued (Popper, 1963; Rubba & Anderson, 1978; Ruse, 1982).

Philosophers of science have classified various scientific methods as well as the way they view scientific knowledge, into inductivism, deductivism, hypothetico-deductivism, and relativism (Kasser, 2006; Matthews, 1994; Paik et al., 2005; Psilla, 1995). Following Kasser (2006) and Paik et al.’s (2005) classification, inductivism includes observational-inductivism, empiricism, positivism and universalism. Hypothetico-deductivism can embrace rationalism, deductivism, falsificationism, and logical empiricism. Relativism comprises
relativism, instrumentalism, multiculturalism and constructivism. None of the scientific methods listed above claims to develop infallible knowledge (AAAS, 1993; Bauer, 1994; Feyerabend, 1975; NRC, 1996; Shapin, 1996). Among these different methods, there are some similarities in approach to describing the nature; the major differences concern how to obtain knowledge.

In this section, these different forms of scientific method will be introduced. Then, current views on scientific method in NOS will be discussed, particularly in their critiques of traditional scientific method.

**Inductivism**: Inductive epistemology can be described as theory-independent observation, the accumulation of observations, and the generalization of observations into a theory that no individual observation contradicts (Kasser, 2006; Hickey, 2005; Lawson, 2005; Markie, 2004). Rigorous observations of natural phenomena distinguish science from other disciplines. Inductively strong means that arguments (theories) are that if the premises (observations) provide good evidence for the conclusion (general statements) then the arguments (theories) are said to be inductively strong. To give an example: an infinite numbers of observations of white swans create the general statement, “All swans are white.” The statement, “All swans are white” is supported by the observations, and then the statement (theory) is an inductively strong. In order to have a predictive ability of the general statement, the inductive inference needs to be based on the assumption that future experiences will be similar to the past; universe is uniform. As a swan was white yesterday, so other swans will be white in the future. This inference generates a general statement, “All swans are white.” The past informs the future.
Inductivists are too permissive to the observational results while too restrictive the theoretical entities in science. They believe that a universal statement is truth because it is known by experience (Chalmers, 1999; Hickey, 2005). Ernst Mach insists that science should not ask “Why?” but “How?” since the “why?” questions are beyond experiences (1893, cited in Psillos, 1995). For Mach, the “why?” questions and answers are regarded as meta-physics arena. Science postulated many unobservable entities, which were beyond tangible experiences and predicted what would happen within error ranges. Inductivists (empiricists) could not deny the accountability of the postulated theoretical entities. Carnap’s (1966) two-tier model is an example of giving observational language to theoretical entities, so that inductivists could admit the existence of theoretical entities. The dilemma of these efforts was that the interpretation of theoretical terms into observation language lost simplicity; the interpretation itself needed another criterion and then the criterion needed others.

Inductivist logic faces problems; how can we predict or justify that which we have not experienced? How can we know the universe is uniform? How many individual observations verify a universal statement? How do scientists deal with theory-laden observations? Inductive epistemology does not provide clear answers to these questions (Cobern & Loving, 2008; Hodson, 1998; Hoyningen-Huene, 2000; Popper, 1963).

Supporters of inductive inference in developing scientific knowledge argue that the logic of induction “determines the truth of scientific theories. To eliminate it from science would mean nothing less than to deprive science of the power to decide [it would no] longer have the right to distinguish its theories from the fanciful and arbitrary creations of the poet’s mind” (Reichenbach, 1930, p. 186, cited in Popper, 1959). Also, Skyrms (2000), supporting inductive methods, argues:
If an argument is inductively strong its conclusion makes factual claims that go beyond the factual information given in the premises. ...An inductively strong argument risks more than a deductively valid one, it risks the probability of leading from true premises to a false conclusion. But this risk is the price that must be paid for the advantage which inductively strong arguments have over deductively valid ones; the possibility of discovery and prediction of new facts on the basis of old ones. (pp. 8-9)

However, for Hume, induction is actually an irrational and indispensable psychological habit (Kasser, 2006; Psillas, 1995). Popper argues that induction is not only an irrational habit, but also a myth in the epistemology of science, because it does not tell us new theories, and because actual science never occurs the ways inductivists argue; scientists generate hypotheses without a specific experience, and the past generally does not inform new and bold conjectures.

**Deductivism** Deductivism is based on the idea that certain knowledge cannot be based on observations (experiences) but must be deduced from clear and self-evident truths independently from sense experiences (Lawson, 2005; Markie, 2004; Reichenbach, 1951; Skyrms, 2000). Rather than accumulating experience leading to knowledge, scientific theories are deductively tested against data (Kasser, 2006, Markie, 2004; Popper, 1963). It is useful to accept the theoretical entities in a theory, which are problematic for inductivism and empiricism.

While inductive epistemology is drawn *a posteriori* from empirical observations, deductive epistemology is *a priori*. Empirical evidence serves to confirm a theory as a non-falsified and valid theory (Hickey, 2005; Skyrms, 2000). Deductivists adopt nontruth-functional logic in its *a priori* validation procedure to avoid invalid tests. In other words, if a
true statement is drawn, but the premise is false, the test cannot be valid (Hickey, 2005; Skyrms, 2000). To illustrate: all A's are B. If X is an A, then X is B. Here, if premise A’s value is not established, then the tests are not valid (Hickey, 2005, p. 38). The logic applied to the case of swans runs: “All swans are white. If X is a swan, then X is white.” An inductivist would have to know whether the bird is a swan, and have to know the color of all swans, before concluding the bird’s color. A true statement (the color of the bird is white) with a false premise (the bird is not a swan) is therefore an invalid test. In contrast to the inductive arguments, a deductivist would not need to know that all other swans are white before verifying the truth of a statement. This point is the benefit of deductive epistemology in developing knowledge from accepted theoretical terms.

In the confirmation procedures for a scientific theory, the theory should be testable under specific conditions, and scientists should attempt to avoid ad-hoc reasoning in order to make a truth of a false statement. Popper (1963) insists that severe tests cannot confirm but "unfalsify" scientific theories. When such a theory stands up under a severe test, it becomes corroborated.

**Falsificationism** According to Popper (1963), the procedures of falsification ensure that scientific knowledge is unbiased and no dogmatic. He accepts theory-laden observation and partially theory-independent observation; he holds that science has been developed through trial-and-error, conjecture and refutation. In light of this view, science is the outcome of falsifiable hypotheses and includes the empirical testing of those hypotheses. The problem with the falsification process is that the actual verification of a scientific hypothesis is impossible. That is because the development of new measurements or new theories always can possibly falsify existing hypotheses (Hoyningen-Huene, 2000; Maxwell, 2002). While Popper
claims only one contradictory observation can falsify a scientific hypothesis or theory, Kuhn (1970) observes that normal science is very conservative, and that abnormal results are often ignored in the history of science. A theory or hypothesis stays in place along with any contradictions until a more applicable scientific hypothesis or theory appears.

Popper values bold conjecture and refutation in scientific theory over its explanatory power (1963). He points out that the theories of Alfred Adler, Max and Freud act as dogmatic systems, full of explanatory power, but only for believers. From his study of Einstein’s theory of relativity in gravitation, Popper concludes that the proper critical attitude toward a theory is to look for a crucial test that can refute the theory rather than searching for verification. Rather than confirming a theory, a crucial test corroborates a theory. Popper’s yardstick of a theory’s falsifiability demarcates science from pseudoscience. If a theory can be falsified by new discovery, it is scientific. He demonstrates this by instancing a statement that cannot be falsified empirically: “The universe is governed by God or by love and hate.” Because this hypothesis cannot be refuted empirically, it cannot be a scientific hypothesis. The main empirical tests for scientific hypotheses involve logical deduction; a scientific hypothesis stands as long as it is not falsified.

A problem with falsificationism is that it overlooks the influence of auxiliary hypotheses while a scientific theory develops (Maxwell, 2002). Auxiliary hypotheses include the conditions of observation and the apparatus of experiments. Therefore, if a theory cannot predict properly a targeted natural phenomenon, it is not clear whether the theory itself is wrong or whether one of the auxiliary hypotheses is wrong. Popper does not suggest a clear explanation in this case. Popper strongly rejects ad hoc reasoning, so he insists that any auxiliary hypothesis can be tested in an isolated situation. However, Quine (1953) insists on
the holistic examination of a theory. Statements do not have empirical significance in isolation. For him the distinction between deductive *a priori* and inductive *a posteriori* is not valuable since all beliefs are contingent. Furthermore, any theory can be revised, if it has been underdetermined by data and observation. He therefore advocates the openness of science toward all possible data for rival theories no matter how much evidence comes in. A piece of counter evidence may occur because of one of many postulated auxiliary hypotheses.

**Relativism** Relativists argue that there is no absolute truth in the world, and that *correctness* is decided by standards or values rather than absoluteness (Hickey, 2005; Morrison, 1990; Psillos, 1995). These standards vary from individuals to historical milieu and culture (Curd & Cover, 1998; Hickey, 2005). Kuhn (1977) admits to certain general criteria in choosing a scientific theory, such as accuracy, simplicity, breadth of scope, fruitfulness, and consistency with related theories. However, he rejects the thesis that scientific theory is value-neutral.

Acknowledging contextual values in science, he notes that standards of correctness have been affected by individual scientists’ backgrounds, society and power; thus, there are no objective criteria. Longino (1990) was similar to Kuhn’s view and defends the view that all judgments of scientific theory have in some degree been affected by contextual values.

Feyerabend (1975), a radical naturalist, asserts that there are no universal, immutable methods that all sciences should follow. Like Kuhn, Feyerabend bases his argument on the history of science. In his book *Against Method*, he declares scientists should follow their own ways, “anything goes” (p.28). Good science has always involved a lot of faith, chaos, play, serendipity and downright irrationality. He criticizes any predetermined method (inductive, deductive methods) that neglects the complexity of nature and makes “science less adaptable and more dogmatic” (p. 295). Scientific knowledge and methods have changed, and patterns of
changes are not consistent rather irregular and unpredictable. Thus, according to naturalist epistemology, science is not different and has no special status distinct from other disciplines. Partially supporting Feyerabend’s non-method, Chalmers (1999) asks, “If we have a conception of science as an open-ended quest to improve our knowledge, then why cannot there be room for us to improve our methods and adapt and refine our standards in the light of what we learn?” (p.161) (italic original). Yet, he also criticizes holding no method at all: “A middle way would hold that there are methods and standards in science, but that they can vary from science to science and can, within a science, be changed and changed for the better” (p.162). While a rationalist’s objective science describes the world and universe as a network of abstract logical relations and rejects intersubjectivity in science, a naturalist tries to connect science with man, and human subjectivity (Hoyningen-Huene, 2000).

Hoyningen-Huene (2000) is skeptical toward radical naturalistic approaches based on the history of science. He says that historical cases can be interpreted in various different ways, so naturalists can interpret history in favor of their ways. Mendel’s law, for example, is interpreted as a case of inductive inference for inductivists, while it is interpreted as a case of deductive method for deductivists; inductivists claim that observation of the number of round and wrinkled peas led to Mendel’s law. On the contrary, deductivists insist that the actual ratio was not exactly 3:1. Mendel needed a hypothesis, and experimented under a controlled condition.

While conceding to the naturalist argument that science may not be pre-eminent over other disciplines, it is clear that science is marked by distinctive qualities. For instance, the arts do not require evidence or prediction, but a scientific theory requires evidence and predictions. No other domain of knowledge can predict natural phenomena as accurately as science does.
Therefore, the scientific community cannot completely accept naturalist arguments for an irrational non-method of science.

Science educators like McComas et al. (1998) pointed out, “the universal scientific method is one of science education’s most pervasive creeping fox terriers” (p. 57). The typical scientific methods presented in most science textbooks are 1. Define the problem, 2. Gather information, 3. Form a hypothesis, 4. Make relevant observations, 5. Test the hypothesis, 6. Form conclusions and 7. Report results. Like relativists, McComas et al. do not agree with such notions of step-by-step methods, complete objective observation or only inductivism or deductivism. Rather, they support creative and imaginative approaches to nature combining different methods at different stages. They suggest a model of a process of knowledge generation in science, including creative leaps (Figure 2-1) (p. 59).

**Figure 2-1. A Process for Knowledge Generation**

It is obvious that scientists observe, compare, measure, test, speculate, hypothesize, create ideas and conceptual tools, and construct theories and explanations. However, there is no single sequence of activities (prescribed or otherwise) that will guarantee scientists discovery of new theories or solutions. Also, there is no absolutely superior method that is alone certain to other methods. As McComas et al. (1998), Hodson (1998) and Lederman et al.
(2002) point out, myths about scientific method common in science education need to be corrected. Scientists pursue diverse ways to understand how nature works rather than following one strict method.

2.2.6. Section summary

This section has discussed general concepts of NOS in some of their complexity. This study focused on the following NOS: the tentative nature of science, empirical bases of scientific knowledge; observation as theory-laden; the variety of scientific method; and scientific knowledge as culturally embedded. Ontologically, scientific knowledge is either the enterprise of discovery or the enterprise of invention. Epistemologically, its ways of knowing involve either rational, logical, systematic methods with theory-free observation or methods affected by theory-laden observations, culture, and history. In NOS education, these diverse views of science and scientific knowledge can deepen students’ understanding of what science is and is not.

2.3 Why is Students’ Understanding of NOS Important?

NOS includes the epistemology of scientific knowledge as a social and historical construct and the ontology of the knowledge as either out-there existing knowledge to be discovered or an invention for describing the nature. Thus, understanding NOS means students understand how science and scientific knowledge have developed, and affected society and what it means. Kuhn’s (1970) studies on the historical development of science lead him to claim that science is situated in society; scientists’ values – what counts as good questions, appropriate methods and good answers – are constructed and negotiated within particular scientific disciplines and communities. McComas et al. (1998) argue that the knowledge of NOS dispels the myths in science and helps students towards a correct understanding of
scientific knowledge. Driver, Leach, Millar, and Scott (1996) advance five reasons for the importance of NOS: 1) science and technology are useful in everyday life (utilitarian); 2) students as citizens need preparation to participate in democratic decision-making processes on scientific and technological issues (democratic); 3) science must be understood as the outcome of culture (cultural); 4) science cannot be seen as value-free (moral); and 5) science is better understood through exposure to NOS concepts (pedagogical). Many other scholars insist that understanding NOS enhances students’ learning activities (Hodson, 1998; Lederman, 1992; Matthews, 1994; Roth & Roychoudhury, 1994; Sandoval & Morrison, 2003) and their understanding of varied concepts within science (Leach, 2006; Reif, 1995). NOS also motivates students to pursue professional careers in science by improving their perceptions of science. They discard dehumanized and decontextualized images of the field and begin to see work in science as humane and relevant to life (Abd-El-Khalick & Lederman, 2000; Matthews, 2000). Furthermore, NOS helps students understand the interdisciplinary connections between science and applied technology (Driver et al, 1996; Holton, 1993; Matthews, 1994). Finally, NOS provides the needed diversification of instructional methods for teachers (McComas, 2004).

Explaining to students the development and characteristics of scientific knowledge demystifies science and clarifies its unique place among other disciplines (Hodson, 1998; McComas et al., 1998). McComas et al. (1998) and Hodson (1998) argue that unfavourable images of science or the myths of science turn many students away from the field in their early years. Both studies point to a number of such myths prevalent in science education. Science is pictured as an inductive formation of universal, timeless knowledge without consideration of ethics. It is presented without recognizing theory-laden observation or the hierarchical
relationships among scientific hypotheses, theories and laws. Outcomes from scientific experiments are presented as value-free. Scientific progress is painted as a universal algorithm ever stepping upward and onward, decontextualized from actual scientific practice. All this moves students to leave science before they can appreciate its real importance. NOS debunks such myths of science since it deals with not only learning science but also learning about science and doing science (Hodson, 1998). Loving (1991) also asserts the importance of students’ having proper images of science: “Students need to be given an accurate picture of what science is and is not, of its relation to technology and of each to society, of good versus poor science, of what it took to get where we are today, and of the continuous struggle in all disciplines to come up with better explanations about natural phenomena” (p, 824). According to Lederman (1992) and Hodson (1998), NOS plays a pivotal role in breaking up stereotypic views of science. Naïve realism and excessive rationalism dominate student perceptions in science classes. NOS accepts the limitations of science and recognizes science as a human enterprise bounded by cultural and historical circumstances. When these stumbling blocks are removed, students can realize, “science [is] carried out by people and that these people, like everyone else, have views, values, beliefs and interests …people (scientists) can be warm, sensitive, humorous and passionate. More importantly, I want them to realize that people who are warm, sensitive, humorous and passionate can still become scientists” (Hodson, 1998, p. 208).

Consistent with the role of NOS in demystifying such prevailing myths, NOS affects students’ epistemological beliefs about scientific knowledge and practice. This in turn bears an important influence on their approach to learning science (Duschl, 1990; Roth & Roychoudhury, 1994; Songer & Linn, 1991). Duschl stresses knowledge about science as well
as knowledge of science in science education. Without teaching knowledge about science, he argues, students are simply taught current science in its final form. Knowledge about science education may alter and correct students’ biased views on science and scientific knowledge. Duschl (1990) warns that if students are educated on such a one-side view, they may falsely conclude: (a) all scientific knowledge claims are considered equally important, (b) scientific knowledge claims do not interact with others, and (c) scientific theories do not change. That is, without providing opportunities to learn knowledge about science (e.g. history of science), students cannot understand how the collective mind of science arrived at the knowledge it holds today. Roth and Roychoudhury (1994) also assert that

[I]f science is presented to students as a body of knowledge, proven facts, and absolute truths, then they will focus on memorizing facts and think that all knowledge can be ascertained through specific proof procedures embedded in the scientific method. If, on the other hand, students experience science as a continuous process of concept development, an interpretive effort to determine the meaning of data, and a process of negotiating these meanings among individuals, then student might focus on concepts and their variation (p.6).

The concepts of NOS are therefore very helpful to understand scientific knowledge (Leach, 2006; Leach & Scott; 2003; Reif, 1995). Leach explains why epistemological knowledge on science is important. Scientists are generators of knowledge; the processes of knowledge generation are relatively well established with the inductive, deductive and hypothetico-deductive methods. Even though most citizens do not engage in generating knowledge, they can apply the established knowledge and methods in their work places and in their everyday life decision-making. NOS learned in science classes can offer an appropriate basis for sensing
the tangibility and limitations of that knowledge. Thus, Leach argues that an epistemological approach to academic science work can prepare students to address the epistemology involved in generating and using scientific knowledge.

Adopting and enlarging NOS in the science curriculum expands students’ understanding of other cultures outside of the classrooms (Abd-El-Khalick & Lederman, 2000; Matthews, 2000; Solomon, 1992). As technology has developed, worldwide economic trade and cultural exchanges are becoming commonplace. These circumstances require us to develop pluralistic thinking and open-minds toward other cultures. Studying the history and philosophy of science can enhance students’ understanding of other cultures and diversify their ways of thinking as they grasp the cultural and historical milieu in which scientific method developed. Kuhn (1977) illustrates how he could understand the Aristotle’s conceptions on mechanics through the understanding of history of science; “I all at once perceived the connected rudiments of alternate ways of reading the texts with which I had been struggling” (p. xi). Learning the history of science as a part of NOS links the world in which students live to the historical setting and paradigms in which past scientists lived but which are now incommensurable to today. Solomon (1992) stresses that understanding of NOS enlarges the science student’s social, moral and spiritual and cultural context:

Pupils should develop their knowledge and understanding of the ways in which scientific ideas change through time and how the nature of these ideas and the uses to which they are put are affected by the social, moral, spiritual and cultural contexts in which they are developed. (p. 68)

NOS enriches school science with perspectives from the history and philosophy of science (Matthews, 2000; McComas et al., 1998). It provides teachers an abundant instructional
source of anecdote and pedagogical approaches to gain students’ attention to motivate their classroom activities (Lederman, 1992; Matthews, 2000; Paik et al., 2005). Without going into difficult philosophical abstractions, NOS can offer vivid, concrete case studies demonstrating the nature of scientific reasoning and the dynamics of theory evaluation and replacement (Matthews, 1994). Those who support including the history and philosophy of science in school curricula find its value in humanizing otherwise arid course material, in bridging cultural gaps, and in instilling an appreciation for some of the great intellectual adventures in human history.

NOS is known for its importance in enhancing the scientific literacy of the public. Studies on the public’s scientific literacy show that few citizens understand the fundamental principles of science or the operations of the scientific enterprise – this, even though science has pervaded and influenced every aspect of modern life (Bell & Lederman, 2003; Hodson, 2008; Holton, 1993; Lee, 2007). This lack of understanding is potentially harmful, particularly in societies where citizens have a voice in funding scientific decisions, in evaluating policy matters in science and in weighing scientific evidence provided in legal proceedings. Public participation is fundamental to a sound society and its proper decision-making processes. The public must have sufficient scientific literacy even to read newspapers or follow media debates on such issues in everyday life.

This section has thus provided the main rationale for teaching and learning NOS. Science is more than a body of assorted, factual knowledge. Learning about science is “something about the character of scientific knowledge, how it has been developed, and how it is used” (Hurd, 1960, p.34). National Science Education Standards (NSES) (National Research Council, 1996) support the idea that learning the history and nature of science is “an
understanding of what science is and what science is not, what science can and cannot do, and how science contributes to culture” (p.21). The following section will review the opinions of those skeptical of NOS’s effects on learning science and public scientific literacy.

2.4 Why NOS Should Not Be in Science Education?

Aikenhead (2008), and Dogan and Abd-El-Khalick (2008) say that instructional strategies must consider students’ cultural backgrounds as well as their physical and mental capabilities. Universalists, however, who argue that science is trans-cultural and trans-historical (Matthews, 1994; Siegel, 2002; Stanley & Brickhouse, 1994), do not always welcome the culturally embedded, subjective aspects of NOS. Universalists do not regard knowledge from different cultures as equally important. Also, other researchers voice the concern that students’ knowledge of science has deteriorated, leaving them less prepared to major in science at post secondary levels (Millar & Osborn, 1998; Turner, 2008).

Universalists contend that science curriculum and science class should deal more with scientific knowledge rather than the cultural, subjective aspects of science since they declare:

Science as an intellectual activity whose truth-finding goal is not, in principle, affected by national, class, radical, or other differences: Science transcends human differences…..the character of the natural world is unrelated to human interest, culture, religion, race or sex. Ultimately, the concept is judged by the empirical evidence, not the other way around. (Matthews, 1994, p.182)

Within limited class time, Universalists argue, teachers should teach only the best of what is known currently. The best science is not limited to any culture or race but is worldwide (Matthews, 1994; Siegel, 2002). They also complain that Multiculturalists’ views of science
(the egalitarian position toward all cultures) undercut the key goal of science, i.e., the parsimonious, accurate description of nature.

Millar and Osborne (1998) likewise emphasize knowledge of science, saying that science education values the products of science and technology, which are embedded in our daily lives. Citizens’ scientific literacy and participation in democratic decision-making should most wisely be founded on current scientific knowledge. Furthermore, scientific knowledge is necessary for all young people whatever their career aspirations or aptitudes. Besides, school science education should emphasize futuristic advances:

We need, as a society, to train and educate new generations of scientists and technologists to maintain the technological tools and systems we value and to develop new and better ones to meet new needs and solve new problems. School science is, for some young people, the start of the process which will enable them to become the scientists and technologists of the future. (Millar & Osborne, 1998, p.2008)

There is a lively debate over the role of NOS in enhancing public scientific literacy to participate in society’s decision-making procedures on science-related issues. Research conducted by Bell and Lederman (2003) and by Lee (2007) reveals that the public’s decision making in their everyday life and on societal issues does not necessary depend on their scientific knowledge. Rather, their prior experiences, and their ethical, personal values play a crucial role. Hodson (2008) also complains that it is naïve to believe that the shallower knowledge of science affected by increased knowledge about science in the curriculum will ultimately help scientists begin or succeed in their careers. The public may not tolerate the
expenditure of their tax money without feasible short-term outcomes that might serve wider society but not their own practical interests.

Within given curriculum time, teaching NOS makes a trade-off of the time for scientific knowledge (Matthews, 1994; Turner, 2008; Turner & Sullenger, 1999). In sum, opponents of NOS education in the classroom warn that the goal of the Benchmarks and the Standards, i.e., “Science-for-All”, is neither realistic nor necessary. Turner and Sullenger (1999) criticize these NOS educational measures thus: “All the framework documents mildly de-emphasized content, university preparation, professional recruitment, and the pursuit of excellence, and shifted the rhetoric of science teaching toward higher minimum standards for all students” (p.8).

Other opponents of NOS in K-12 classrooms point to teachers’ naïve understanding of NOS. If NOS is added on to the existing curriculum (Turner, 2008), the content will overload teachers working within a limited time. When balancing the pros and cons of these arguments, science educators have to decide whether or not they will teach the concepts of NOS, and, if they will, how. A review of previous NOS studies will be helpful to understand possible pedagogical approaches.

2.5 What Has Been Done So Far?

Through his comprehensive review, Lederman (1992, 2007) classifies NOS related research into four distinct categories: (a) assessment of student conceptions of NOS; (b) development, use, and assessment of curricula designed to “improve” student conceptions of NOS; (c) assessment of, and attempts to improve, teachers' conceptions of NOS; and (d) identification of the relationship among teachers' conceptions, classroom practice, and
students’ conceptions. The literature review in this section focused mainly on (b) and on one aspect of (d), i.e., what are the factors of affecting the improvement of understanding.

### 2.5.1 Measuring the concepts of NOS

Influenced by dominant educational and philosophical paradigms, the formats of instruments and the ways of asking for content have changed. Research methods have shifted from quantity focused standardized paper-and-pencil tests to quality focused open-ended questions and interviews; Realistic, rationalistic questions have given ground to constructivist, naturalistic questions (Aikenhead, Fleming & Ryan, 1987; Lederman, Wade, & Bell, 1998; Lederman et al., 2002; Liang et al., 2008). Table 2.1 is a summary of a few instruments, which have been widely used.

Reviewing previous instruments in the area is beneficial to reduce trial-and-error in developing an instrument for this research. In 1976, Rubba invented *Nature of Scientific Knowledge Scale (NSKS)*. It consists of 48 items with a Likert scale of five choices (strongly agree, agree, neutral, disagree, and strongly disagree). The target participants were secondary school students. The constructs measured the creative, parsimonious, moral, developmental, testable and unified nature of science. NSKS was attacked in for wording problems and for redundant questions (Lederman et al., 1998, 2002). Most questions were paired with positive and negative wording, i.e., question 1 is “Scientific laws, theories and concepts do not express creativity” and question 20 is “Scientific laws, theories and concepts express creativity.” This problem can inflate the scale’s reliability and create misplaced confidence in its results.

Forced-answer instruments including multiple choice and Likert scales have been criticized for reflecting too patently the views of their developers on NOS (Aikenhead et al., 1987; Lederman et al., 1998; Liang et al., 2008). Often the NOS concepts presented are
oversimplified (Alters, 1997; Liang et al., 2008), and it is impossible to identify the reasons for a given answer or choice (Aikenhead et al., 1987). Alters (1997) points out that the developers of forced-answer instruments do not represent adequately the perspectives of scientists, philosophers and science educators. Aikenhead et al. (1987) find the fact that different students choose a particular answer for widely different reasons. These considerations have led Aikenhead et al. to develop an assessment instrument entitled *Views on Science-Technology-Society (VOSTS)* working over a six-year period with Canadian students in Grades 11 and 12. It consists of 114 multiple choices and students can express alternative views. Aikenhead and Ryan (1992) promote *VOSTS* because “[the] items focus on the reasons that students give to justify an opinion” (p. 480). Critics attack *VOSTS*’s redundant questions and strong Canadian context. So, users in other countries need to adjust it to avoid its nationalistic sentiments (Lederman et al., 1998; Liang et al., 2008).

A more recent and influential instrument is entitled *Views of Nature of Science (VNOS)*, a questionnaire developed by Lederman, Abd-El-Khalick, Bell and Schwartz (2002). It has several versions such as VNOS-A, B, C and D, which show variations in the number of questions and in the complexity of language used in the questionnaires. All questions are open-ended and VNOS’s developers strongly recommend administering follow-up interviews. It is quite demanding to answer the questions, which assume both considerable knowledge of NOS and mature writing skills. For instance, the VNOS-C target participants are mostly undergraduate or graduate students and to complete the ten questions takes about 45-60 minutes. Liang et al (2008) point out that such a demanding task can result in many blanks or very brief responses. In addition, results can be significantly affected by the participants’ writing skills.
Liang et al. (2005, 2008) have in turn developed a mixed-format instrument entitled, *Student Understanding of Science and Scientific Inquiry (SUSSI)*, which focuses on the concepts of observations and inferences, change of scientific theories, scientific laws and theories, social and cultural influence on science, imagination and creativity in scientific investigations and methodology of scientific investigation. Each aspect of these NOS concepts is investigated with four close-ended (“strongly disagree” to “strongly agree”) questions and with open-ended follow-up questions for each scale. The developers have investigated and established the reliability of the instrument with international pre-service teachers and undergraduate students (USA, China and Turkey).

Table 2-1

*NOS Instruments*

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Authors and Year</th>
<th>Target sample</th>
<th>Structure</th>
<th>Focus area</th>
<th>Characteristics/ Advantages/ Disadvantages</th>
</tr>
</thead>
</table>
| TOUS (Test on Understanding Science) | Cooley & Klopfer (1961) | Broad         | 60 items multiple choice -forced response | -Understanding about the scientific enterprise; -scientist; -Methods and aims of science | -The first widely used instrument  
-TOUS covers science and science enterprise not scientific knowledge.  
-too many items embrace a negative viewpoint of science (Lederman et al., 1998)  
-lack of validity (Welch, 1996)  
vague words result in the failure of sorting items (Hukens, 1963 cited in Lederman et al., 1998) |
| NOSS (Nature of Science Scale) | Kimball (1968)     | Science teachers/ scientists | 29 items with agree/neut ral/ disagree | Philosophical aspects of science (goals of science/ scientific methods/ knowledge) | well covered the philosophical aspects of science  
lack of validity of high and lower grade level students (Aikenhead, 1973) |
<table>
<thead>
<tr>
<th>Scale Description</th>
<th>Author(s)</th>
<th>Age Group</th>
<th>Scale Details</th>
<th>Judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSKS (Nature of Scientific Knowledge Scale)</td>
<td>Rubba and Anderson (1978)</td>
<td>Grades 9-16</td>
<td>48-item Likert scale (five levels) six subscales Amoral/creative/developmental/parsimonious/testable/unified</td>
<td>-the higher graders the higher reliability (grade 9: the lowest) -reflection of falsificationism in developmental aspects -wording problems (redundant: asking the same question positively and negatively)</td>
</tr>
<tr>
<td>VOSTS (Views on Science-Technology-Society)</td>
<td>Aikenhead, Ryan, &amp; Fleming (1987)</td>
<td>Canadian secondary school graduates (grade 11~12 at least)</td>
<td>114 multiple choices/Open-ended questions</td>
<td>STS issues - Questionnaire about views of science and science/technology/society -providing rooms to express alternative students’ viewpoints -questions include Canadian specific situations, so to use other nationalities it needs some adjustment</td>
</tr>
<tr>
<td>Modified NSKS</td>
<td>Meichtry (1992)</td>
<td>Grade 6, 7 and 8</td>
<td>32 statements from NSKS Creative/developmental/testable/unified</td>
<td>-reworded from NSKS -paired positive and negative statements</td>
</tr>
<tr>
<td>VNOS-C (View of Nature Of Science-C)</td>
<td>Lederman, Abd-El-Khalick, Bell, Swartz (2002)</td>
<td>Undergraduate/graduate students</td>
<td>10 open-ended questions Philosophy of science based on Kuhnian (1970) views (Tentative/empirical/creative/subjective/social embeddedness/scientific theories and laws/</td>
<td>-VNOS should not be used to label learners’ views adequate or inadequate and no numerical score. -length and complexity of language used in questionnaire -demanding writing skills</td>
</tr>
</tbody>
</table>
2.5.2 Science Textbooks and Students’ Understanding of NOS

Since science textbooks are the major resource of teaching and learning in K-12 classrooms, how the textbooks were written and what the textbooks have are important in students’ understanding of NOS (Bensaude-Vincent, 2006; Hodson, 1982; Kuhn, 1970; Matthews, 1994). In Kuhn’s opinion, textbooks are very conservative tools, allowing students only to explore what has already been established and to engage in solving puzzles in normal
science. They do not encourage students to challenge or to create new, different ways of thinking or solving problems from the previous ways. Much in the same vein, Latour (1987) also criticizes the conservatism of science textbooks. He argues that science textbooks should emphasize two aspects of science: ready-made science data and science-in-the-making method. However, most textbooks heavily weigh the ready-made science; therefore, students are deprived of sufficient opportunity to explore how science is made. Textbooks, as the basis of all curricula, have been training students to focus on “answers rather than the exploration of questions, memorization versus critical versus critical thought, information than understanding, recitation as opposed to argument, and reading in lieu of doing” (Johnson & Stewart, 1991, p. 208)

Textbooks are often devoted to describing factual knowledge using jargon (Clough, 2006; Matthews, 1994; Yip, 2006) and to introducing one or other of the mainstream views on scientific knowledge, predominantly the empiricist and inductivist schools (Hodson, 1982; Kim, Jeon, & Pak, 2007; Yip, 2006). Rather than helping students to develop a meaningful understanding about scientific knowledge and the conditions by which it develops, science textbooks provide students a stock of unconnected facts and concepts. Bensaude-Vincent (2006) says that

[S]cience textbooks do not include anything about painstaking efforts, about enthusiasm fierce struggles involved in the construction of scientific facts. More exactly they tell something only in so far as they emphasize the divorce between science creators and the anonymous crowd of science transmitters. (p.667)

Though widely criticized, science textbooks remain the major instructional resource of teachers and students. Kim et al.’s research (2007) throws light on how one science teacher
used her assigned textbook. First, she presented what the textbook described science and scientific knowledge (which was cast in empiricist and inductivist terms), adding some explanations on the meaning of the descriptions. She then provided a historical example, which actually verified the views. She essentially transferred what the textbook described to her students without critical reflection, which might filter these processes or introduce other views. After the unit study, the students’ views on NOS were examined. Twenty-two out of twenty-eight (78.6%) expressed inductivist and empiricist views. Other studies (Paik, 2000; Cho, 2001) on textbooks reached similar results. In effect, regardless of teacher education, teachers avoid expressing any views, which differ from the positions found in their science textbooks.

2.5.3 Teachers’ understanding of NOS and students’ understanding of NOS

Most science teachers are the products of textbooks and curricula founded on the prevailing schools of empiricist, inductivist and realist philosophy (Matthews, 1994; Paik et al., 2005; Turner & Sullenger, 1999). Thus, research reports that teachers have heavily realistic and rationalistic views of NOS (Cobern & Loving, 2001, 2008; Kimball, 1967; Lederman, 2007; McComas et al., 1998). Also they rarely gain the opportunity to keep abreast of the current trends in education involving constructivism and the re-conceptualization of the nature of knowledge (McComas et al., 1998). It is therefore hardly surprising that science teachers fail to emphasize new and diverse views of science to their students.

General assumptions are teachers’ beliefs on science and scientific knowledge will reflect in their classroom instruction, and that these instructional activities will shape their students’ learning. Regarding academic achievement, Becker and Aloe’s study (2008) supports the assumption of the direct causal relationships in students’ academic achievement and
teachers’ pedagogical content knowledge. They conducted a meta-analysis on teachers’ content knowledge and students’ understanding of the subject and found a positive correlation. Fraser and Fisher (1983) also corroborate the concept of teacher influence on student outcomes and found similar results to Becker and Aloe’s study.

However, the direct linear causal relationship between students’ understanding of NOS and teachers’ instructions has been controversial through comprehensive research (Ahn & Jung, 1996; Becker & Aloe, 2008; Lederman, 1992, 2006; Lucas & Roth, 1996; Mackay, 1971; Waters-Adams, 2006). Some studies on NOS (Cho, 2001; Gallagher, 1991; Palmquist & Finley, 1997) support the positive relationships between teachers’ understanding of NOS and students’ understanding of NOS while neither the impact of a teacher’s knowledge of NOS on student knowledge of NOS, nor the influence of a teacher on student attitudes towards science is significant (Ahn & Jung, 1996; Lederman, 1999; Lederman & Druger, 1985; Waters-Adams, 2006). Lederman (2007) reviewed the relation between teacher of NOS and teacher instructional activities, as well as teacher understanding of NOS and student understanding of NOS, finding no valid correlation in either case. The current view has emerged that teacher’s knowledge is necessary, but not sufficient for improving students' conceptions of NOS.

### 2.5.5 Pedagogical approaches to NOS

To teach NOS concepts effectively, researchers have adopted explicit and implicit methods. Explicit methods can teach NOS concepts directly through discussion or lectures. Implicit methods approach NOS concepts indirectly by using inquiry-based science (Moss, Abrams, & Robb, 2001) or an interdisciplinary science curriculum (Matthews, 2000; Pedretti, 2003). There is controversy over the relative merits of these two approaches. Explicit methods seem to be more immediately effective than implicit methods in communicating NOS concepts.
However, such direct teaching can give students the impression that NOS concepts are simply more knowledge of science rather than knowledge about science. Explicit methods thus produce results exactly opposite to the original intents of NOS education (Matkins et al., 2002; Moss, Abrams, & Robb, 2001).

Many studies have reported that explicit strategies are more effective than implicit strategies (Abd-El-Khalick & Lederman, 2000; Akerson et al., 2000; Bell, Lederman, & Abd-El-Khalick, 1998; Bell et al., 2003; Feldman, 2003; Khishfe & Abd-El-Khalick, 2002; Scchwartz & Lederman, 2002; Schwartz et al., 2004). Explicit strategies employ discussions and questions on a scientific theory development and its changes, which make the concepts of NOS obvious. Abd-El-Khalick et al. (1998) reported that using a reflective explicit approach to a pre-service teachers’ class, more than 90 % reached informed views of NOS concepts. On the other hand, such implicit approaches to NOS as inquiry–based science with its hand-on activities, cannot impart NOS concepts directly; rather they demonstrate practically the ways by which science works (Matkins et al., 2002).

Alternatively researchers try interdisciplinary approaches to enhance class understanding of NOS. Some use the history of science (Abd-El-Khalick & Lederman, 2000; Irwin, 2000; Kang, 2005), science, technology, society and environment (STSE) in science curriculum (Bentley & Fleury, 1998; Pedretti, 2003) or broad domains of subject areas such as physics, chemistry, and music (Matthews, 2000). Studies of history of science and STSE in science education are here reviewed.

**History of science:** The Standards (NRC, 1996) emphasize the history of science in the science curriculum:

History provides another avenue to the understanding of how science works . . . it is equally important that students should come to realize that much of the growth of
science and technology has resulted from the gradual accumulation of knowledge over many centuries. (p. 4)

In dealing with the history of science, Kuhn (1970) emphasizes that any specific scientific discovery should be understood within its social context since the sociocultural environment cultivates paradigms, defining the types of questions that scientists ask and critical aspects of the subsequent investigations. Allchin (2002) warns that the isolation of a historical case from its social situation may mislead students about the epistemological foundations of science. In his view, if a teacher presents only well-known, successful discoveries or well-organized ways of scientific discovery students can gain the impression that science unfolds via a special method, in which well-designed experiments invariably predict and reveal truth. Students thereby get a dichotomous view of science as a matter of correct or incorrect.

Studies of the effect of history of science on NOS are not consistent. Kang (2004) and Rudolph and Stewart (1998) show positive effects while Abd-El-Khalick and Lederman (2000) and Irwin (2000) report no significant effect. In Kang’s (2004) and Rudolph and Stewart’s (1998) studies, those students who have the opportunity to learn the history of science class had clearer views on the tentative characteristics of scientific knowledge than did the control group students. Lonsbury and Ellis (2002) examined the effects of historical figures and events in science class on students’ understanding of NOS, and reported the results that without dropping down of the required acquisition of knowledge of science (in their content, Mendel’s genetic law), students acquired scientific knowledge is testable rather than absolute. However, in Abd-El-Khalick and Lederman’s (2000) study of the impact of college history of science courses on students’ understanding of NOS, most students showed little increase in their understanding of NOS while the students who had a strong initial understanding of NOS themes illustrated somewhat improved views at the end of the course. Their interpretation of
the results confirms Allchin’s (2002) warning: the students were unable sufficiently to remove themselves from their current reality to study from a strict historical perspective.

**STSE and NOS:** Pedretti (2003) and Bentley and Fleury (1998) emphasize NOS concepts, particularly the interaction with society and science, using STSE education. These studies seek to demonstrate that science is not an isolated examination of natural phenomena, but actively interacts with society. Pedretti recommends that teacher education provides teachers with concrete resources, both to help them develop a theoretical underpinning and justification for STSE and to create opportunities to discuss and practice the principles of science in society. Matthews (2001) suggests interdisciplinary instruction can be developed across science, with history, philosophy, music and religion on one hand, and actual science content on the other. He demonstrates that the pendulum motion in physics can be explored in class through an interdisciplinary discussion involving history, physics, geometry, music and social studies:

> Pendulum motion, if taught from a historical and philosophical perspective, allows connections to be made with topics in religion, history, mathematics, philosophy, ....as well as other topics in the science programme. And such teaching promotes greater understanding of science, its methodology, and its contribution to society and culture.

(p.366)

Teaching science in this manner will certainly make science more relevant and easy to connect to everyday life and expect to improve students’ understanding of NOS.

**2.6 Cross-country Studies**

Aikenhead (2008) opposes spreading curricula based on the *Standards* (NRC, 1996), and the *Benchmark* (AAAS, 1993) across cultures because many countries accept the curricula despite their different educational cultures. Cross-cultural studies can identify such differences
and expand the cultural worldview. International comparative studies on students’ or teachers’ NOS have revealed significant differences (Cobern, 1989; Halai & McNicholl, 2004; Liang et al., 2009). In Cobern’s study, more Nigerian pre-service science teachers viewed science as a way of producing useful technology than their American counterparts. Cobern interpreted this finding as a reflection of the Nigerian Government’s policy on science as a stimulant for national development. Halai and McNicholl (2004) studied science teachers’ NOS in two different contexts: Pakistan, a Muslim-dominant society, and England, a Western society. Both societies possessed many commonalities in understanding NOS except in the religious aspects. However, when the researchers probed these similarities, they found that the reasons given for them were quite different. For instance, both countries’ science teachers agreed on tentative nature of scientific knowledge. However, teachers from England said that people’s ways of thinking, new evidence, and technological development bring changes to scientific knowledge while a few teachers from Pakistan answered that according to the Quran, God knows everything, but people are imperfect, hence, scientific knowledge is imperfect and changeable. Liang et al. (2009) examined Chinese, American, and Turkish pre-service teachers’ views of NOS. The Chinese participants held the most contemporary views, while Turkish participants possessed more traditional views. The authors’ interpretation was consistent with Cobern’s (1989) that teachers’ views were influenced by their national policies and educational cultures.

2.6.1. NOS Studies in Canada and Korea

2.6.1.1 Canada

Aikenhead (1987) reported the results of nationwide research on secondary school graduates’ conceptions of NOS using VOSTS. This report provides not only numeric results from large samples but also deeper explanations of students’ responses. It marks a watershed
from quantitative to mixed method of NOS research. Griffiths and Barry (1991) and Griffiths and Barman (1995) used interviews in their studies because they argued that the standardized large-scale NOS instruments only scratched the surface, but failed to measure students understanding of NOS deeply. The students in Griffiths and Barry's study considered scientific knowledge to be tentative but held scientific factual information to be absolute and irrefutable. Many of them thought that as a theory accumulates evidence, it becomes a scientific law. Griffiths and Barman’s (1995) comparative study with Australian, Canadian and American students aged 16-18 found that students had hierarchical conceptions concerning scientific theories and laws. The views of American students were more traditional and those of the Australians were more contemporary. The views of Canadian students were somewhere between these two.

Roth and Roychoudhury (1994) examined the effects of teacher and student interaction on student epistemologies. Forty-two Grade 10 and Grade 11 students in a physics course participated in the study. During a course on climate change, the teacher actively engaged the class in open-ended inquiry and whole-class discussion with a disinterest for “the right answer”. In a pre-test, students positioned strongly with objectivist points of view. After the intervention, two-thirds of the students still believed that scientific knowledge is exact, not tentative, and independent from human conceptualization. The result is consistent with other findings that teachers’ NOS is not a direct cause of students’ understanding.

Lucas and Roth (1996) and Pedretti (2003) conducted longitudinal studies on NOS. Lucas and Roth studied two students using a book Inventing Reality for 15 months and found that students’ understanding of NOS was neither easily changed nor closely related to their teachers. A student who focused on high marks in science cared little for NOS and wanted
only an effective environment for transmitting learning. In Pedretti's (2003) study, pre-service teachers learned the concepts of NOS embedded in STSE. After the participants became in-service teachers, how they practiced the concepts in their classrooms was examined. Pedretti found gaps between the intended curriculum and the enacted curriculum because class time and resources were insufficient to cover NOS.

2.6.1.2 Korea

Before the late 1980s, studies on NOS were rarely conducted, but since the 1990s studies on NOS have been conducted to assess both student and teacher understanding of NOS (Soh, 1998; Paik, 2000). As most science classes heavily cover the knowledge of science, teachers' as well as students' understanding of NOS are very poor (Hong, 2001; Kwun & Park, 1995). Middle and high school students have strongly inductivist and empiricist perspectives regardless of their grade level, gender, and academic orientation (Ban, Lee, Kim, & Park, 2000; Hong, 2001; Im, 2002; Jahng, 1995; Kwun & Park, 1995; Seoh, 1996).

A few studies try to implement an instructional strategy to change heavily inclined inductivist and empiricist views on science into more informed views (Kwun & Park, 1995; Hong, 2001). Kwun and Park’s study taught NOS concepts directly to pre-service elementary teachers, but there was no statistically significant difference between pre and post-test results. The participants’ NOS concepts were quite solid and were not easy to change within one semester of instruction. Middle and high school teachers who majored in physics, chemistry, biology or earth science take mandatory science education method courses, which include the concepts of NOS. They have inclined epistemologically to falsificationism and ontologically to realistic views (Cho, 2001; Jahng, 1995). Hong (2001) carried out interpretive and analytic
reading materials about scientific knowledge and historic events. The students’ views have changed from extreme to moderate inductive views.

Likewise, many studies in Western research on NOS over the past thirty years (Lederman et al., 1998, Lederman, 2007), show that the typical Korean student’s understanding of NOS falls far short of currently recommended views (Kang, Kim, & Noh, 2004). Kang et al. argue that K-12 science education over-emphasizes knowledge of science (the products of science) and has not paid sufficient attention to knowledge about science, or to doing science.

2.6.1.3 Comparison of NOS studies between Canada and Korea

Unlike NOS studies in Canada, the NOS studies in Korea reviewed above used large-scale standardized instruments invented by Korean scholars, (e.g., Soh (1998)) or translated and modified versions of NSKS (Rubba & Anderson, 1978). Except for a few studies like Kwun and Park (1995) and Hong (2001), most studies have focused on identifying current understanding of NOS, not on the effects of instructional methods, e.g. implementing a constructivist approach. This research trend may reflect the product-centered emphasis of Korean educational culture and its test-driven curriculum (Kim et al., 2007). Like the student in the Lucas and Roth (1996) study, who was oriented to high marks in achievement tests, and did not care much on NOS, Korean teachers will not risk impairing student performance on achievement tests by spending many class hours teaching the concepts of NOS.

2.7 Summary

For decades, science educators have conducted extensive research to improve student understanding of NOS. There are still debates whether NOS should be included in science education at all, and which factors in NOS are truly influential. Cross-country studies identify
several differences among countries. Perhaps these inconsistencies arise from the complexity of learning and teaching environments, the abstract concepts of NOS itself, the different measuring instruments, and the researchers’ concepts of NOS. Cross-cultural studies revealed differences in the understanding of NOS, which reflect educational culture and societal differences. There are no completely wrong or right ways to improve students’ understanding of NOS.
3. CLASSROOM LEARNING ENVIRONMENT,
ACHIEVEMENT ASSESSMENT AND CONCEPTUAL FRAMEWORK

3.1. Overview

Learning and teaching make up a complex enterprise involving many factors that interact with each other. Selecting a specific factor to measure for its learning effects is not easy, for it is not possible to isolate it to the exclusion of all other factors. Finding all potential factors and considering all the factors in a study are both impossible and meaningless. Nevertheless, an effort to identify a few potential elements is crucial for the development of learning and teaching enterprises.

This research proceeds on the assumption that if understanding NOS is part of the expected outcomes of science education, then, students’ learning environment and achievement assessment can be significant factors in achieving that outcome. Thus, section 3.2 is devoted to the definition of learning environment and section 3.3 reviews the studies on the field. Section 3.4 includes a review of instruments assessing learning environment and of studies associated with these instruments. Section 3.5 attempts to trace back how learning and teaching paradigm shift has influenced the studies of NOS and instruments. Section 3.6 describes an overview on the effects of standardized large scaled assessments on learning and teaching in Canada and in Korea. Section 3.7 reviews the instruments of students’ perceptions of achievement assessment. Based on these reviews Section 3.8 describes the conceptual frameworks for this study. Lastly, section 3.9 provides a summary of the Chapter.
3.2. Definition of Learning Environment

Since the 1960s, much attention has been given to gauge the quality of learning environments from students’ perspectives (Fraser, 1998, Kim, Fisher & Fraser, 2000). “Learning environment” includes not only the physical infrastructure but also the quality of the students, of the teachers, and of the educational curriculum (Fraser & Walberg, 1995; Robins et al., 1995). Emphasizing students’ achievement of and attitudes towards a subject matter, Fraser and Walberg (1995) argue social, psychological and pedagogical contexts that nurture learners’ growth should be considered. Hannafin, Land and Oliver (1999) suggest that a successful learning environment provide learners with abundant resources, which learners can use to build a scaffold for their learning. Besides learning resources, Robins et al. (1995) stress that an appropriate assessment such as timely evaluation of students’ performance with constructive feedback is an important component for successful learning.

Given the wide scope of students’ over-all learning environment, this research has confined its focus to student perceptions of learning activities and tasks within science classes. These activities and tasks were broadly classed as teacher-directed (TD) or student-directed (SD). Typical TD activities are: listening to explanations, copying what the teacher wrote on the board, watching teachers demonstrates experiments, etc. In these experiences, students are passive recipients of knowledge from their teachers. In contrast, SD activities give priority to students’ involvement, and put them in charge of their learning. Typical approaches to encourage student involvement are lab activities (Hodson, 2005; NRC, 2000) and inquiry based learning (Schwartz, Lederman, & Lederman, 2008). However, cookbook-style lab activities cannot be classed as student-directed since teachers decide almost entirely what to achieve and how to perform the processes. Peer interactions such as class discussions and
presentations are included as student-directed learning activities from a socio-cultural perspective of learning (Vygotsky, 1978). Students’ perceptions of these TD and SD learning activities were measured by posing direct questions how often students have engaged in them, and how often they are required (almost always, often, sometimes, seldom and almost never).

3.3 Reviews of Learning Environment

Studies have asserted that the quality of learning environment is a strong indicator for the development of potential and the improvement academic achievement (Allodi, 2007; Dorman, Adams & Ferguson, 2002; Fraser, 1998; Fraser & Walberg, 1995; Goh & Fraser, 1998). Students who evaluate their learning environment positively achieve a higher level of outcomes than those who evaluate it negatively. Haertel, Walberg and Haertel (1981) conducted a meta-analysis on learning environments and revealed that students’ achievement and their perceptions of learning environment are positively correlated. Results from Program for International Student Assessment (PISA) 2000 (PISA, 2006) also demonstrate that the learning environment of schools strongly influences students’ performance in literacy. Recently, the outcomes measured have been extended; not only cognitive domains but also affective domains such as attitudes toward science or mathematics are being assessed and the relationships are positive (Dorman, 2009; Dorman & Knightley, 2005; Dorman et al., 2006; Lee & Fraser, 2001).

This review highlights the relationships between the learning environment and cognitive and affective outcomes in terms of class size, students’ characteristics, and several instruments. The last part of the review deals with Canadian and Korean studies.

Class sizes are one of the indices of the interactions between teachers and individual students and of individual student involvement in learning (Dorman, 2009; Mulline et al.,
Dorman uses a multilevel analysis to identify how individual, class and school levels affect students’ perceptions of their learning environment. Although the class size was not the main point of the research, it clearly showed that the class differences were predictive for the variances of students’ perceptions across all constructs of the learning environment. TIMSS reports were a good example. For the most part, small class size countries achieved high scores in TIMSS science tests except for Asian countries. However, for Ontarian and Korean classes, class size and student achievement had no significant relationship. For 65% of Ontario schools, the average class size ranged from 25 to 40 and 33% of the schools had less than 24 students; yet both group achieved 527 in the TIMSS science test. Korean class sizes were larger than Canadian classes. For 75% of Korean schools, the average class size was 25–40, and achieved a TIMSS score of 554. Some 21% of the schools had more than 40 students, and achieved a TIMSS score of 555. The class size effects on student achievement in TIMSS science tests were not significant in these two countries.

Studies in Canada show positive relationships between learning environment and academic performance (Dorman, 2003; Dorman, Adams, & Ferguson, 2003; O’Reilly, 1975; Roth, 1999). O’Reilly (1975) investigated the relationship between classroom environment and students’ mathematics achievement in Ontario using Learning Environment Inventory Scales, developed by Walberg and Anderson (1968). He found that 67% of variance in raw achievement scores was accounted for by the learning environment inventory scale. Roth (1999) conducted a small-scale study using Constructivist Learning Environment Survey (Tayler, Dawson & Fraser, 1995) in order to reform a high school science classroom. He investigated the relationships between students’ perceptions of learning environment and academic achievement. For him, a mixed method to diagnose and reform a classroom,
involving a survey questionnaire, student interviews, and test results, was crucial. Dorman et al.’s (2002) study attracts attention because it probes students’ psychological learning environments and student academic self-efficacy in a Canadian context among others. They conducted a cross-country study using *Constructivist Learning Environment Survey (CLES)* (Fraser, 1986) and *What Is Happening In this Classroom (WIHIC)* (Aldridge & Fraser, 2000; Fraser, 1998) with 3602 secondary school students from Australia, the United Kingdom and Canada. They reported that those students who perceive their learning environment positively achieve higher scores than those who do not.

Learning environment studies in Korea have been conducted since the late 1990s. Kim, Fisher and Fraser (2000) used the *Constructivist Learning Environment Survey (CLES)* (Tayler, et al., 1995). Kim, Fisher and Fraser’s (2000) study used *WIHIC*, and Baek and Choi (2002) used a modified and translated version of the *CLES* to investigate the learning environment and academic achievement. Fraser and Lee (2009) examined the unique learning environment Korea’s secondary school students using *Science Laboratory Environment Inventory (SLEI)*. The participants were science-independent, science-oriented, and sociology-oriented students. Science-independent schools are quite unique. The students in these schools are highly filtered (their academic achievement in middle school should be in the upper 1% and each province has only one science-independent school). Science-oriented students and sociology-oriented students are not greatly different. As expected, the results showed that science-independent school students had more open topics and freedom during science lab activities and than the other two groups. In the Korean context, the validity of *WIHIC* was established.

Kim, Fisher and Fraser (2000) examined Grade 8 students’ perceptions on their learning environment and their attitudes to science using *WIHIC* and *QTI (Questionnaire on*
Teacher Interaction) (Wubbels & Levy, 1993). They also examined gender differences between boys- only and girls-only schools in students’ perceptions on their learning environment, teacher interpersonal behavior and attitudinal outcomes. The Korean students in the study showed high means in the scales of Students Cohesiveness, Task Orientation and Cooperation while low means in the scales of Teacher Support, Involvement, and Investigation. School differences from different provinces were presented, but the effect sizes were small. The researcher inferred that the small effect sizes were due to the nationwide uniform curriculum. Gender differences were significant across all scales. Furthermore, the associations between the perception of the learning environment and students’ attitudes towards science were significant, particularly teachers’ helping and friendly behavior had a strong association with the positive attitudes towards science.

3.4 Instruments for Assessing Learning Environment

Many instruments for assessing the learning environment have been developed. Among these are the Learning Environment Inventory (LEI) (Walberg & Anderson, 1968), the Questionnaire on Teacher Interaction (QTI) (Wubbels & Levy, 1993), the Constructivist Learning Environment Survey (CLES) (Taylor, Dawson & Fraser, 1995; Taylor, Fraser, & Fisher, 1997), the Science Laboratory Environment Inventory (SLEI) (Fraser, Giddings, & McR Robbie, 1995; Fraser, McR Robbie & Giddings, 1992), the Classroom Environment Scale (CES) (Moos & Trikett, 1987), and What Is Happening In this Classroom? (WIHIC) (Aldridge & Fraser, 2000; Fraser, 1998). The major categories of assessment in the learning environment field are the comparison of actual and preferred environments, school psychology, and teacher education (Fraser, 1998). In this section, QTI, CLES and WIHIC are reviewed, particularly,
those which have been developed for junior and secondary school students, and have been used in Canada and Korea.

*QTI* (Wubbels & Levy, 1993) was developed for primary and secondary school students to examine the interactions between students and a teacher; i.e., how a teacher distributes students’ responsibilities, freedoms and restrictions. It has been used extensively over the world. In Korea, Kim et al. (2000) and Lee and Fraser (2001) used this instrument for secondary school science classes, confirming its validity and reliability for this context. As reviewed in the previous section, *QTI* was predictive of the variances of students’ attitudes toward science for Grade 8 Korean students.

*CLES* (Taylor et al., 1995; Taylor et al., 1997) reflects a constructivist epistemology of learning. It is intended to help teachers and researchers assess the degree to which a classroom reflects constructivist epistemology, and if desired, to change teaching practice to a more constructivist style. The focusing scales are called Personal Relevance, Uncertainty of Science, Critical Voice, Shared Control and Student Negotiation and each scale has 6 items ranging from Almost Never to Almost Always. It has been administered in many countries including Canada (Roth, 1998) and Korea (Kim et al., 1999; Lee and Fraser, 2001) and has been validated for elementary, middle and secondary school students and teachers.

*WIHIC* (Aldridge & Fraser, 2000) is a relatively new instrument but has been tested various countries and become the most widely used instrument in the field in the last 10 years (Dorman et al, 2002). It highlights current classroom settings and considers academic interest in constructivism. It covers student cohesiveness, teacher support, involvement, investigation, task orientation, cooperation, and equity with a five-point frequency response scale. In Korea, Lee and Fraser (2001) used it for secondary science classes with *CLES*. Dorman et al., (2002)
used *WIHIC* in its cross-national study of Canada, Australia and the United Kingdom. It accounts for the variance of academic efficacy and of attitudes towards science. Fraser and Lee (2009) administered the instrument to Korean secondary school students, and confirmed its validity and reliability.

**TIMSS**: Designed for large-scale comparative studies, TIMSS has been known for the highest standards of quality for conducting research into educational practices and outcomes (Robitaille, 1998; Martin, Mullis & Foy, 2008). It covers broad ranges from students’ attendance of school, students’ backgrounds, mathematics and science in school, computers to in-depth consideration of students themselves (TIMSS 2007 Student Questionnaire). Highlighting science parts, the questions are about individual students’ attitudes towards science, reasons of learning science and activities in science classrooms. The questionnaire for the science teachers also includes their typical class instructional activities, which can triangulate whether the students’ and teachers responses’ are congruent or not. Forty-eight countries (in 2007) including Ontario Canada and Korea and thousands of the fourth and eighth grade students participated in TIMSS since 1995. Because the questions on students’ learning environment of science and achievement assessment were referred from TIMSS questionnaire, further explanation would be present in the measurement instrument section.

### 3.5 Learning Environment and NOS Studies

#### 3.5.1 Behaviouristic science education and research on NOS

When behaviourist and cognitive learning theories prevailed in learning environments, the dominant instructional strategies were teacher-centered (Fraser & Walberg, 1995). Lectures and teacher demonstrations emphasized memorization and retention of the learned contents. Students were required to be quiet, passive receptors of knowledge (Fraser &
Walberg, 1995; Lawson & Suurtamm2006). A teacher-centered learning environment means “the teacher being active and the students being less active, but not necessarily intellectually passive” (Fraser, & Walberg, 1995, p.70). During those days, most research on NOS employed experimental designs and standardized quantitative paper-and-pencil instruments: for instance, the *Test on Understanding Science (TOUS)* (Klopfer & Cooley, 1961), the *Nature of Science Scale (NOSS)* (Kimball, 1967) and the *Nature of Scientific Knowledge Survey (NSKS)* (Rubba & Anderson, 1978).

After reviewing the literature on NOS research Kimball (1967) announced that for “whatever reason, American secondary school students typically have not acquired a valid understanding of what it meant by the nature of science” (p.110). The studies he reviewed criticized the teachers and teacher educating programs for not teaching NOS to their students. Klopfer and Cooley (1961) used *TOUS*, which includes an analysis of scientists at work and of diverse domains including the history and philosophy of science. They examined the relationships between teachers’ and students’ understanding of NOS. Mackay (1971) tested Australian secondary school students’ changing understanding of NOS after two years of three traditional physics courses and two pilot physics courses. The students regarded scientists as specifically able people. He found overall no significant differences between the experimental and control groups in their grasp of the creativity of scientific knowledge, of the role of scientific models, and of the distinctions between theories, hypotheses, and laws. These studies illustrate that teachers are the most important factor for students’ understanding of NOS. On the other hand, based on his own broad review of *TOUS*, *NOSS*, and *SPI*, Aikenhead (1973) pointed out these measuring instruments are problematic. He argued that purely quantitative measuring instruments do not reflect the actual learning activities and goals of science
education. Accordingly, he recommended mixed methods of measuring understanding of NOS.

### 3.5.2 Constructivist Science Education and Research on NOS

Since the late 1980s, the paradigm for teaching and learning environments has shifted from behaviorism to constructivism (Bates, 1999; Fraser, 1998; Matthews, 1994). Relatively new teaching strategies in the laboratory and classroom have been employed, in which students are becoming the center of learning (Bates, 1999; Fraser & Walberg, 1995; Sfard, 1998; von Glaserfeld, 1990). Student-centered instructional strategies are the opposite of teacher-centered instruction. Students construct their knowledge and are expected to participate in learning activities (Sfard, 1998; Vygotsky, 1978). These instructional strategies include laboratory, inquiry-based learning, small-group activities and individualized learning. Teachers are expected to treat their students as co-workers in classroom learning activities (Fraser, & Walberg, 1995). A key tenet of constructivism is that knowledge is not transmitted directly from a knower to a learner, but is actively built and created by each learner (Cobb, 1994; Driver et al., 1994; Lin & Hsieh, 2001).

Adopting constructivist’s learning theory, newer NOS research environments have emphasized learner-centered inquiry, lab activities and diverse approaches. Also, newer instruments for measuring NOS have been developed with open-ended questions: *Views on Science-Technology-Society-Technology (VOSTS)* (Aikenhead, Fleming & Ryan, 1989), *Views of Nature of Science (VNOS)* (Lederman et al., 2002), and *Student Understanding of Science and Scientific Inquiry (SUSSI)* (Liang et al., 2005, 2008).

The effects of constructivist learning strategies in enhancing students’ understanding of NOS are controversial (Driver et al., 1996; Lederman, 1992). Lederman reports that involving
students in an inquiry-based classroom rarely results in changes in their views of NOS. He
observes that students’ beliefs about science and their views of themselves as science learners
may not consistent with goals for scientific inquiry and inquiry-based science learning. On the
contrary, NRC (1996) and Smith (2000) claim that constructivist inquiry-based science result
in informed student views. In Smith’s (2000) study, two classes of sixth-grade students were
divided into experimental and control groups. The constructivist inquiry-based learning class
students practiced designing their own experiments, generating their own problems,
collaborating with fellow students and using multiple representations to learn concepts. The
control group students learned science with the conventional textbook- and lecture- based
methods. The researcher found that students working in the inquiry-based constructivist
classroom had acquired a concept of the diverse aspects of scientific knowledge and social
interaction of science rather than a dichotomous concept of mere right-or-wrong ideas. The
constructivist class students, he concluded, were better able to delineate the nature and purpose
of scientific experiments. The substantial group work of the class, with its many opportunities
for the exchange of views and the development of shared norms enabled students to accept
diverse aspects of science.

Empirical studies suggest that in order to improve students’ understanding of NOS,
educators must provide an authentic learning environment such as real scientists have when
they engage in open-ended laboratory work (Bell et al., 2003; Kim & Kang, 2007; Matkins et
al., 2002; Schwartz, Lederman, & Crawford, 2004). In open-ended lab activities, students
practice diverse trial-and-error experiments, for which there is no right-or-wrong answer.
There is no cookbook style of step-by-step procedures; students interpret the results and
conceive how to present the results (Hodson, 1998; Meichtry, 1992). Kim and Kang’s (2007)
study employed hypothetico-deductive experimental procedures for middle school students in Korea. After a series of experiments, students had more sophisticated views of theory-laden observation, and could distinguish the meaning of scientific theories and laws. In contrast, the study of Schwartz et al., (2004) placed pre-service secondary science teachers into a 10-week science research internship course. The participants who had the greatest amount of research experience showed little or no gain in their understanding of NOS as well as a poor overall understanding of NOS themes. Meichtry (1992) implemented BSCS (Biological Science Curriculum Study): a program representing NOS themes more explicitly, with different grade levels (6th, 7th, and 8th) and across wide ranges of different schools’ biology classes. The result showed no significant improvement in the experimental group; the creative, tentative, and developmental characteristics of NOS were inadequately understood. Meichtry concluded that the mere use of a program designed to develop student understanding of NOS was no guarantee that such understandings would develop. Furthermore, students can gain the mistaken impression that more experimental data are a way of improving the accuracy of results and the origin of knowledge (this reflects empiricist and inductivist assumptions).

Many studies have attempted to identify students’ perceptions on their learning environment and the relationships between the perceptions and learning outcomes. Regardless of the instruments researchers used, students who positively evaluated their learning environment tended to achieve higher than those who did negatively. Studies on NOS were dependent on the major tends of teaching and learning theories. When behavioristic perspectives were dominant, students’ understanding of NOS was measured close-ended questions, which strongly reflected developers’ ideas. On the other hand, as recently constructivist-learning theories are prevailed, NOS studies have used flexible formats that
provide more rooms for students’ ideas. This literature review provides a foundation that students’ perceptions of their learning science should be included as a significant factor for understanding of NOS.

3.6 Review of Achievement Assessment

An assessment is a decisive factor for students in determining how to study (learning strategies), what to study (content) and how much to study (Biggs, 2003; Ross & Siegenthaler, 2006; Scouler & Prosser, 1994). However, little has been done on the relationship between achievement assessment and students’ understanding of NOS. This section takes a general overview of student achievement assessment considering how it affects student learning and how students perceive it. This would in turn explain why the present research considers achievement assessment such an important factor in students’ understanding of NOS although there is not yet sufficient evidence of its actual effects. The review foci were how students perceive achievement assessments and their learning strategies and how large-scale and high-stake assessments affected teaching and learning activities.

3.6.1 Achievement Assessment Formats, Content and Learning Strategies

The goals of an assessment for both schools and students are to ensure competence, provide feedback, guide learning, and evaluate curriculum (Brown & Hirschfeld, 2008; Fulcher, 2009; Hall & Øzerk, 2008). Some gaps exist between the intentions of an assessment and students’ conceptions of an assessment (Brown & Hirschfeld, 2008; Fisher, Waldrip, & Dorman, 2005). Students regard assessments as a judgment on their learning rather than a means to improve and help it. In addition, students adjust their learning habits to handle assessment formats (Biggs, 2003; Entwistle & Entwistle, 1992; Scouller, 1998; Struyven,
Dochy & Janssens, 2005). Struyven et al. (2005) also emphasize the importance of assessment in students’ approaches to learning:

[T]he learner’s experience of evaluation and assessment determines the way in which the student approaches (future) learning. Assessment is thus logically, but also empirically, one of the defining features of students’ approaches to learning. (p. 326)

The more students endeavor to achieve high score, the more conscientiously they adjust their ways of study.

Formats of an assessment are a decisive factor in student learning (Biggs, 2003; Crooks, 1988; Entwistle & Entwistle, 1992; Scouller, 1996, 1998). According to Crooks’s (1988) report, when students were required to memorize a particular fact, they focused on short-term memorization; when essay-type questions were given, they tried to integrate concepts and to study the deeper meaning of the concepts. Biggs (2003) also reports that when assessments had a multiple-choice format, students’ strategies to prepare for a test were memorization, while when assessments were of an essay type, students tried to search for more information to prepare for them. Entwistle and Entwistle’s (1992) and Scouller’s (1996) studies obtained similar findings. Students were influenced significantly by the formats of assessment. Multiple-choice questions emphasizing factual knowledge pushed students to memorize and reproduce accurate details while open, essay-type questions encouraged students to think critically. In Scouller’s study (1998), students were asked about their learning strategies in preparation for a multiple-choice test and an essay assignment in a course. Preparation and study strategies for multiple-choice types involved surface learning (memorization and reflection) while preparation for an essay assignment required deep learning strategies (comprehension and application of the knowledge). Tang’s study (1992) showed no
differences in test outcomes, but a shift from surface to deep learning strategies to meet essay type assignments.

Additionally, the formats and use of a classroom assessment determines students’ satisfaction with their school experience (Hall & Øzerk, 2008; Karagiannopoulou & Christodoulides, 2006; Sambell, McDowel, & Brown, 1997). In Sambell et al.’s study, students regarded traditional paper-and-pencil based assessment as irrelevant and arbitrary because such tests only measure their memorization ability. Alternative methods such as portfolio and performance assessments were believed to be fairer because the methods measured diverse skills, knowledge and abilities. If they are impressed that an assessment is not fair, or does not reflect their learning and diverse abilities, students think it irrelevant or useless. Thus, not only what, why and when, but also how a teacher assesses students’ learning bears a significant impact on the quality of what they learn and how they study. What and how is assessed becomes crucial to students’ learning and mastery of a subject domain.

Studies on the relationships between an assessment measuring different levels of intellectual ability and skill and students’ learning strategies have recorded positive correlations (Entwistle & Tait, 1990; Scouller, 1998; Tang, 1992). That is, if an assessment employed lower levels of cognitive abilities (e.g. Bloom’s taxonomy), students prepared for it by memorizing and recalling content; if an assessment examined students’ higher levels of cognitive abilities, students approached their preparation with analysis, synthesis and application of content. Relevant to the present study are the relationships discerned between students’ perceptions on their assessment and their attitudes toward the subject area. As with assessment format, Scouller (1998) study reported that students’ perceptions of what would be assessed (either low or high levels of intellectual ability) were positively related to their
learning strategies. When students perceived their assessments would test low levels of intellectual ability they prepared factual knowledge while when they perceived their assessment would address high levels of cognitive process, their learning strategies were oriented to the synthesis and application of knowledge.

Comparing the levels of achievement in a test shown in high or low scores with strategies of learning, studies have reported inconsistent results (Scouller, 1998; Scouller & Prosser, 1994; Tang, 1992). In Scouller (1998) the strategies of memorization of factual knowledge gained higher scores in multiple-choice tests; however, Tang’s study (1992) did not show any difference among test preparation strategies.

3.6.2 Large Scale and High-Stake Assessment in Teaching and Learning

Nation-wide or province-wide standardized assessments have been shown to determine teaching practice (Shohamy, 2001; Struyven et al., 2005). Shohamy (2001) argues that centralized governments use a standardized large-scale assessment (SLA) to control educational systems by defining what kind of knowledge is prestigious. In these systems “the primary goal is to make teachers teach and students study specific topics” (p. 34). Teachers become concerned with what is covered on the test; in effect, the SLA becomes the “curriculum” because teachers adjust their instructions to enhance test scores rather than to respond to student interests and needs. Such an assessment forces a teacher to narrow the curriculum to fit the test, and to concentrate on memorization rather than critical thinking and creativity. In other words, the centralized test policy affects teachers’ practice and students’ learning and perceptions on a subject. When the primary purpose of high-stakes tests is to filter students for higher education, the control of achievement assessment becomes extremely powerful.
According to Ben Jaarfar and Anderson (2007), the current trend in Ontario schools is to expand LSA through the work of the Education Quality and Accountability Office (EQAO). EQAO testing is conducted for Grades 3 and 6 reading, writing and mathematics, for Grade 9 mathematics, and for Grade 10 literacy, in order to track students’ basic literacy and numeracy. The consequences of EQAO tests for students up to Grade 9 are not severe, but EQAO’s Grade 10 literary exam is a high-stakes test. Its results comprise a significant percentage of a secondary student’s final grades, and also serve as a graduation requirement or a compulsory entry requirement for university attendance in Ontario, New Brunswick and Quebec (Volante & Ben Jaafar, 2008). As a result, teachers teach test-taking skills. Many educators feel a pressure to secure good results in EQAO tests because the raw scores of the test are reported in local newspapers commenting on school environments.

Like EQAO, Korea has similar LSAs: the Basic Achievement Tests (BAT) and Suhakneonglyuk. The BAT is conducted for Grade 3 literacy and numeracy, and for Grades 6, 9 and 10 Korean language, mathematics, social studies, and science. The BAT is not a high-stakes test. It is regarded as an important formative assessment for students and a resource for policy makers (Lyu, 2008). Students, teachers, and parents do not feel much pressure from the results. However, the Suhakneonglyuk test, similar to the Scholastic Aptitudes Tests in the United States, is an extremely high-stakes test. It cannot be overstated that all students and all school education for K-12 students focus on and prepare for this test (Kang, 2005; Kwak, 1998). High schools in particular teach what will be in this test and teach skills for taking this test. Teachers must cover the possible content of the test and transfer their knowledge to students so to find the correct answers. Students rely on the authority of teachers and textbooks; teachers are required to demonstrate their mastery of the knowledge and skills prescribed in
the standardized curriculum (Kang, Cha, & Kim, 2008). Critics note that the test emphasizes an extremely objective evaluation using the function of judgment; thus, it largely measures how much knowledge a student remembers, not his problem-solving, creativity or logical thinking abilities (Kwak, 1998).

3.7 Instruments of Measurement for Student Perceptions of Assessment

There are not sufficient measuring instruments for students’ perceptions of assessment tasks (Cabanagh et al., 2005; Fisher et al., 2005; Schaffner et al., 2000; Zhang & Burry-Stock, 2003). In this section, four instruments are reviewed, which measure how students perceive their achievement assessments and how teachers perceive their skills in the practices of assessment in their classrooms.

Schaffner et al. (2000) developed an instrument called Perception of Assessment of Teachers by Students (PARTS), and sought to validate it with 174 students. From Kindergarten through Grade 3 they used a pictorial scale and for Grades 4 through 12 they employed a senior version with a 5-point Likert scale. The instruments asked students to respond to 55 questions on “how you feel about the way your teacher finds out how much you have learned”. The factors they extracted were: 1) Internal Locus of Control, 2) Fairness Issues, 3) Positive Grade Impact, 4) Teacher’s Job, and 5) Negative Grade Impact. They found that as participants’ grade level became higher, students were more aware of how their teachers assessed them. The study established the reliability and validity of both the pictorial scale and the senior Likert scale.

Zhang (1995) and Zhang and Burry-Stock (2003) developed an instrument, Assessment Practice Inventory (API), for teachers’ self-reported perceptions of competence in assessment practice, which was intended to reflect the Classroom Assessment Standards (American
Federation of Teachers, National Council on Measurement in Education & National Education Association, 1990). It measures teachers’ perceived assessment skills concerning the development of instrument, the use of various types of instruments, the analysis of items and ethics using Likert-type scales from one to five. It consists of 67 questions with 7 scales: 1) perceived skill in using paper pencil test, 2) perceived skill in standardized testing, test revision and instructional improvement; 3) perceived skill in using performance assessment; 4) perceived skill in communication assessment results; 5) perceived skill in non-achievement-based grading; 6) perceived skill in grading and test validity; 7) perceived skill in addressing ethical concerns. Zhang and Burry-Stock (2003) found that as grade levels became higher, teachers relied more on objective tests in classroom assessment and showed an increased concern for assessment quality. The item convergent and discriminating validity of API for teachers was established.

The Student Learning Preferences Questionnaire (SLPQ) was tested for its validity and reliability in England (Dorman & Knightley, 2005; Dorman, Fisher & Waldrip, 2006) in Canada (Dorman et al., 2006) and in Australia (Dorman et al., 2006; Fisher, Waldrip, & Dorman, 2005). The study of Dorman, Fisher and Waldrip (2006) worked with 449 students from Grades 8, 9 and 10 at secondary schools in Canada, Australia and England to investigate whether students’ perceptions of their achievement assessment could be a predictor of their academic achievement (academic efficacy and attitudes to science). They used the Student Perceptions of Assessment Questionnaire (SPAQ) consisting of 30-35 items with 5 scales, which had been validated through previous studies (Cavanagh, Waldrip, Romanoski, & Fisher, 2005; Dorman & Kinghtley, 2006; Dorman et al., 2006). They explained the five scales adopted as follows: Congruence with Planned Learning is “the extent to which assessment
tasks align with the goals, objectives and activities of the learning program.” Authenticity means “the extent to which assessment tasks feature real life situations that are relevant to the learner.” Students Consultation includes “the extent to which students are consulted and informed about the forms of assessment tasks being employed.” Transparency concerns “the extent to which the purposes and forms of assessment tasks are well-defined and clear to the learner.” Diversity means “the extent to which all students has an equal chance at completing assessment tasks” (p. 9). They adopted a four-point Likert response format for each item (Almost Never, Sometimes, Often, and Almost Always). The reliability coefficient of Cronbach alpha was satisfactory in internal consistency. The results showed students’ perceptions on their assessments predicted academic performance.

The Perceptions of the Assessment Demands Questionnaire (Scouller, 1998; Scouller & Prosser, 1994) has been used to measure what and how students or teachers perceive assessments demand. In their questionnaire scales, Scouller and Posser classified learning into surface and deep learning. They assumed that depending on an assessment type, students’ learning approaches would be different. For instance, multiple choice questions or simple answering questions can make students’ learning strategies emphasize rote study and reproduction of key terms and definition to find a correct answer, so students would spend more time on memorization rather than problem solving, critical and creative thinking or further development of related knowledge. On the contrary, essay or project types of assessment tended to make students’ learning approaches integrate from a variety of sources and spend their time on obtaining more information relating interesting topics of class. Thus, the questionnaire consists of 12 items with a 5-point Likert response format. Two scales are deduced from the questionnaire: one for low-level surface skills and another for high-level
intellectual skills. For instance, a question asking the perception of surface learning skills is “I expect the Over All Test to assess my ability to reproduce key terms and definitions.” On the other hand, a question for deep learning skills is “I believe a test should assess the student’s ability to integrate from a variety of sources.” Several different studies have established the questionnaire’s reliability and validity (Segers, Dierick, & Dochi, 2001; Segers et al., 2008).

In short, assessments are decisive in students’ learning as well as teachers’ teaching practices. Studies revealed that when an assessment emphasized factual knowledge, memorization, or choosing one correct answering formats, the strategies in learning were surface. In contrast, when an assessment measured understanding, diverse views on a topic, or writing an essay, students tried to search diverse sources beyond the class covered and deep learning strategies were used. Large scaled assessment and high stakes assessment strongly confine teachers’ teaching practice. Teachers adjust the content and teach test taking skills. Therefore, how to measure and what to measure for students’ learning outcomes should be considered when a study examines how students’ perceive a subject and what can be the outcomes of learning. Accompanying with the perceptions of learning science, the perceptions of assessment should be important factors how students view science and scientific knowledge.

### 3.8 A Theoretical Model for This Study

The major goals of this research were to identify how students perceived their learning activities and tasks (PLAT), assessment formats (PAFORMAT) and content (PACONT), and their understanding of NOS, and consequently to determine which of these factors might affect students’ understanding of NOS.

Previous literature has pointed toward several potential factors in formal education. Carey and Smith (1993), Clough (2006), Duschl (1990), Roth and Roychoudhury (1994) argue
that the way science is presented to students is a decisive factor how students perceive science. Bell and Lederman (2003) also argue, “Students experience a wide range of direct instruction and conformational, cookbook-style laboratory experience in their [students’] science instruction. It is not surprising that in such an environment, students often develop the misconception that scientific knowledge is portrayed as the results of steady and unproblematic accumulation of confirmed hypotheses” (p. 369). Kuhn (1970) and Laudan (1987) criticizing the conservative science textbooks, argue that school science allows students to learn science that has been already established and ready-made knowledge rather than to challenge or to create new ways of thinking or solving problems, so students become a well trained puzzle solvers in a normal science.

In accordance with the NOS studies, Canadian studies on students’ learning environment (Dorman, 2004; O’Reilly, 1975) have reported that students who positively evaluated their learning environment achieved higher than did those students who negatively evaluated it. In agreement with these studies, Korean research has also reported similar results: positive evaluators of learning environment have attitudes toward science (Kim & Kang, 2007; Fraser & Lee, 2009).

Perceptions of the methods and content of their assessments also had a decisive influence on students’ learning strategies. The use of close-ended multiple-choice formats emphasizing factual knowledge tends to promote surface learning, while essay-type assignments emphasizing knowledge construction enhances deep learning strategies (Biggs, 2003; Black & William, 1998; Crooks, 1988; Scouller, 1998). Especially a high-stakes test administered in nationwide or provincialwide strongly changed not only students learning strategies but also teachers’ instructional approaches (Ben Jaafar & Anderson, 2007; Kang,
2005; Lyu, 2008). These research results imply that students’ perceptions of assessment can be an effective indicator to predict learning outcomes.

This research attempted to identify the relationships of these factors with student understanding of NOS. The following assumptions were postulated:

(1) If science class activities and tasks underscore knowledge transmission from an authoritative “knowers” (science teachers) to passive “non-knowers” (students), students could come to regard scientific knowledge as a body of factual knowledge;

(2) If science class activities and tasks encourage students to challenge or to create new and different ways of thinking and solving problems from previous ways, students could come to see that science is both science-in-the-making methods and knowledge about pain taking efforts and about enthusiasm fierce struggles involved in an interpretive effort to understand natural phenomena;

(3) If students’ learning of science is assessed so as to select only one “correct” answer, students could gain the perceptions that scientific knowledge is about the “correct” statement of absolute truths; whereas, if science tests permit more than one correct answer, students could come to a more pluralistic view of science; and,

(4) If the learning content addressed in science assessment emphasizes the knowledge of science and treats empirically supported theories as proven facts, students presumably perceive science as a body of accumulated knowledge that is difficult to change.

These assumptions provide a theoretical model that describes the relationships among these constructs. The constructs, PALT, PAFORMAT and PACONT are set as predicting variables (exogenous variables) and the concepts of NOS are regarded as predicted variables
(endogenous variables). The study therefore measures how much these three constructs can predict the variances of the NOS concepts.

Figure 3.1 represents relationships between the observed variables and their targeting constructs, as well as relationships among the three constructs and the concepts of NOS. Since this research is not experimentally designed, a significant relationship, if any, cannot be confirmed as cause and effect. In the figure, the blue circles represent constructs and the yellow rectangles show observed variables. The arrows connecting the blue circles and the yellow rectangles (observed variables) are outer loadings. PLAT, PAFORMAT and PACONT are defined by the observed variables (formative measurement models) and the magnitudes of the outer loadings indicate how much an observed variable contribute to the construct. The five NOS concepts control the observed variables (reflective measurement models) and the outer loadings were reflected the concepts. The inner coefficients connecting blue circles (constructs) depict the structural model (relationships among latent variables). The coefficients correspond to beta (slope) in a regression analysis.

Figure 3-1. Theoretical Model
3.9. Chapter Summary

This study examined what factors affect students’ understanding of NOS. The relation of learning to achievement assessment was highlighted. Hence, the first part of this chapter included a literature review on learning environment studies and instruments. Some Canadian and Korean studies were reviewed. The second part discussed how students’ achievement assessment affects student learning and teacher instruction, and then reviewed instruments, which measure student and teacher perceptions of achievement assessment. No NOS studies have yet examined relationships between understanding NOS and the types, content or student perception of achievement assessments. Thus, this review gave only a general overview of the Canadian and Korean external assessment systems. The next Chapter will show how this research was conducted to gain the answers needed to achieve this goal.
4. METHODOLOGY

4.1 Introduction

This chapter outlines the methodology for this research. A mixed method (Creswell, Plano, & Clark, 2007; Greene, Caracelli, & Graham, 1989) was employed using a survey questionnaire and semi-structured interviews. The data were collected from Canadian and Korean Grade 8 students and analyzed with the methods, which this chapter explains. This study sought the answers to the following three major research questions:

1. What are Canadian and Korean junior middle school students’ perceptions of their learning activities and tasks (PLAT), their perceptions of assessment formats (PAFORMAT) and content (PACONT) and their understanding of NOS?
2. To what extent are the students' perceptions similar or different in the two contexts?
3. What are the relationships, if any, among the PLAT, PAFORMAT, PACONT and their understanding of NOS?

This chapter is organized into eight sections: 1) the first section introduces and outlines the research design; 2) the second section gives a general descriptive overview of the contexts for the research, i.e. typical Canadian and Korean junior middle school classrooms and the participant schools; 3) the third section explains the data collections focusing on the instrument for this study; 4) the fourth section explains the examination of the instrument validity and reliability; 5) the next section discusses data analysis strategies; 6) the sixth section deals with quantitative and qualitative data integration methods; 7) the some ethical considerations was followed; and 8) the final section summarizes the chapter.
4.2 Rationale for a Mixed Method and Research Design

Considering the advantages and disadvantages of quantitative and qualitative methods across two distinct research contexts, it seems best to adopt a mixed method. The two strands of methodology can compensate for each other, and the synergy of their strengths can achieve a much closer description of reality for this research. Empirical research on NOS (Aikenheads 1987; Liang et al., 2008) supports the importance of adopting a mixed-method. For instance, Aikenhead's (1987) study illustrates why NOS studies adopt a mixed method; at times responses to close-ended questions and open-ended written responses prove inconsistent. In such cases, open-ended questions can give clues as to whether a respondent understood the question correctly or not, as well as providing keys to further information. Krathwohl (1997) also stresses that open-ended elements in a study make participants come alive. Follow-up semi-structured interviews brought in-depth understanding of participants' points of views on NOS, and educational culture differences (Lederman et al., 2002).

Standardized close-ended questions can cover broad ranges of the research domain, and reduce the effects of respondents' other erroneous abilities such as writing abilities or cultural differences. Brown (2007) identified important cultural differences in writing and answering questions. For example, Asian cultures value indirect descriptions while Western culture prefers a direct approach when students answer open-ended questions. Liang et al. (2008) pointed out that in a questionnaire that consists largely of open-ended questions, many blank answer sheets or answers of only a few words could undermine the quality of a study. The standardized, close-ended questions, however, cannot fully explain the reasons of students’ answers. Further, if a participant has different ideas or may not understand questions properly, his/her random answers can lead to a biased conclusion.
On the other hand, open-ended questions allow researchers a more comprehensive understanding of students’ views, reasons and resources informing the beliefs that students have and the way in which students’ views affect their learning (Aikenhead, 1987; Lincoln & Guba, 1985). In spite of these benefits, interview or open-ended questions cannot include as many participants as the standardized close-ended questions can. The richness of interview data and the answers of open-ended questions can also be disadvantages when summarizing the data in a concise form. Categorizing, sorting and coding data not only consume time and energy but also have a high probability to introducing biased results (Sudman, & Bradburn, 1983). Further, for open-ended questions, which require students to write full sentences with paragraphs can be affected by writing skills and cultural differences too.

This research made its investigation three phases (Figure 4.1). Phase I involved the recruitment of samples and administrating the survey questionnaire (Refer to Appendix A). In Phase II, quantitative data analyses were performed using statistical software, Excel, SPSS and SmartPLS. The quantitative data analyses examined 1) the instrument’s validity and reliability, 2) the differences between two countries and among schools and 3) the associations between the perceptions and understanding of NOS. In Phase III, semi-structured interviews were conducted and analyzed, and auxiliary data sources were consulted (See Appendix B for interview protocols for students). Figue 4-2 presents research questions and related data sources as well as analytical methods.
Figure 4-1. Research Design and Process
<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Data</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Question 1:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. What are the students’ perceptions on learning activities and tasks and</td>
<td>-Survey answers</td>
<td>Quantitative data</td>
</tr>
<tr>
<td>assessment formats and content?</td>
<td>-Examples of science tests</td>
<td>a. Descriptive statistics</td>
</tr>
<tr>
<td>b. How do the students understand NOS?</td>
<td>-Students’ interviews</td>
<td>(means, frequencies, SD)</td>
</tr>
<tr>
<td><strong>Purpose:</strong> To identify characteristics of each research setting’s learning and assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Research Question 2:</strong></td>
<td>-Answers to the close-ended questions</td>
<td>Quantitative data</td>
</tr>
<tr>
<td>What are the differences/similarities between Canada and Korea?</td>
<td>-Answers to the open-ended questions on NOS</td>
<td>a. Independent samples t-tests</td>
</tr>
<tr>
<td><strong>Purpose:</strong> To examine differences of the three domains between the countries and among schools</td>
<td></td>
<td>b. Multivariate Analysis Of Variances (MANOVA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Software: SPSS/Excel</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Qualitative data</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Coding (identifying major themes and categories based on literature review)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Conceptually clustered matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Research Question 3:</strong></td>
<td>-Answers to the survey</td>
<td>Quantitative data</td>
</tr>
<tr>
<td>Can the perceptions of learning and of assessment predict student understanding of</td>
<td>-Students’ interviews</td>
<td>a. PLS</td>
</tr>
<tr>
<td>NOS?</td>
<td>-Examples of science tests</td>
<td>b. Software: SmartPLS</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td></td>
<td><strong>Qualitative data</strong></td>
</tr>
<tr>
<td>To examine relations between the independent variables and dependent variables.</td>
<td></td>
<td>a. Coding (major themes and categories)</td>
</tr>
<tr>
<td>To verify the factors for understanding of NOS</td>
<td></td>
<td>b. Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Conceptually clustered matrix</td>
</tr>
</tbody>
</table>

Figure 4-2. Research Questions, Data and Analysis
4.3 Participants

In total 536 Grade 8 students from 4 schools in Toronto Canada and 3 schools from Busan, South Korea participated in the study voluntarily. As the concepts of NOS are abstract and require a high-level of thinking about the epistemology of science, the questions, especially the open-ended questions, were fairly demanding. If participants are at very low grades, a considerable number of them may answer some questions randomly. This could create false results and distort the reality. In Korea, Grade 9 students are the highest grade in a middle school so they are ready to be streamed into Academic or Vocational schools. In Canada, Grade 9 students attend either Applied or Academic Science classes. If one group of participants is from an Academic class and another from a Vocational school or an Applied class, the result of this study could reflect not the general school students but specially oriented class or individuals. For these reasons, Grade 8 students were recruited to participate in this research.

Regarding the number of participants, Garson (2010), and Schumacker and Lomax (2010) suggest that the minimum number of participants should be is five times of the number of observed variables. Generally the structural equation modeling (SEM) requires a large number of participants; Schumacker and Lomax (2010) recommend ten times the number of participants to establish external generalizability. This study contains 27 close-ended observed variables; thus, the required number of participants was satisfied in total. However, the samples consisted of two countries, and when the analyses were done in individual countries, the sample sizes did not fully meet this condition. To meet the 10 times sample size, each country should have more than 270 participants. Due to the difficulties in recruiting participants, the Canadian sample was 217, which did not meet the condition.
The participants were recruited by a convenient sampling method (Warner, 2008). Having difficulties in recruiting volunteers randomly, personal acquaintances and friends who were working in junior middle schools were asked to participate in the research or to introduce it to their school science teachers. Because of this convenient sampling, the sample cannot represent of the population of all Canadian or all Korean Grade 8 students. Therefore, caution must be used in seeking to generalize the findings of this research beyond its own context.

4.3.1 Canadian Participants

The total effective number of participants from Canada was 217. Ten classes from 4 schools of Toronto participated in the study. School 1 is a religion-based school and 53 participants were from two classes. School 2 was a public school and 54 students participated from two classes. School 3 was also a public school and 2 science teachers and with their students (65 students from 3 classes) volunteered. Forty-five students from School 4 participated. The participation rates of all 4 schools were over 90%. Twenty-one cases were excluded because more than 6 items were missed or outliers (if two constructs were located beyond 3 SD, the cases were excluded).

4.3.2 Korean Participants

The total effective number of participants from Korea was 319. Thirteen classes from 3 schools participated in the study. The schools are public schools and located in Busan (the second largest city in South Korea with a population size similar to Toronto). More than 90% of students in the classes participated in the study. School 5 was a boys’ school while the other two schools were for both boys and girls. Therefore, an even gender distribution was not established. Seventy-two cases were excluded since they missed more than 6 items or outliers.
4.3.3 Interviewees

The first condition for selecting interviewees was students’ willingness to join the interview; the second condition was diversity in their survey responses. When the survey questionnaires were collected, I chose students who wrote the answers sincerely and gave the list to the science teachers. The number of interviewees for each school was about 7 to 10 students except School 2. However, when actual interviews were attempted, recruiting enough number of interviewees was very difficult. Although many students wrote their names on the survey, they did not want to join in the interview when I actually asked. So I asked some help to the science teachers. Science teachers recommended and encouraged their students to participate in the interviews. School 1 teacher recommended interviewees because the interviewees involved active learning during class time, not because the students were high achievers. I visited the school and did interviews in the school science lab. It took about 25 minutes for each interviewee.

For the Korean interviewees, I selected 17 interviewees (8 from School 5, 4 from School 6 and 5 from School 7) and gave the list to the science teachers. The science teachers asked me to give general questions and my email address, so they distributed them to the students on the lists. Only 5 students emailed me from School 5 and School 7; thus, the number of Korean interviewees is 5. Four out of 5 interviewees wanted to apply for a science focused high school, which means they were within upper 1% in all subjects. This selectiveness of the interviewees limits in their representativeness to the population. The interviewees responded to the questions and probing questions were done two round emails and online chatt.

A total 9 students participated in the semi-structured interviews. Two Canadian
students (one boy and one girl) from School 1 and two students (both girls) from School 4 participated. Some students from School 3 wrote their names on the survey, but when contacted through their science teachers, they declined to join. The time in the school year might have prompted this refusal. The survey was collected in early June 2011, and interviews began just before summer vacations. The teachers at School 2 agreed to conduct the survey only on a condition of not participating in the interviews. Three boys from School 5 together with one boy and one girl were from School 7 served as Korean Interviewees. Students from School 6 who wrote their names on the survey, also declined to participate once contacted. Still their surveys were effective. The withdrawing policy was stipulated in the consent and ascent forms.

4.4 Research Context

This research context provides an overview of the two educational systems emphasizing classroom features and NOS in science curricula. This should highlight the potential factors affecting student understanding of NOS. Regarding the cultural differences between Canada and Korea, whether or not one accepts Universalist or Multiculturalist claims of knowledge generation (Matthews, 1994; Siegel, 2002), comparing students’ understanding of NOS between the relatively mono-cultural learning environments of Korea and the multi-cultural learning environments of Canada will be meaningful in terms of their perceptions of cultural aspects of knowledge formation. Also reviewing the effects of high-stakes assessment and of teacher-student interactions in learning activities will be helpful to clarify the research context.
4.4.1 Canadian Context

4.4.1.1 Typical features of Canadian junior middle school classrooms

Students in Canada’s classrooms come from many different countries and ethnic groups. According to documents from the Ministry of Education in Ontario (MoE Ontario) (2007), typical junior middle school classes consist of students whose parents represent many different ethnic groups and nationalities. Province-wide, about 22% of Ontario students learn a first language at home other than English; in urban schools, such as those under the Toronto District School Board, the range is from 20% to 75%. The distribution of students’ first languages implies a diversity of ethnic, national and cultural minorities in a classroom. TIMSS 2007 (Martin et al., 2008) data also support the diverse ethnicity of a class, particularly in Grade 8, reporting that 57% of students were not born in Canada. Despite these ethnic and cultural differences, 54% of Canadian Grade 8 science teachers did not think that diversity placed limitations on their instruction while 16% believed that student factors placed high limitations on their instruction. This may be because teachers are obligated to consider their students’ differences in their instructions (Aikenhead, 1973, Ben Jaafar & Anderson, 2007; Volante & Ben Jaafar, 2008).

Class sizes are about 25~30 students and the number of girls and boys are balanced (MoE Ontario, 2007). According to TIMSS 2007, the average class sizes were 26, smaller than in previous assessments in Ontario (30 in 1999). Ninety-six percent of schools and households had computers with the Internet connections, so students were available in their schools and at home.

TIMSS 2007 reports (Martin et al., 2008) show that about 45% of Ontario students have about 3 assignments a week and spend 30 minutes on each assignment; 55% students
have less than two. More than 50% of students have positively valued science and have positive attitudes towards science. When asked about their self-confidence in learning science, about half of Canadian Grade 8 students expressed full confidence while 18% boys and 14% girls expressed low self-confidence.

### 4.4.1.2 NOS in Science curriculum

According the report of TIMSS 2007 (Martin et al., 2008), science class was 96 instructional hours, about 11% of total school learning hours in Ontario Grade 8 curriculum. A general science curriculum was employed across the whole province (MoE Ontario, 2007; Martin et al., 2008). Students learn science from a science teacher not from different teachers of biology, physics, chemistry and earth science. Although the four elements of the curriculum were not separated, the learning content covered evenly the four domains: biology (69): chemistry (57): physics (68): earth science (73). The Ontario Ministry of Education provides science teachers with Ontario Science Curriculum, but teachers probably have flexibilities in teaching its content to their students. Of course, the textbooks are not absolute yardsticks if teachers follow the curriculum, but the use of textbooks only as a reference. Only 43% of teachers use science textbooks as the primary basis for their lessons while 54% use them as supplementary resources; the rest do not use textbook at all. Decentralized curricula and educational systems (e.g., diverse textbooks, term/semester systems) result in a diversity of individual schools or districts.

Ontario’s science and technology program seeks to make future citizens scientifically and technologically literate (MoE Ontario, 2007). NOS is explicitly mentioned in the curriculum: “Science as a way of knowing that seeks to describe and explain the natural and physical world. An important part of scientific and technological literacy is an understanding
of the nature of science” (MoE Ontario, 2007, p.4). The fundamental concepts of science from Grade 1 to 12 are sequentially developed covering matter, energy, systems and interactions, structures and function, sustainability and stewardship and change and continuity. The sequentially designed curricula help students expand and deepen their understanding and application of science with each higher-grade level. NOS concepts are embedded in the fundamental concepts of the curricula, so teachers need the pedagogical content knowledge of NOS to intertwine scientific concepts and NOS properly. The achievement chart also explicitly states that a successful NOS student “makes connections between science, technology, society, and the environment with considerable effectiveness” (MoE of Ontario: The Ontario Curriculum, 2007, p. 27). The STSE curriculum, as Aikenhead (2006) points out, reflects Canadian culture in distinction to the Standards and Project 2061 of the United States.

4.4.1.3 Schools in This Research

Three schools (School 2, 3 and 4) were public schools while School 1 was a religion-based small school. Like the TIMSS report, class sizes were from 24 to 32 and all classes consisted of both boys and girls. As to student ethnicities, teachers did not know exactly who was or was not born in Canada. All teachers were females who majored in science. School 1 and 4 teachers explained the school features. Regardless of students’ achievement, all students in the schools learned the same content; no differentiated curriculum was available for advanced or low achieving students in the schools. Students had around three class periods in a week and one teacher taught four subjects of science. There were no pre-set assessments as mid-term or final tests. Science tests were administered when the teacher thought proper; quizzes were held mostly in the middle of a unit, and tests mostly the ends of a unit.
School 1 has one science lab classroom and only science teachers use the lab when they did actual lab experiments. School 4 used science teachers’ homeroom classrooms for lab experiments, and the classes were designed to do experiments.

4.4.2 Korean Context

4.4.2.1 Typical Korean Middle School Classrooms

In Korea, the new school year starts in March. The school system consists of 6 elementary, 3 middle, and 3 high school years, followed by 4 years of university or 2 years of college. The average class size for a middle school in an urban area is 32~34 (MoE of Busan, 2009). According to TIMSS 2007 report, the average class size of nationwide was 37, which increased 1 from the year of 2003 and decreased 7 from the year of 1999. In comparison with the nationwide class size, the schools in Busan District School Board had smaller classes. Almost all public schools adopt a mixed class of boys and girls, but there are some for only boys or for only girls. Ethnic differences and cultural diversities of students are not big issues either for teachers or for students because international students are rarely found. According to TIMSS data of 2003 and 2007, 100% of students were born in Korea and their parents were also born in Korea. It meant that all students speak Korean as their first language.

Geographically South Korea is a small country; one-tenth size of Ontario, so there is no notable difference among provinces.

The top-down curriculum fosters a unilateral learning environment. TIMMS report 2007 showed 73% teachers use textbooks (that should have passed the verification process of MoE Korea) and 24% use as supplementary resources. The dominant instructional strategies are lectures and laboratory experiments, sometimes with computer-assisted activities (Park, Khan, & Petrina, 2009). Regarding the laboratory activities, MoE Busan very strongly
recommends teachers to do the hands-on lab activities at least 30% of the total instructional hours (31 hours out of 104 hours a school year). Science class laboratory activities are highly focused on verifying theories, and using ‘ready-made’ procedures and results (Fraser & Lee, 2009).

Most Korean students (77%) do not have much homework in science (less than 2 times a week), and their achievement scores of TIMSS test do not relate to the amount of homework (Martin et al., 2008). One-third of the students value science highly and their achievement scores are significantly higher (586) than students who give science a medium or low value (526). One third of the students express confidence in learning science while one third do not. The difference of achievement scores between these two groups is large: 603 (highly confident group) versus 516 (low-confidence group).

### 4.4.2.2 NOS in science curriculum

After the Seventh Curriculum Reform in 1998, the concepts of NOS have explicitly been placed in science curricula emphasizing the tentativeness of scientific knowledge, the empirical nature of science and the procedures of scientific knowledge formation (MoE Korea, 2000, 2001). The chapter on NOS, entitled "Exploration", is the first in the Grade 10 textbook, *General Science*. It consists of sections entitled, “What Science Is” and “What a Scientist Does” and “The Relationships between Science and Society”. However, the description of science and scientific knowledge heavily inclines to an inductivist and realist epistemology (Jeon, 2006; Kim, Jeon & Paik, 2007). There is no unit for NOS in middle school science curriculum (MoE Korea, 2007), but the goals of science education include the aim to educate students to understand the mutual relationships between science, technology and society (p. 152).

Jeon (2006) criticizes the gap between goals and actual content in Korea’s science
education. There are goals but no explanation of how to attain them—which is embarrassing for teachers. The actual science class, according to Jeon, has been directed by a product-centered and large-scale assessment-driven curriculum over the past 20 years. Thus, the practice of science education has not met the goal of educating healthy and creative members of society with some necessary scientific literacy. Furthermore, while Korean science teachers think of the science lab curriculum as practical work aimed at enhancing creativity, in the actual lab, teachers barely focus on creativity. This ignorance of creativity in education is explained by the test-driven nature of Korean science education, which forces teachers to focus instruction on the body of scientific knowledge.

4.4.2.3 Schools in This Research

Three schools from the Busan District School Board participated. All students were Korean (with no other ethnicity represented) from public schools. School 5 was a school for boys only while Schools 6 and 7 were for both girls and boys. Thus, the gender balance was not established. Science teachers were females and majored in science (1 in physics, 2 in chemistry and 1 in earth science). All three schools have 3 science classes a week. Schools 5 and 7 have two science laboratory classrooms and School 6 has three science laboratory classrooms; science teachers in the schools share the rooms. Mostly science lab activities were conducted in small groups, usually 4 to a group, but for some topics 2 to a group.

As TIMSS 2007 report showed, there was no weekly-based homework except in a few special cases (e.g., a science fair). There were 4 pre-set tests a year in traditional paper-and-pencil formats. The test results were counted toward a maximum 70% of total student achievement. The remaining 30% was assessed from lab reports, science fair results and other activities. These weights were uniform - across the three schools. Occasional quizzes were
administered but not included in the students’ final test scores.

**4.4.3. Summary of Research Context**

There were a few noticeable differences in the learning environments between Canadian and Korean schools in general. First, the average class sizes of Ontario Canadian schools were smaller than the average class sizes of Korean schools. Another difference was the ethnic and cultural cohesion of the student bodies; 43% of Canadian students were born in other countries rather than Canada while 100% of Korean students were born in Korea. Ethnicity was not a concern in either country, since Canadian teachers regarded diversity in their context as natural, with no great effect on their instruction, while Korean teachers did not have any diversity to address. Science tests were different. Four pre-set tests per school year were crucial to Korean students’ science scores. Both countries shared similar goals related to NOS in science education, but the Korean curriculum lacked a specific unit or defined guidelines to inculcate NOS concepts.

**4.5 Data Collection**

The following data were collected: 1) a survey questionnaire on students’ NOS, PLAT, PAFORMAT and PACONT, which took about 45-50 minutes to complete for the average grade 8 students, 2) semi-structured interviews, and 3) examples of achievement tests and rubrics. Also the open sources of each Ministry of Education (Ontario Ministry of Education: http://www.edu.gov.on.ca/eng/ and Busan Metropolitan City Office of Education: http://www.pen.go.kr/M_Eng/main/main.php) were used as a reference. The survey and other data sources provided prompts for questions during the semi-structured interviews.

**4.5.1. Catetogies of the Instrument**

The criteria for selecting and developing the questions include 1) matching up with this
research focused NOS concepts, 2) covering broad ranges of the learning activities and tasks and assessment, 3) keeping statements neutral, 4) achieving age appropriateness, and 5) avoiding test fatigue. In developing the items of learning activities and tasks and assessment, TIMSS’s (2007) questions and results for students and science teachers provided general ideas of both countries’ science classes. In selecting NOS questions, the previous NOS studies were referred, so the focusing concepts (*Tentativeness, Subjectiveness, Empirical Evidence Based Scientific Knowledge (EMP), Social Cultural Embeddedness, Diverse Scientific Research Method*) guide the selection of NOS questions. Details concerning the individual questions are found in Appendix A.

### 4.5.2 Emerged concerns

Two issues emerged in the course of developing the measuring instrument for this research. Firstly, since this study covers three different domains, the number of questions is a key concern. The measuring instruments reviewed in Chapter 2 and 3 were designed to minimize the measurement errors, so several questions were assigned to measure one construct. Those instruments have been validated worldwide. However, they have too many questions (e.g. WIHIC, 56 questions). If a study intends to examine only one of the domains, the number of questions cannot be a key concern, but if a study examines more than one domain as in this study, adopting the questionnaires demands some fundamental changes. Otherwise, the total number of questions would exceed one hundred. A relatively small number of questions can avoid test fatigue and to assist students to concentrate. It is also easy to meet the statistical requirements of the number of participants. However, with more than one hundred questions the liability of the results would be questionable.

The parsimony of the number of questions may cause measurement errors in an
abstract construct (Gorsuch, 1983; Schumacker & Lomax, 2010; Warner, 2008) and may not cover the complex and diverse activities in learning science. That is, if an instrument adopted a one-item assessment of abstract concept (e.g., NOS) would have serious limitations; it could not provide enough information about individuals’ understanding of different aspects of science. A reliable and valid questionnaire should include some ranges of an abstract construct in order to measure the concept as well as to engage enough participants.

Therefore, the issue became what minimum number of questions the survey should use to cover PLAT, PAFORMAT and PACONT and of their understanding of NOS. As a solution, two different measurement models were used. Formative measurement models (Bollen & Lennox, 1991; Chin, 1998; Diamantopoulos & Winklhofer, 2001) were created to gage perceptions on learning activities/tasks and on the assessment formats and content. Reflective measurement models were applied to the NOS concepts. In a formative measurement model, the construct of the model is defined by its observed variables rather than controlling the observed variables by the construct (Bollen & Lennox, 1991; Diamantopoulos & Winklhofer, 2001) without the restriction of high correlation among the observed variables. Thus, the model becomes more flexible in the selection of observed variables than a reflective model. Further discussion about the advantages of adopting formative measurement models will be followed in Section 4.5.3. In addition, to reduce the potential test fatigue, the six open-ended questions on NOS concepts are divided into two groups. Participants can choose either set, and so handle only three open-ended questions on NOS.

Second, as the research will be conducted in two countries, the questionnaires need to be identical in meaning across two different languages. If the questions do not carry the same meaning clearly, the results can lead to a wrong conclusion and mislead future studies (Liang
et al., 2008). The author translated the English questionnaire into Korean, and then, an English teacher in Korea checked whether the questions in Korean asked the same thing as the English questions. A teacher specializing both in Korean language and science refined the Korean questionnaire. These steps ensured that the meaning of the two language questionnaires was identical.

### 4.5.3 Measuring instrument

The questions were stated teacher-directed/student-directed learning activities and relativist/realist views of science. Converse and Presser (1999) and Warner (2008) state the importance of reverse worded questions in an instrument. To avoid the yea-saying bias Warner says

Self-report responses are prone to many types of bias, including **yea-saying** or **nay-saying bias** (some respondents tend to agree or disagree with all items) and social desirability bias (many people tend to report behaviors and attitudes that they believe are socially desirable) some scales include reverse-worded items.” (pp. 844-845, bold original).

The balance of ordered and reversed items was maintained. When analyzing the data, the reversed items’ scores would be rearranged for the conceptual consistency.

As mentioned before, due to the parsimony of the item numbers, formative measurement models (Bollen & Lennox, 1991; Diamantopoulos, 2006) for PLAT, PAFORMAT and PACONT and reflective measurement models for NOS concepts were employed. Without considering the meaning of a construct, if the measurement model were reflectively designed, misspecification would occur. The misspecification measurement model leads to over or underestimation of parameters, and the outcomes result in either Type I or Type II error (Diamantopoulos & Siguaw, 2006; Jarvis et al., 2003). The theoretical
framework for this research guided that the constructs of NOS concepts were pre-defined and less room to change the meanings. However, the perceptions of learning activities science and assessment were adjustable for the objectives of this research, and measuring the constructs with a formative model was an adequate method.

4.5.3.1 Students’ perceptions of dominant learning activities and tasks in science class (PLAT)

*PLAT* in science class focused two perspectives, teacher-directed (e.g., lecture) or student-directed (e.g., science project) learning activities and tasks. The construct covers what students do during science classes; i.e., how individuals perceive their learning activities, and what students think that their science teachers ask them to do during the classes. The questions of learning activities were “How often do you do these things in your science class?” and the questions of learning tasks required by their teachers were “How often does your teacher ask you to do the following?” Figure 4.3 includes the items to measure students’ perceptions on the learning activities and tasks.

As learning and teaching is a complex human enterprise, examining what happen in a class is not an easy task within a limited number of items. Based on 18 years of teaching experience in Korean middle schools, and on previous studies on learning activities in science classroom (Martin et al., 2008), five frequent classroom activities were selected, which could reflect the conceptual framework of this research. The items ask students how frequently, in their views, the activities and tasks occurred in their science classes. The activities are 1) listening to teachers’ explanation, copying what the science teacher wrote on the board, 2) teacher guided or demonstrated lab activities, 3) answering and solving questions, 4) memorizing scientific facts, theories and laws and 5) discussing or presenting scientific ideas...
and issues. Figure 4.3 includes the items to measure students’ perceptions of these learning activities and tasks.

| Your experience of science classroom activities: How often do you do these things in your science class? |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| 1       | I listen to the teach explaining the lesson or copy out what she/he writes on the blackboard. |
| 2       | I do experiments following teachers’ guidance. |
| What does your science teacher ask you to do? |
| 3       | My teacher asks me to memorize facts, theories and laws of science. |
| 4       | My teacher asks me to connect what I learn in science to everyday life. |
| 5       | My teacher asks me to plan how I will perform a project or present my ideas about a topic in science. |

Figure 4-3. Items for PLAT

4.5.3.2 Students’ perceptions of dominant assessment formats and content in science tests

(PAFORMAT & PACONT)

This research also considered how science test questions are formatted (open-ended, multiple-choice, constructive responses) and what kinds of cognitive demands they try to measure (factual knowledge, application, or explanation). Previous research on test formats (Biggs, 2000, 2003; Scouller, 1998) and examples of science tests from both Korean schools and Canadian middle schools guided the selection of items. The selected items were classified into knowledge-centered and understanding/application-centered assessments. For instance,
the ways in which questions in an assessment are formatted (true/false question types) or asking factual knowledge could reflect the fact that the science learning assessed has emphasized memorization. In contrast, a constructive format or questions applying a concept for a new situation could indicate that the science learning assessed has stressed the thorough understanding of content.

Questions pertaining to test formats (PAFORMAT), ask students which types of test questions they think most significantly affect their science scores. The questions are designed in a five-point Likert response format (from most significant to least significant). Questions regarding test content (PACONT) ask how often, they think, the content appears in their science test. Figure 4.4 and 4.5 represent these perceptions of science assessments.

<table>
<thead>
<tr>
<th>Which types of questions do you think most significantly affect your science test scores?</th>
<th>1 most significant</th>
<th>2 significant</th>
<th>3 moderate</th>
<th>4 little significant</th>
<th>5 least significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>True or false questions and multiple choice questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Write short answers by your own</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Presentation, discussion or an essay about scientific issues or projects.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4-4. Items for PAFORMAT*
How often do you think you get the following kinds of questions in your science tests

<table>
<thead>
<tr>
<th></th>
<th>1 most often</th>
<th>2 often</th>
<th>3 sometimes</th>
<th>4 seldom</th>
<th>5 least often</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Questions explaining science facts, theories, and laws</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Questions applying what I know to new situations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Questions about what I did during science lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Questions about what I think about how science affects and is affected by a society, and how it works.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-5. Items for PACONT

4.5.3.3. The Concepts of NOS

In constructing questionnaire items, reference was made to Nature of Scientific Knowledge Scale (NSKS) (Rubba & Anderson, 1978), Student Understanding of Science and Scientific Inquiry (SUSSI) (Liang et al., 2008), Views of Nature of Science-B (Abd-El-Khalick, Bell, & Lederman, 1998), VNOS-C (Lederman et al., 2002) for NOS. Figure 4.5 represents an example of items for tentative nature of science.

As reviewed in the section on the measuring instrument for understanding NOS (2.5.1), NSKS has been used worldwide for secondary school students. It includes tentative nature of science, subjectivity and empirical evidence based scientific knowledge of which concepts this study focused. However, NSKS repeatedly asks the same concept with positively and negatively stated questions, so those redundant questions were removed. SUSSI includes the concepts of cultural embedded scientific knowledge and diverse scientific research methods. Originally the constructs consist of 4 similar items, but this research chose three items for the
parsimony and modified a few words to suit the participants’ age level (e.g., double negative sentences were modified). The selection of open-ended questions was limited by the selected concepts of NOS. They were drawn from mostly VNOS-B (Abd-El-Khalick et al., 1998) and C (Lederman et al., 2002) and SUSSI (Liang et al., 2008).

The domains and questions were therefore selected and adapted from widely used, relatively proven instruments dealing with central issues within the grasp of the children studied, so that this research might build upon and readily relate its findings to previous studies.

<table>
<thead>
<tr>
<th><strong>Tentative nature of science</strong></th>
<th><strong>SA</strong></th>
<th><strong>A</strong></th>
<th><strong>N</strong></th>
<th><strong>D</strong></th>
<th><strong>SD</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A new interpretation of data can change our present scientific knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 We accept an idea as scientific knowledge only if it does not have any error.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 With the help of technology development, scientific knowledge approaches to absolute truth.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open-ended question After scientific theories developed (e.g. atomic theory/ cell theory), do the theories ever change? If you think that theories do change, explain why they change?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-6. Items for Tentativeness

4.5.3.4 Rationale for adopting formative measurement models for exogenous variables and reflective measurement models for endogenous variables

The central concepts involved in this research are PLAT, PAFORMAT, PACONT and NOS, and the meanings are broad and abstract. They cannot be measured directly using one item. Regarding the measurement errors, multiple items for a construct can theoretically cancel out the measurement errors. The idea derived from the measurement theory is the following: for instance, $X_1$ includes the true score and $error_1$, and $X_p$ includes the true score and $error_p$. 
When the scores are summed up, the errors will theoretically approach to zero, so they can be
cancel out: $X = p \times T + e$ (here, $e = 0$ and $X$ is a true measurement of an abstract construct).

In measuring several constructs with multiple items (observed variables),
aforementioned, the large number of questions was unavoidable. Formative (cause, causal)
indicators “are observed variables that are assumed to cause a latent variable. For effect
indicators the latent variable causes the observed variables” (Bollen, 1989, p.65). Concerning
the choice of a formative vs. a reflective measurement model, (Fornell & Bookstein, 1982)
provides detail explanation:

Construct such as "personality" or “attitude” are typically viewed as underlying
factors that give rise to something that is observed. Their indicators tend to be realized,
then as reflective. On the other hand, when constructs are conceived as explanatory
combinations of indicators (such as "population change "or "marketing mix") that are
determined by a combination of variables, their indicator should be formative. (p. 292,
emphasis original)

In other words, from the broad concepts of learning activities and tasks and test formats and
content, the meanings of those concepts in this research were defined by a combination of the
observed variables. Figure 4.7 illuminates the concepts of formative measurement model of
$PLAT$. The relationships between the observed variables and the construct $PLAT$ can be
expressed by the following equation:

$$\xi(PLAT) = \lambda_1 \times X_1 + \lambda_2 \times X_2 + \lambda_3 \times X_3 + \lambda_4 \times X_4 + \zeta(disturbance)$$

![Figure 4-7. Formative Measurement Model](image)
Reflective models were employed to measure students’ understanding of NOS. The concepts of NOS had been well defined and established by previous studies and the philosophers of science. The observed variables are influenced by the concepts of NOS (latent variables: constructs) not the opposite direction, so they are the effect variables. An observed variables which represented the underlying construct in reflective models, is expected to have correlations with the other observed variables. Figure 4.8 depicts the measurement model of Tentativeness. Due to the high correlations, dropping out one of the observed variables does not significantly change the meaning of the construct (Bollen & Lennox, 1991; Javis et al., 2003). The relationships between the observed variables and the construct can be expressed by the following equations:

\[
\begin{align*}
y_1 &= \gamma_1 \times \eta + e_1 \\
y_2 &= \gamma_2 \times \eta + e_2 \\
y_3 &= \gamma_3 \times \eta + e_3
\end{align*}
\]

Figure 4-8. Reflective Measurement Model (Tentativeness)

4.5.4 Semi-structured interviews

Nine volunteer students participated in the interviews. Four Canadian students did face-to-face interviews while emails and phone calls were facilitated for Korean students. The time for each interview depended on the interviewees’ personality, willingness to talk and
experiences. Generally the interview lasted about 25 minutes were recorded with a digital voice recorder, and were fully transcribed. The basic interview protocol was designed in advance to indicate experiences of science classes and assessments, and issues with NOS concepts. Questions emerging from the analysis of the participants’ survey responses were followed up. The questions addressed in the interview are detailed in Appendix B.

4.5.5 Other data resources

The science teachers’ clarification answers, samples of achievement tests and rubrics were gathered. The teacher clarification answers were not major data sources since the focus of this research was not teachers’ understanding of NOS or their perceptions on their science classes or assessments. If there were some aspects that were not clear in students’ responses regarding systematic levels, teachers were asked for clarification. Samples of science tests could be a backup for students’ perceptions on their assessment; what the tests actually measured (e.g., students’ memorized factual knowledge, process emphasized knowledge, knowledge transfer, conceptual understanding) and students’ PAFORMAT and PACONT as actually reflected in their science class practice.

Both Ontario Canada and Busan Korea Ministry of Education websites and TIMSS 2003 and 2007 data provide comprehensive general information over curriculum content, students’ attitudes to science, achievement levels among international countries, teachers’ perceptions on subject areas, the physical and administrative learning environment. Each school and classroom has its idiosyncrasies, but there must be commonalities in same countries. When the idiosyncrasies cannot provide enough explanations, the TIMSS data and both Ontario and Busan MoE open sources were referred to for the general information.
4.6 Reliability and Validity

Before analyzing the data, the quality of the measuring instrument was checked with three steps: 1) both reflective and formative measurement models, 2) structural models and 3) comparisons between the quantitative results and the qualitative results.

4.6.1 Assessing the NOS measurement models: Reflective models

The reflective models should be checked both for their internal consistency reliability and construct validity (convergent and discriminant validity). The first work was checking the measurement models’ internal consistency reliability. The traditional criterion for internal consistency is Cronbach alpha\(^1\) (Cronbach, 1951, cited in Warner, 2008), which provides an estimate for the reliability based on the indicator intercorrelations. Cronbach alpha for this research, however, was not a proper index of the internal consistency reliability. First it is strongly dependent on the number of observed variables (refer to the footnote equation), particularly when the number of observed variables is small, Cronbach alpha underestimates the model’s reliability considerably. Each construct for NOS concepts had three observed variables, and it could severely underestimate the model’s reliability. Another difficulty in adopting Cronbach alpha is that Cronbach’s alpha assumes that all indicators are equally reliable, whereas PLS takes into account that the observed variables have different loadings. Therefore, a composite reliability offers a better estimate of variance and can be interpreted in the same way as Cronbach’s alpha (Hair et al., 2006). The composite reliability (\(\rho_c\)) (Werts, Linn, & Joreskog, 1974) was a proper index for examining internal consistency reliability for

\[
\alpha = \frac{pr}{[1 + r(p-1)]} \quad (p: \text{# of observed variables})
\]

\(^1\)
the NOS concepts. The composite reliability can be calculated by the following equation:

\[ \rho_c = \frac{(\sum \lambda_i)^2 \text{var}F}{(\sum \lambda_i)^2 \text{var}F + \sum \Theta_{ii}} \]

where \( \lambda_i \) is the standardized component loading of a manifest indicator on a latent construct, \( F \) is factor variance, and \( \Theta_{ii} \) is unique/error variance. Mostly \( F \) is set at 1, \( \Theta_i \) is 1 minus the square of \( \lambda_i \) (Chin, Marcolin & Newsted, 1996). Then, the formula becomes following:

\[ ICR = \frac{(\sum \lambda_i)^2}{[(\sum \lambda_i)^2 + \sum(1 - \lambda_i^2)]} \]

The internal consistencies of 0.70 or higher are considered adequate (Agarwal & Karahanna, 2000; Barclayetal.1995) whereas a value below 0.6 indicates a lack of reliability.

Another way is the magnitudes of outer loadings; i.e. the absolute correlations between a construct and each of its observed variables should be higher than 0.7 in reflective models. But some scholars in psychology (e.g., Churchill, 1979) recommend eliminating reflective indicators from measurement models if their outer standardized loadings are smaller than 0.4. Henseler et al., (2009) and Diamantopoulos and Winklhofer (2001) recommend researchers to take into account PLS’ characteristic of consistency at large; therefore, one should be careful when eliminating indicators. Only if an indicator’s reliability is low and eliminating this indicator goes along with a substantial increase of composite reliability, it makes sense to discard this indicator. For this research an observed variable of which outer loading is smaller than .4 would be removed rather than following the conventional standard (.7) because each reflective measurement model has only three observed variables, removing one of the variables could not measure the construct properly.
To check the validity of the reflective measurement model, convergent and discriminant validity were examined. For assessing the convergent validity of a construct, I employed Fornell and Larcker’s (1981) Average Variance Extracted (AVE) criterion. An AVE value greater than 0.50 indicates that a latent variable is able to explain more than half of the variance of its indicators on average (Henseler et al., 2009). The PLS output provides the values of AVE or they can be calculated using the following equation suggested by Chin (1998):

\[
AVE = \frac{\sum \lambda_i^2}{\left[ \sum \lambda_i^2 + \sum (1 - \lambda_i^2) \right]}
\]

Discriminant validity of the measurement models was tested through Fornell and Larcker’s (1981) AVE test and cross loadings criterion (Chin, 1998). Evidence of discriminant validity occurs when the square root of the average variance extracted estimation exceed the correlations between the factors making each pair (Fornell & Larcker, 1981). Each latent variable shares more variance with its own block of indicators than with another latent variable representing a different block of indicators. According to cross loading criterion (Chin, 1998), the loadings of an indicator to its targeting construct are expected to be greater than all of its cross-loadings.

**4.6.2 Assessing PLAT, PAFORMAT and PACONTENT:**

**The formative measurement model**

The internal consistency reliability of formative measurement models in traditional test theory is not as crucial as that of the reflective measurement model. Diamantopoulos (2006, p.11) says, “Reliability becomes an irrelevant criterion for assessing measurement quality.” Diamantopoulos and Winklhofer (2001) suggest that construct reliability of formative models
should be performed by multicollinearity and a test of indicator validity (path coefficients’ significance). Formative measurement models are based on a multi-regression so that multicollinearity should not exist (Diamantopoulos & Winklhofer, 2001). That is, the equation for a formative construct is the same as a multiple regression equation (refer to Figure 4.7):

\[ \hat{Y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + a \]

Thus, the reliability evaluations for the formative models are to assess the assumption of no multicollinearity (Diamantopoulos & Siguaw, 2006); Variance Inflation Factor (VIF) is evaluated. If the values of VIF are less than 3.3, the indicators of a construct are excellent. Also VIF is less than 10 that no collinearity is commonly accepted; however, greater than 10 has multicollinearity problems (Warner, 2008).

Because the limitation of applicable quantitative quality checks of the formative measurement models, a validity assessment is a controversial issue (Diamontopoulos & Winklhofer, 2001; Henseler et al., 2009; Javis et al., 2003). Validity assessments can be done at two levels: indicator and construct validity. If the estimated indicator’s weight (\( \lambda_i \) in Figure 4.7) is not significant, an indicator does not contribute to the targeting construct. A nonparametric bootstrapping method provides \( t \)-values of the formative indicators’ weights. The significance of the \( t \)-values is the basis for judging whether a weight is significant or not. However, Bollen and Lennox (1991) and Diamontopoulos and Winklhofer (2001) warn against the decision, noting that if the weight for an indicator is insignificant and the indicator might be eliminated, the meaning of the construct would be changed. Each formative indicator carries its meaning, and an elimination of the insignificant indicator risks to change the theoretical perspective of the constructs (Bollen & Lennox, 1991; Diamantopoulos, 2006; Jarvis et al., 2003). Thus, when an observed variable has a low outer loading, the decision to
eliminate it should be made carefully.

4.6.3 Structural model

For structural model validity, inner path coefficients were checked the signs, inner coefficients’ magnitudes, t-values and the significance. The nonparametric bootstrapping procedure (Chin, 1998; Davison & Hinkley, 2003) using 1000 subsamples was performed to evaluate the statistical significance of each path coefficient. In addition, how much the variances ($R^2$) of the latent dependent variables (endogenous variables of NOS concepts) were accounted for by the latent independent variables (exogenous variables of $PLAT$, $PAFORMAT$ and $PACONT$). Since Chin’s (1998) standards were widely used for $R^2$, this study also used them; 0.67, 0.33, or 0.19 are described as substantial, moderate or weak (p. 323). The effect size, $f^2$, for each independent latent variable was examined; i.e., $f^2$ is a gauge for whether the latent independent variables have weak (0.02), medium (0.15) and large effect (0.35) on the endogenous variables: $f^2 = (R^2_{\text{included}} - R^2_{\text{excluded}}) / (1 - R^2_{\text{included}})$.

For the prediction relevance, Stone-Geisser’s criterion ($Q^2$) was used since it was dominantly used in PLS (Henseler et al., 2009). The predictive relevance ($Q^2$) is one of the major aspects of PLS analysis. Geisser (1975) emphasizing the importance of prediction role of PLS says, “the prediction of observables or potential observables is of much greater relevance than the estimation of what are often artificial construct-parameters” (p.320). $Q^2$ is “a measure of how well-observed values are reconstructed by the model and its parameter estimates” (p.318). If $Q^2$ is positive, the model has the predictive relevance. On the contrary, if $Q^2$ is negative, the model lacks the predictive relevance. The values ($Q^2$) were measured by the

$$Q^2 = 1 - \frac{\sum_D E_D}{\sum_D O_D}$$

where $D$ is omission distance, $E$ is the sum of squares of prediction error and $O$ is the sum of squares errors.
blindfolding procedures (Tenenhaus et al., 2005). The prediction relevance ($Q^2$) indicates that “the observed values are well reconstructed and the model has predictive relevance” (Henseler et al., 2009, p. 303). The effect size of prediction relevance, $q^2$, indicates that a latent variable has small (0.02), medium (0.15) and large (0.35) predictive relevance (Henseler et al., 2009).

### 4.6.4 Comparison between quantitative and qualitative Data

The comparisons between the quantitative results and the qualitative results would also provide the validity and reliability of the data. Since the instrument had the open-ended questions on the concepts of NOS, the comparisons were done while the responses to the open-ended questions were analyzed; the consistency and diversities of the answers were examined.

### 4.7 Data Analysis

Four distinct steps of analysis were taken in this study. As a first step, the mean differences in the quantitative data between countries and among schools were examined. The second step examined the relationships between NOS concepts and $PLAT$, $PAFORMAT$ and $PACONT$. The third step involved qualitative analysis of open-ended questions; data categorizations of common themes, the reasons for various answers to closed-ended questions and the consistency between close-and open ended answers and unique answers. The last step addressed the combination and triangulation of quantitative and qualitative data, and other data sources.

#### 4.7.1 Quantitative data

##### 4.7.1.1 Missing Data and Outliers

A case was excluded from the data set if it had more than 6 missing items (20%) although the missing items were completely random. However, when a case had less than 20%
missing items, it was included and each missing item was replaced with the item mean. Several methods are available in treating missing cases such as listwise or pairwise elimination, and replacing the missing value into the item mean. Since the analytical method of structural equation modeling requires a large number of participants, the method of the mean replacement seemed preferable.

In dealing with statistical outliers, Warner (2008) says that cases greater or less than 3 SD are regarded as extreme. However, when Warner’s standards were compared to items with a 5-point-Likert scale, 3 SD could not directly be applied at each item level. Therefore, the outliers were checked at the construct level. If a student had more than two constructs greater or less than 3SD (+/-2.58 standardized score), that student would be classified as an outlier. In other words, a construct consisted of at least three items; more than two constructs greater/less than 3 SD meant, possibly, more than 6 items were extreme scores. The construct scores (latent variables) were calculated using the software, SmartPLS.

4.7.1.2 Differences

Differences between the two countries and among schools were examined on the concepts of NOS and PLAT, PAFORMAT and PACONT. Descriptive statistics and Multivariate Analysis Of Variance (MANOVA) were performed using the Statistical Package for Social Sciences (SPSS) version 19. The two factors in MANOVA were country and school and the outcome (dependent) variables are the observed variables of PLAT, PAFORMAT, PACONT and the NOS concepts. Many outcome variables were involved, MANOVAs were more suitable to examine the differences in means of several outcome variables and different groups than a multiple comparison of univariate ANOVA.
According to Warner (2008), MANOVA has at least three advantages when more than one outcome variable in a study involves comparisons of means across groups as opposed to conducting separate one-way ANOVAs for each individual outcome measure. First, it does not inflate the risk of Type I error, which can arise when multiple significance tests are performed. The probability of making at least one Type I error in the set of multiple tests of one-way ANOVA is likely to be substantially higher than the alpha level (e.g., conventional alpha level is $\alpha=.05$). The second advantage of using MANOVA is that it concerns the linear intercorrelations among the Y outcome variables whereas multiple univariate ANOVA tests do not take this into account.

However, to obtain more informative results from MANOVA than univariate ANOVAs, no violations of the assumptions of MANOVA are required. If the assumptions were violated, the obtained significance value could be a poor indication of the true level of risk of Type I error. MANOVA assumptions are: 1) the independence of participants’ Y scores from each other, 2) the normality of the outcome, Y, variables, 3) the linear associations between pairs of Y variables and 4) the homogeneity of variance and covariance matrices for Y variables across the population groups. As the data were not repeatedly measured, participants’ Y outcome variables were independent from each other. The other violation assumptions were tested using histograms, skewness and kurtosis of each outcome variable for the normality, and Box’s M tests for homogeneity tests. Further, Box’s M tests are very sensitive to normality of the distribution of scores. Particularly when the sample size is large, the results may be statistically significant although the initial patterns are not severely violated. Thus, since the total number of participants was 536, the significance level for Box M tests
would be .01 rather than .05. In reducing the significance level of Box’s M, Warner (2008, p.668) says that Box’s M testing can be problematic when the number of cases is quite large:

“Box’s M may indicate statistically significant violations of the homogeneity of the variance/covariance assumption, even when the departures from the assumed pattern are not serious to raise problems….To compensate this, when the overall number of cases is very large, it may be preferable to evaluate the statistical significance of Box M using a smaller alpha level” (p.668).

If the Box’s M tests were significant, Pillai’s trace would be used for MANOVA instead of Wilks’s lambda because Pillai’s trace is robust to the violation of homogeneous assumptions.

The null hypotheses for students’ perceptions on their Learning Activities, the perceptions on assessment and understanding of NOS are:
In $\mu_{kp}$, $p =$ the outcome variables (5 for learning activities, 7 for assessment observed variables; while NOS concepts were compared on the construct level, it had 3 from 5 construct.) and $k =$ the number of groups (7 schools).

If MANOVA analysis results were significant, it means there is at least a difference among schools or between countries, or a difference from one of the outcome variables. A one-way between-subject ANalysis Of VAriance (ANOVA) was adopted as the follow-up analysis. The results of ANOVA provided two significant tests: a significance test for the main

$$PLAT : H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \\ \mu_{15} \end{bmatrix} \begin{bmatrix} \mu_{71} \\ \mu_{72} \\ \mu_{73} \\ \mu_{74} \\ \mu_{75} \end{bmatrix} \text{ and } H_1 \neq H_0$$

$$PAFORMAT : H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \begin{bmatrix} \mu_{71} \\ \mu_{72} \\ \mu_{73} \end{bmatrix} \text{ and } H_1 \neq H_0$$

$$PACONT : H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \begin{bmatrix} \mu_{71} \\ \mu_{72} \\ \mu_{73} \end{bmatrix} \text{ and } H_1 \neq H_0$$

$$NOS : H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \begin{bmatrix} \mu_{71} \\ \mu_{72} \\ \mu_{73} \end{bmatrix} \text{ and } H_1 \neq H_0$$
effect of Factor A (country difference), and a significance test for the main effect of Factor B (school difference). To assess the pattern of outcomes in a nested ANOVA, two separate null hypotheses were tested:

First null hypothesis: Test of main effect for Factor A encompasses that the population means on the outcome variables (learning activities/assessment/NOS concepts) are equal across all levels of factor A. Using an SS (sum of squares) term obtained from the sample means for Y for the A1 and A2 groups, I can have information about the effect of A on Y.

\[ H_0 : \mu_{A1} = \mu_{A2} \]

Second null hypothesis: the population means on Y are equal across all levels of B

\[ H_0 : \mu_{B1} = \mu_{B2} = \mu_{B3} = \mu_{B4} = \mu_{B5} = \mu_{B6} = \mu_{B7} \]

Assumptions for an ANOVA are that Y variable is approximately normally distributed and the variances of scores are reasonably homogeneous across groups (the Levene test). ANOVA is fairly robust to violations of the normality assumption and the homogeneity of variance assumption unless the numbers of cases in the cells are very small. In this study, the minimum sample size for School 4 is 45 and it is not seriously small (Warner, 2008).

4.7.1.3 Associations

As one of the major objectives of this research is to identify factors, which affect students’ understanding of NOS, Structural Equation Modeling (SEM) has advantages in examining various types of hypothesized (theorized) models to depict the relationships among observed variables (Bollen & Lennox, 1991; Chin, 1998; Henseler et al., 2009). If the hypothesized models are supported by the gathered sample data, then further complex theoretical models can be postulated to explain more accurately the real situations. If the
sample data do not support the theorized model, then either the original model can be modified and tested or other theoretical models need to be developed and tested.

**Rationale for Employing PLS:** Learning and teaching is one of the most complex of human enterprises. To explain and predict this process correctly, it would be better to seek to understand multiple observed variables rather than a few independent variables and one dependent variable, which cannot allow for the development of sophisticated theories.

Haenlein Kaplan (2004) pointed out the limitation of regression analysis:

(a) the postulation of simple model structure (at least in the case of regression-based approaches), (b) the assumption that all variables can be considered as observable; and (c) the conjecture that all variables are measured without error, which may limit their applicability in some research situations (p.284).

Jacoby (1978), in discussing the simple model structure of regression, was concerned that an analysis based on one or two variables in isolation would be relatively artificial and inconsequential to describe a complex reality. In addition, this limitation makes it difficult to consider the effects of mediating or moderating variables. Furthermore, it is far from realistic to conjecture a measure of variables without error. Warner (2008) advises that researchers have to bear in mind that each observation of the real world is accompanied by some measurement errors, which could be either random, systematic, or both.

In order to overcome the limitations of regression analysis, Partial Least Squares (PLS), one of the methods of Structure Equation Modeling (SEM), was used. This was determined because several independent and dependent variables and unobserved variables (latent variables) were involved in this research. The methods postulated take measurement errors for the observed variables into account.
PLS was introduced by Wold (1978, cited in Bollen & Lennox, 1991). Chin and Newsted (1999) explain that this method focuses on maximizing the variance of the dependent variables explained by the independent variables, so it can maximize the model’s predictive ability. This is the major different from covariance based SEM (e.g. LISREL analyses). Both covariance based SEM (CBSEM) and PLS models consist of structural components, which reflects the relationships between the latent variables, and a measurement component, which shows how the latent variables and their indicators are related.

Unlike CBSEM, PLS does not make strict assumptions about distribution. If an analytical method were not robust in the violation of normality, the results would be biased and unreliable. This research was conducted over two countries and seven schools. With such data, the distributions may not be easy to meet the normality assumption. While CBSEM requires large sample sizes, PLS requires a relatively smaller number of participants (Loehlin, 2004). As mentioned in the participant description section, the Canadian sample size did not meet the standard of 10 times the observed variables. This small sample size was one of the hindrances to employing CBSEM for this research. However, PLS requires a smaller number of participants; ten times the number of items was comprised by formative constructs (Barclay, Higgins, & Thompson, 1995). The number of items that comprised the formative constructs for this study was 12, so the number of participants met the recommendation. The third advantage of PLS over CBSEM for this research was that it applies to situations where knowledge about distribution of the latent variables is limited. PLS requires estimates to be more closely tied to the data rather than to the hypothesized model as in CBSEM (Fornell & Cha, 1994). Very rarely have studies been conducted to gauge the relationships between students’ perceptions on science assessment and their understanding of NOS, so not much
knowledge has been discovered about the distribution of latent variables. The last, but most importantly, was the analytical software. SmartPLS (Ringle, Wende, & Will, 2005) allows for estimating both measurement and structural model simultaneously with no limitation of the number of observed variables.

**Assessing PLS models:** There is no global goodness-of-fit in PLS analysis (Henseler et al., 2009); thus, the model adopted was evaluated by examining the validity and reliability of the model and its measurements. According to Henseler et al. (2009), “It only makes sense to evaluate the inner path model [structural model] estimates when the calculated latent variable scores show evidence of sufficient reliability and validity (p.298).” Thus, the reliability and validity of outer measurement models were examined first and then the structural model was assessed. Details of this model evaluation were set out in the section on reliability and validity.

### 4.7.2 Qualitative data

General strategies for analyzing qualitative data were 1) transcribing the semi-structured interviews, 2) reading through the texts of the interview transcripts and open-ended responses, 3) coding the raw data in accordance with predefined categories and separating the emergent categories, 4) reconstructing the coded data to make conceptually and theoretically coherent data and 5) comparing and supplementing reconstructed data with the results of the quantitative data analyses and other data sources such as the examples of science tests.

#### 4.7.2.1 Interview Data

The main reason for conducting semi-structured interviews was to enhance the credibility of research findings as to of the associations from the quantitative data analysis. In order to avoid participants’ fatigue, there were no open-ended questions on students’ learning
activities and tasks and science assessments on the survey. Therefore, the interview questions probed these aspects.

**Transcription and translation**: First, a denaturalized approach was taken toward the transcription process (Oliver et al., 2005). Oliver et al. (2005) defines denaturalism as practices where “idiosyncratic elements of speech (e.g., stutters, pauses, nonverbal, involuntary vocalizations) are removed.” (p. 1273-1274). The denaturalized perspective holds that “within speech are meanings and perceptions that construct our reality” (p. 1274). Since the interest of this research is the informational content of what students said rather than the conversation or language itself, the strategy of denaturalism is appropriate to this research. Korean students’ responses were translated into English by the author, and proofread by a university student who speaks and writes both languages fluently.

**Coding**: Coding is an analytic and synthetic processes. The raw data were fractured, analyzed and grouped by predefined categories so as to reconstruct their meaning within a category. These processes involve “taking constructions gathered from the context and reconstructing them into meaningful wholes” (Lincoln & Guba, 1985, p. 333). The data were sorted into 1) major categories, 2) sub-categories and 3) emerging categories. The major categories were the four main domains of this research: Perceptions of Learning Activities and Tasks (PLAT), Perceptions of Assessment Formats (PAFORMAT), Perceptions of Content (PACONT), and the concepts of NOS. The subcategories were divided into: teacher-directed learning or student-directed learning, learning of science or learning about science; test formats emphasizing one correct answer or test formats emphasizing opening alternatives/diverse possible answers; and tests stressing memorization of content or tests stressing application of content. The subcategories of NOS were divided between naïve realistic or informed views
(epistemological naturalistic or rationalistic views, and ontological relativistic or realistic views) (refer to Figure 4-9 Conceptually clustered matrix). The emerging categories could be unique, atypical responses that were not predefined.

**Reconstruction:** Reconstruction involves collapsing the coded data into common themes, finding connections and interrelated themes among them, and seeking answers to the research questions. These processes make meaningful wholes of the data.

**Comparison and Supplementation:** The interview data were compared both within each interviewee’s input (interview data and survey data) and between interviewees. A further comparison was considered between the results of the interview data with its large number of open-ended responses, and the close-ended data as a whole. While this might have been meaningful, it had to be borne in mind that the interviewees could not represent the whole sample of participants. The number of interviewees was a few, and not randomly selected from the sample populations, so that their representativeness was very low. Therefore, direct comparisons of interview results with quantitative analysis results (whether from close-ended or open-ended responses) could not provide unbiased and reliable meanings. These interview results then would have the role of supplementing the quantitative analysis results.

**4.7.2.2 Open-ended responses (Understanding of NOS)**

Although the conceptually clustered matrix was in hand before the data were collected, it could not predict what responses would come or how they should be classified. Once actual data were collected, it was important to read them without preconceptions or coding schemes, so as to understand the students’ actual ideas. This meant the conceptually clustered matrix is not an absolute standard but there are some rooms to adjust to the students’ level of comprehension and expression. A second reading of the actual data categorized material into
the major concepts of NOS, not only where questions addressed them as major categories, but also where questions overlapped several concepts, as Lederman et al. (2002) suggested.

The next step was to classify the material in detail based on the matrix (Figure 4-9). Common themes and atypical answers under the subcategories were found and highlighted. In this study, a “common theme” means that more than 10% of students answered similarly, while atypical or unique answers were the responses, which did not belong to any category in the conceptually clustered matrix. Overall, the examples students gave to support or explain their opinions and ideas were quite profound relative to their grade level. Erickson (1998) stresses, “[G]ood qualitative research reports the range and frequency of actions and meaning of perspectives that are observed, as well as their occurrence narrative” (p.1155). Thus, the frequencies of responses in each subcategory were counted and presented in a table.

Naturally, the open-ended answers were not as straightforward as the close-ended answers. Some students answered with one or two words, some expressed their opinions without providing reasons and explanations, and a few agreed to both views. Due to these diversities, the classification of the answers as either “informed” or “naïve realist” could be more or less arbitrary. It was not proper to classify one or two words responses without an adequate explanation of opinions and atypical responses with others; therefore, no explanation responses put under the “no explanation” category, and atypical responses were put under “others”. Some answers were vague, others showed two perspectives and so were double counted, and yet a further, considerable proportion of answers went blank. For these reasons, the classification table could not cover all the information provided in the open-ended responses.
Direct quotations of representative or unique statements, even from both countries, were extracted to supplement the limited information in the classification table. Comments on the quotations proceeded. Martin-Dunlop (2004) says that prefatory comments can tend to direct readers’ attention in examining the text. However, these direct quotations also enable readers to validate or invalidate the interpretations given.

<table>
<thead>
<tr>
<th>NOS Aspect</th>
<th>More realistic &amp; rationalistic views (naïve realistic views)</th>
<th>More naturalistic and relativistic views (informed views)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical NOS &amp; theory laden observation</td>
<td>Science is something that is straightforward and isn’t a field of study that allows a lot of opinions, personal bias, or individual views—it is fact based. Science is concerned with facts. We observed facts to prove that theories are true Without direct seeing something, scientists can reach different conclusions. Scientists are very objective observers while artists are very subjective.</td>
<td>Much of the development of scientific knowledge depends on observation… [But] I think what we observe is a function of convention, I don’t believe that the goal of science is (or should be) the accumulation of observable facts. Rather … science involves abstraction, one step of abstraction after another. Scientists are humans. They learn and think differently, just like all people do. They interpret the same data sets differently because of the way they learn and think, and because of their prior knowledge.</td>
</tr>
<tr>
<td>Scientific Method</td>
<td>Science deals with using an exact method…. That way we know we have the right answer. Science has a particular method of going about things, the scientific method. Scientists are very objective because they have a set of procedures they use to solve their problems. Artists are more subjective, putting themselves into their work. Science would not exist without scientific procedure that is solely based on experiments…. The development of knowledge can only be attained through precise experiments.</td>
<td>Scientists investigate research questions with prior knowledge, perseverance, and creativity. Scientific knowledge is constructed and developed in a variety of ways including observation, analysis, speculation, library investigation and experimentation. Scientists are human. They learn and think differently, just like all people do. They approach a problem with different ways, and interpret data sets differently because of the way they learn and think and because of their prior knowledge. Different scientists may come up with different explanations based on their own education and background. Experiments are not always crucial. It is limited by scientists’ background knowledge.</td>
</tr>
<tr>
<td>Tentative NOS</td>
<td>The same results over and over again, then the results become a law. Facts are truths, not change. Comparing to religion… science demands definitely….. right and wrong.</td>
<td>Everything in science is subject to change with new evidence and interpretation of that evidence. We are never 100% sure about anything because… negative evidence will call a theory or law into question, and possibly cause a modification. Having confidence in scientific knowledge is reasonable while realizing that such knowledge may be abandoned or modified in light of new evidence or reconceptualization of prior evidence and knowledge. The history of science reveals both evolutionary and revolutionary changes.</td>
</tr>
<tr>
<td>Social and cultural embeddedness</td>
<td>Science is about the facts and could not be influenced by cultures and society. E.g.,</td>
<td>Culture influences the ideas in science.</td>
</tr>
<tr>
<td>ss of science</td>
<td>atoms are atoms in the USA, and in Asia. Well, society can sometimes not fund some scientific research. So in that sense it influences science. But scientific knowledge is universal and does not change from one place to another.</td>
<td>All factors in society and the culture influence the acceptance of scientific ideas. People from all culture contribute to science. As a human endeavor, science is influenced by the society and culture in which it is practiced. The values and expectations of the culture determine what and how science is conducted, interpreted, and accepted.</td>
</tr>
</tbody>
</table>

*Figure 4-9. Conceptually Clustred Matrix*

### 4.7.2.3 Other data sources

The qualitative data for *PLAT, PAFORMAT* and *PACONT* were drawn from the semi-structured interviews of students and samples of science tests. Unclear points were clarified by asking the teachers. In addition, just as TIMSS data outlined the research context, so also TIMSS reports supplemented the data sources. Furthermore, the documents from the MoE of Ontario and of Busan and two school district boards (Toronto District of School Board and Busan District of School Board) provided significant supplementary references, especially where discrepancies occur among different data sources.

### 4.7.3 Integration of Quantitative and Qualitative Data

The purpose of integrating qualitative and quantitative data is to enable each method to complement the other’s weaknesses. In the first stage of data integration, the results of quantitative data analysis were used to formulate probing questions for the student interviews. In the second stage, the results of close-ended questions and the open-ended questions of NOS

---

3 from AAAS, 1993; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Liang et al., (2008)
were combined. Each construct of NOS concepts was accompanied with one open-ended question. The open-ended questions and the items of the construct were integrated using the clustered content matrix and the definition of NOS. At the third stage of integration, the results of the survey were compared with the interpretation of interviews as well as with other educational documents and outside resources (TIMSS results, and school board website resources). Triangulation of different data sources and integration enhanced the accuracy of this research.

4.8. Ethical Consideration

Before commencing this research written consents were obtained for the Ethical Review Process between the Ontario Institute for Studies in Education of University of Toronto (OISE/UofT) and the district school boards to which participating schools belonged. Also, all of the work involved in this research was conducted after gaining written consents. Special care to do this was required since the research took place in school environments where the participants were under age. The researcher gave participants sufficient information in such necessary areas as:

- What the research is,
- What kinds of constructs are measured,
- What participants’ rights are, and
- How confidentiality will be kept.

Before administrating the questionnaire, the researcher of this study had confidence that reciprocity would be included in the guidelines for the questionnaire. The outcomes of this study have to be helpful for the participants to learn science and for the teachers to practice their instruction.
Attention must be paid to protecting individual students from potential harms, weighed against the benefits of conducting the research with honesty and justice. In reporting research outcomes, pseudonyms were used. In conducting this research, the maxim can be best described in Kant’s words (cited in Flinders, 1992), “We [researchers] should treat others [participants] not solely as a means, but as ends in themselves” (p. 104). Shank (2002) also stressed the protection of participants. He said that the followings promises should be required in any ethical contract: 1) not to harm; 2) to be open; 3) to be honest, and 4) to be careful. Emphasizing the importance of accuracy, Mills (2003) warns not to leave any space for deception. In this research, the yardsticks for qualitative data analysis will not be so “flexible” as to adjust the predictability of the factors.

### 4.9 Summary

The focus of this chapter was to explain the methods this research used. First it outlined the research questions and design (a mix methods approach). The participants and general context of Canadian and Korean junior middle school classrooms and the schools were described. In the data collection section, the procedures of instrument development, the major constructs and their observed variables were explained in some detail. Then, the plan for examining validity and reliability was included because the instrument had not been used in other research. Then, the data analytical strategies were followed in terms of the differences (MANOVA) and associations (PLS) for the quantitative data and in terms of a conceptually clustered matrix and coding strategies for the qualitative data. The integration of quantitative and qualitative methods was described with the intention of adopting both methods to compensate and verify each other. The last section discussed relevant ethical considerations.
Overall, this chapter has mapped out the methodology, which provides the guidelines for conducting the research.
5. RESULTS FOR THE SURVEY QUESTIONNAIRE

5.1 Introduction and Overview

A survey was used to obtain information about students’ Perceptions on the Learning Activities and Tasks (PLAT) in science classes, their Perceptions on Assessment of science Test Formats (PATFOMAT) and Content (PATCONT), and their understanding of the Nature Of Science (NOS). As mentioned in the chapter on methodology, the survey consisted of qualitative (open-ended questions) and quantitative (Likert scaled items) elements. Having both qualitative and quantitative elements allowed them to compensate for each other’s methodological weaknesses.

The process of quantitative data analysis started with a general description of the samples. Descriptive statistics are basic for all statistical analyses since they provide an outline of sample data such as distributions and centrality (Spatz, 2002; Warner, 2008). Thus, the first analyses examined the frequencies, means and standard deviations (SD) for each item and the outliers for the each construct. The second step was an examination of the reliability and validity of the questionnaire before inferential statistics were applied. Multivariate Analysis Of Variance (MANOVA) examined the effects of the country and school differences on multiple dependent variables of PLAT, PATFORMAT, PATCONT and NOS concepts. As a last step of quantitative analysis, Partial Least Squares (PLS) were employed to identify the relationships among the variables. As the samples for this study were from Canadian and Korean students, the results were reported by 1) individual countries, 2) the combination of the two countries, and 3) the comparisons and contrasts between two countries.
5.2 Variable Codes

Table 5.1 shows detailed information about all the variables used in this study and their codes. The country and school names are nominal although the codes were assigned with numbers. Canada was labeled as 1 and Korea was labeled as 2. Canadian schools were School 1, 2, 3 and 4 while Korean schools were labeled School 5, 6 and 7. The perceptions on learning activities and tasks (PLAT) and assessment (PAFORMAT, PACONT) and understanding of NOS consisted of five point Likert scales and treated as scale data. Three categorical variables were used for the identification of individual participants, and in total 27 variables were deployed to examine the content of PLAT, PAFORMAT, PACONT and NOS.

Table 5-1

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable name</th>
<th>Type</th>
<th>Value labels</th>
<th>Descriptions of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification</td>
<td>Country</td>
<td>Nominal</td>
<td>1=Canada 2=Korea</td>
<td>Participants’ countries</td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>Nominal</td>
<td>1<del>4=Schools from Canada 5</del>7=Schools from Korea</td>
<td>Participants’ schools</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>Nominal</td>
<td></td>
<td>Individual participants</td>
</tr>
<tr>
<td>PLAT (Perceptions of Learning Activities and Tasks during science class)</td>
<td>1.PLATlecture</td>
<td>Scale</td>
<td>Possible range: 1~5</td>
<td>PLAT 1~5 are used to measure students’ perceptions on their science class activities and tasks</td>
</tr>
<tr>
<td></td>
<td>2.PLATquestion</td>
<td></td>
<td>1: almost always</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.PPLATlap</td>
<td></td>
<td>5: hardly ever</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.PLATmemo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.PLATpresent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>Possible range: 1-5</td>
<td>1: most significant</td>
<td>5: least</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>----------</td>
<td></td>
</tr>
</tbody>
</table>

**PAFORMAT**
(Perceptions of Assessments Formats)

1. **PAFORMATclose**
2. **PAFORMATshort**
3. **PAFORMATalter**

**PACONT**
(Perceptions on Assessment Content)

1. **PACONTknowledge**
2. **PACONTapplication**
3. **PACONTlab**
4. **PACONTtphs**

**NOS**
(Nature of Science)

1. **TEN1, 2, 3**
2. **SUB1, 2, 3**
3. **EMP1, 2, 3**
4. **CUL1, 2, 3**
5. **METH1, 2, 3**

**5.3 Data Rearrangement and Exclusion**

As the items were stated positively/negatively or from relativist/realist points of view, the rearrangement of some items was required for conceptual consistency; a higher score
indicates a greater degree of relativistic viewpoint for NOS. For instance, item A was stated from the relativist point while item B was stated from the realist point:

**Item A:** *A new interpretation of data can change our present scientific knowledge.*

**Item B:** *Scientific knowledge approaches the absolute truth.*

Responses reflecting item A should be arranged inversely so that relativist viewpoints have high scores. *PLAT, PAFORMAT* and *PACONT* were also rearranged; thus, high scores were interpreted to indicate that students perceived their science class activities or tasks were more student-directed, while low scores were interpreted as teacher-directed. For instance, *PLAT*presentation (My teacher asks me to plan how I will perform a project or present my ideas about topics in science) was classified as more of student-directed learning activities and tasks, so the item was inversely arranged. In case a student responded to this item with number 1 (“almost always”), the reverse of number 1 would be 5. Consequently, the conceptual consistency with other items was achieved in interpretations of the data.

With respect to data exclusion, 21 cases from the Canadian sample and 72 cases from the Korean sample were excluded because more than 20% (6 items) of the answers were missing. The cases that had less than 6 missing items were replaced with the item means. In addition to the missing data exclusion, there were a total of 20 cases whose constructs were more or less than 3SD (standard deviation). Nineteen cases had only one extreme construct, but one case of School5 had two extreme constructs; *Sociocultural Embeddedness* was -3.80 and *Diverse Methods in Scientific Research* in was -4.54. This meant that more than 6 items were relatively located in extreme scores. And the case was excluded. Therefore, the effective sample sizes were 217 (Canada) and 319 (Korean).
5.4 Descriptive Statistics

5.4.1 Perceptions of Learning Activities and Tasks (PLAT)

There were 5 items under the theoretical construct of PLAT. Two items, PLATquestion and PLATpresentation, were inversely arranged. School 2 had one missing case in Item 2 (PLATquestion) and School 1 had one in Item 3 (PLATlab), and 2 in PLATpresent. PLATpresentation distributed as positively skewed but not as seriously skewed as PLATlecture. Other items were relatively evenly distributed.

In PLATlecture (lecture type learning activities), most students answered that listening to teachers’ explanation or copying what the teacher wrote on the board happened almost always (number 1) and very often (number 2). The means of PLATlecture for Canadian and Korean samples were 1.40 (SD=0.64) and 1.53 (SD=0.82), respectively. There existed some variations among schools (refer to Figure 5.1), but about 90% of the students answered with “almost always” or “often”. Regardless of the schools and countries, the dominant learning activities that the students perceived were lecture-type. Canadian schools (1,3, and 4) had almost the same distributions while the distributions of the three Korean schools varied.

Since about 90% of the answers of PLATlecture belonged to either almost always (score 1) or often (score 2), it lacked capacity to discern why the participants had different understandings of NOS. Its distribution was also heavily positively skewed and leptokurtic, which violated the assumption of normality required in many statistical analyses; the skewness was 2.02 and the kurtosis was 4.4 (greater than the conventional acceptance criterion, 3).
As to PLATquestion (answering questions using formulas or theories by students), most students (70%) from School 1 and Korean schools responded that the activities did not occur while 15% of the school students perceived that the activity occurred often. Meanwhile, about 30% students from the three Canadian schools (2, 3, and 4) regarded the activities quite often. Also, about 30% student of these schools answered the question with “sometimes” and that was different from other schools. Their means were School 2 \((M_2=2.23)\) School 3 \((M_3=2.88)\) and School 4 \((M_4=2.89)\). Figure 5.2 indicates the distributions of PLATquestion. School 1 and School 7 had similar distributions although they were from different countries.

*Figure 5-1. Distribution of PLATlecture*

*Figure 5-2. Distribution of PLATquestion*
The mean of *PLATlab* (the activities of science lab experiments) was 2.31 and SD was 1.17. Examining Figure 5.3, Canadian and Korean schools distributed differently. More Korean students thought that they did teacher-guided or teacher-demonstrated lab experiments than Canadian students did. About 80% of the students from School 6 responded that they had lab activities quite often. The students from this school perceived that they did not have many opportunities to have time to solve questions by themselves but they had frequently lab activities or teacher demonstration of lab activities. About 40% of students from School 1, 3 and 4 students responded to the question with “3, sometimes.” Less than 5% of students from School 4 perceived that they had lab activities very often, while 25% thought that the activities occurred rarely. The means for Canadian and for Korean schools ranged from 2.69 to 2.96 and from 1.78 to 2.24, respectively.

![Figure 5-3. Distribution of PLATlab](image)

The grand mean of the item, *PLATmemo* (the activities memorizing scientific laws, theories and factual knowledge) was 2.17 and SD was 1.03, which meant that memorization was occurred quite often. Particularly, School 5 had a lowest mean ($M_5=1.94$) of 7 schools whereas School 2 had the highest, $M_2=2.70$. As Figure 5.4 show, more Korean students
thought that they were asked to memorize scientific laws and theories than did Canadian students.

In Item 5 (PLATpresentation), School 3 and School 4 had high means ($M_3=2.92$, $M_4=2.82$) while School 6 had a low mean ($M_6=2.23$). Canadian schools seemed to have higher means than the Korean schools. That is, Canadian school students perceived that they performed more often the activities and tasks of presentations/discussion than the Korean students did ($M_{Ca} \ vs \ M_{Ko}=2.73 \ : \ 2.21$). A noticeable thing was a comprehensive portion of students from Canadian schools answered the question with “sometimes.” On the other hand, about 40% of the Korean students answered “hardly ever.”

The responses of PLATpresentation were a skewed distribution, but it was not as severe as PLATlecture. Its skewness was .0.57 and the kurtosis was -0.50, which was much less than absolute value 3. Thus, PLATlecture (Item1) was excluded in further analysis but PLATpresentation (Item5) was included.
The figures 5.1~5.5 roughly showed students’ distributions across 5 items in school levels. Table 5.2 also includes the details such as means, standard deviations and frequencies for the items in the schools. There exist mean differences among schools and countries, and to confirm whether these differences are statistically significant or not, further analyses like multivariate analysis of variance (MANOVA) are required.

**Table 5-2**

*Frequencies, Means and Standard Deviations for PLAT*

\[
N_1=217 \text{ (Canadian participants)}, \ N_2=319 \text{ (Korean participants)}
\]

<table>
<thead>
<tr>
<th>Item</th>
<th>Country</th>
<th>School</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture 1</td>
<td></td>
<td></td>
<td>1</td>
<td>40</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>30</td>
<td>17</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>44</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>30</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>104</td>
<td>10</td>
<td>18</td>
<td>4</td>
<td>2</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>54</td>
<td>15</td>
<td>16</td>
<td>5</td>
<td>3</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>69</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1.33</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>371</td>
<td>100</td>
<td>47</td>
<td>10</td>
<td>8</td>
<td>1.48</td>
<td>0.85</td>
</tr>
</tbody>
</table>

| Solving questions 1 |         |        | 1  | 16 | 21 | 7  | 6  | 3    | 2.23| 1.17|
|                     |         |        | 2  | 8  | 14 | 17 | 8  | 6    | 2.81| 1.21|
|                     |         |        | 3  | 9  | 19 | 19 | 7  | 11   | 2.88| 1.28|
|                     |         |        | 4  | 8  | 10 | 14 | 5  | 8    | 2.89| 1.34|
### 5.4.2 Perceptions on Assessment

Seven items measured how students perceived different types of assessment formats (three items: close-ended, open-ended questions and alternative assessments) and test content (four items: questions asking about scientific laws, theories and factual knowledge, applications, lab experiments and social, cultural aspects of science). As with the students’ perceptions of learning activities and tasks (PLAT), the items were rearranged for conceptual consistency so that knowledge-focused assessment answers would be ranked with lower scores.
while understanding- and- application- focused assessment answers would have higher scores. Thus, Item 3 (PAFORMalternative: assessment format: alternative methods of assessment), Item 5 (PACONTapplication: assessment content asking to apply the knowledge which learned during science classes), and Item 7 (PACONTphs: assessment content asking social and cultural aspects of science) were arranged in reverse for the inferential statistics.

Taking the item, PAFORMATclose, which asked students’ perceptions on their close-ended questions (multiple-choice, true/false or matching correct answers), most students across the schools responded that they regarded the assessment formats as “most important” or “important” (70%), whereas few considered them “little” or “least” (6.7%). The mean over all schools was 1.94 and the standard deviation was 1.02. Figure 5.6 illustrates students’ responses to the item. The students from School 3 and 4 had small portions of students who agreed that the item was most important, but still they had a large number of students that the item importantly affected their science test scores. School 4 had the highest mean ($M_4=2.82$ and $SD=1.21$) while School 2 and School 6 had very low means, $M_2=1.67$ and $M_6=1.65$, respectively.

![Figure 5-6. Distribution of PAFORMATclose](image)
Responses to *PAFORMATshort*, short answering test format, were similar to the responses of the close-ended format. The overall mean was 2.16 and the standard deviation was 1.13. There were very few cases answering as “least important” (2.0%). Figure 5.7 shows the distribution of the responses; about 80% of students from Schools 1, 2 and 3 were likely to think that short answering format importantly affect their science test scores. About 70% of the students from Schools 5 and 6 students and 55% of the students from Schools 4 and 7 thought that the format affected their test scores.

![Figure 5-7. Distribution of *PAFORMATshort*](image)

*PAFORMATalternative*, asking alternative formats, meant the assessments such as lab-reports and science projects assessment. There were two unique distributions in this item. One was that the participants from School 4 (67%) showed the opposite of answers from other schools. A significant proportion of this school’s students believed this format had affected their science scores, in contrast to the students of other schools (the average percentage of “most important” and “important” from the other schools was 20%). Another conspicuous feature was that high portions of Korean students responded to the item with “neutral.” As
mentioned in the analysis of *PLAT* presentation, the responses needed a clarification whether the student really meant “neutral” or an easy picking answer.

![Figure 5-8. Distribution of PAFORMATalternative](image)

The overall means for the three items were $M_{\text{close}}=1.97$, $M_{\text{short}}=2.16$ and $M_{\text{alternative}}=2.58$. That is, alternative format was not as important as traditional paper-and-pencil test format (close-ended or short answering questions) in most schools except School 4.

For *PACONT* knowledge, perceptions of science test content asking scientific laws, theories, and scientific factual knowledge, the grand mean was $M=2.47$, and $SD=1.24$. Figure 5.9 depicts the distributions of students’ responses. Relatively, the Korean students weighed more on the content on knowledge of science ($M_{Ko}=2.14$) than the Canadian students did ($M_{Ca}=2.96$). Particularly School 4 students had the highest mean ($M_4=3.31$), which meant that they regarded the content less important in their test scores than other school students.
Concerning PACONT application, the content of knowledge application to diverse situations, the overall mean was 3.23 and the standard deviation was 1.32. As this item was ordered in reverse, answers 1 and 2 meant “least important” and “little important” whereas answers 4 and 5 meant “most important” and “important.” As the diagram shows (Figure 5.10), the Korean students seemed to perceive that the test content of applying knowledge into diverse situations were less important than the Canadian students were. The students from School 3 and 4 had more scores of 4 and 5 in comparison with other schools. Schools 1 and 2 had similar distributions while Schools 3 and 4 had similar trends. Unlike Canadian schools, the three Korean schools had similar distributions.
For Korean students, lab activities occurred more frequently than did the Canadian students, and the students responded that the test content about science experiments was very important. The students from School 1 and School 2 perceived similarly whereas the students from Schools 3 and 4 did respond similarly. About 65% of the Korean students regarded the content was important while about 30% of Canadian students did. The results were consistent with the perceptions of lab activities.

Figure 5-11. Distribution of PACONTlab
Testing on diverse aspects of science, i.e., historical, cultural and social aspects of science, \textit{\textsc{PACONT}phs}, was not important across all schools. The distribution was heavily skewed. Figure 5.12 shows that 95\% of Korean students (304 out of 319) answered that the content did not significantly affect their science test scores. Also, 72\% of Canadian students (156 out of 217) answered that the content was not important. With respect to the explicity of the item, the test content of social, cultural and historical aspects of science could not explain why students’ understandings of NOS were different. Like \textit{\textsc{PLAT}ecture}, this item was excluded in the further analysis.

\begin{center}
\textbf{Figure 5-12. Distribution of \textsc{PACONT}phs}
\end{center}

\begin{table}[h]
\centering
\caption{Frequencies, Means and Standard Deviations for \textsc{PAFORMAT} & \textsc{PACONT}}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\multicolumn{2}{|c|}{Item} & \multicolumn{5}{|c|}{School} & Mean & SD \\
\hline
\multicolumn{2}{|c|}{Country} & 1 & 2 & 3 & 4 & 5 & \\
\hline
1 & \textit{\textsc{PAFORMAT}close} & 1 & 27 & 14 & 7 & 5 & 0 & 1.81 & 1.00 \\
& & 2 & 27 & 16 & 8 & 1 & 0 & 1.67 & 0.81 \\
& & 3 & 12 & 27 & 17 & 6 & 3 & 2.40 & 1.04 \\
& & 4 & 6 & 14 & 12 & 8 & 5 & 2.82 & 1.21 \\
& & 5 & 69 & 34 & 32 & 2 & 1 & 1.78 & 0.90 \\
& & 6 & 56 & 18 & 16 & 2 & 1 & 1.65 & 0.92 \\
& & 7 & 40 & 13 & 33 & 2 & 0 & 1.97 & 0.96 \\
& Total & 237 & 136 & 125 & 26 & 10 & 1.94 & 1.02 \\
\hline
\end{tabular}
\end{table}
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>26</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>2.08</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>24</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>2.17</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>33</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>2.23</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>2.58</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>38</td>
<td>27</td>
<td>13</td>
<td>2</td>
<td>2.01</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>30</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td>1.95</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>22</td>
<td>25</td>
<td>14</td>
<td>2</td>
<td>2.39</td>
</tr>
<tr>
<td>Total</td>
<td>175</td>
<td>190</td>
<td>90</td>
<td>68</td>
<td>11</td>
<td>2.16</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>9</td>
<td>13</td>
<td>5</td>
<td>5</td>
<td>2.32</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>18</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>2.46</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>18</td>
<td>10</td>
<td>16</td>
<td>10</td>
<td>2.94</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>8</td>
<td>22</td>
<td>3.80</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>13</td>
<td>62</td>
<td>26</td>
<td>4</td>
<td>2.67</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>25</td>
<td>27</td>
<td>11</td>
<td>1</td>
<td>2.25</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>16</td>
<td>49</td>
<td>7</td>
<td>1</td>
<td>2.58</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>107</td>
<td>177</td>
<td>81</td>
<td>45</td>
<td>2.66</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>9</td>
<td>18</td>
<td>14</td>
<td>3</td>
<td>2.87</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>17</td>
<td>16</td>
<td>10</td>
<td>2</td>
<td>2.60</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>10</td>
<td>17</td>
<td>15</td>
<td>5</td>
<td>3.17</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>7</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>3.31</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>43</td>
<td>32</td>
<td>7</td>
<td>4</td>
<td>2.04</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>30</td>
<td>16</td>
<td>8</td>
<td>7</td>
<td>2.23</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>28</td>
<td>22</td>
<td>7</td>
<td>2</td>
<td>2.15</td>
</tr>
<tr>
<td>Total</td>
<td>147</td>
<td>144</td>
<td>130</td>
<td>74</td>
<td>41</td>
<td>2.47</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>19</td>
<td>25</td>
<td>3.74</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
<td>12</td>
<td>19</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>14</td>
<td>12</td>
<td>3.58</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>33</td>
<td>43</td>
<td>28</td>
<td>17</td>
<td>2.99</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>12</td>
<td>19</td>
<td>30</td>
<td>21</td>
<td>3.42</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>15</td>
<td>25</td>
<td>17</td>
<td>26</td>
<td>3.50</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>113</td>
<td>112</td>
<td>130</td>
<td>115</td>
<td>3.23</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>25</td>
<td>4</td>
<td>10</td>
<td>3</td>
<td>2.42</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>24</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>2.14</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>20</td>
<td>2</td>
<td>12</td>
<td>20</td>
<td>3.15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>11</td>
<td>7</td>
<td>13</td>
<td>11</td>
<td>3.40</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>41</td>
<td>20</td>
<td>3</td>
<td>1</td>
<td>1.68</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>21</td>
<td>16</td>
<td>5</td>
<td>7</td>
<td>2.03</td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>18</td>
<td>19</td>
<td>6</td>
<td>2</td>
<td>1.93</td>
</tr>
<tr>
<td>Total</td>
<td>198</td>
<td>160</td>
<td>72</td>
<td>56</td>
<td>50</td>
<td>2.25</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>14</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>1.85</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>1.94</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1.97</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>17</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2.07</td>
</tr>
<tr>
<td>5</td>
<td>63</td>
<td>66</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1.67</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>55</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1.78</td>
</tr>
</tbody>
</table>
# Explanation of items
- **PAFORMATclose**: assessment format-multiple choice/true or false questions
- **PAFORMATshort**: assessment format-short answers to open-ended questions
- **PAFORMATalternative**: assessment format-alternative methods
- **PACONTknowledge**: assessment content-scientific laws, theories and factual knowledge
- **PACONTapplication**: assessment content-application of scientific knowledge in diverse situations
- **PACONTlab**: assessment content-lab activities
- **PACONTphs**: assessment content-philosophical/historical/social aspects of science

## 5.4.3 Understanding of NOS

Guided by the definition of the NOS concepts, five major concepts of NOS were measured using 15 items. The concepts were: 1) the tentative nature of science, 2) theory-laden observations, 3) empirical evidence based science, 4) socio-culturally embedded science and scientific knowledge and 5) diversity in scientific methods. Each concept was measured by three observed variables. Negative, relativistic statements were inversely arranged for conceptual consistency in the analyses. The data were distributed normally and no heavily skewed items were found; thus, all items were included in the analyses (Refer to Figure 5.13 and Table 5-4). The analysis procedures for NOS resembled the previous analyses of **PLAT**, of **PAFORMAT** and **PACONT**. Descriptive statistics provided general information of the samples.

The tentative nature of science (**Tentativeness**) asked students whether scientific knowledge 1) has been changed due to reinterpretation of previous observation, data or evidence (**TEN1**), 2) can be error-free (**TEN2**), and 3) approaches absolute truth (**TEN3**). The grand means of **TEN1**, **TEN2** and **TEN3** were 3.78, 3.80 and 3.56. These mean scores meant that most students agreed that scientific knowledge has been changed and will change by new
interpretations of existing or newly observed data. Students also agreed that scientific knowledge includes some errors and does not approach absolute truth. In discussing change in scientific knowledge, Lederman et al. (2002) and Lederman (2007) argue that a naïve view of the tentative nature of science supposes that only new discoveries or new evidence makes scientific knowledge change rather than new interpretations of existing data. In this research, 58% of the students answered to the first question, TENI (inversely arranged), a new interpretation of data can change current scientific knowledge, with strongly agree or agree. Because the question did not emphasize “interpretation of data,” the students might focus on “change” of scientific knowledge. The unclear result could be clarified by their answers to the open-ended questions; i.e., what the students meant by the “change” in the close-ended question. In Figure 5.13 (a) the distributions of the three observed variables skew slightly negatively.

The items concerning Subjectivity (theory-laden observations) asked whether 1) scientists’ prior knowledge affects their observation of data (SUB1), 2) scientific observations are independent from theories (SUB2) and 3) scientists are objective (SUB3). The answers distributed evenly, and the means of SUB1, SUB2 and SUB3 were 2.98, 2.78 and 2.78; their SDs were 1.40, 1.33 and 1.40, respectively. The overall means were over 2.50 (the middle score); students seemed to admit to theory-laden observations although the trends were not obvious. In all three variables of Subjectivity, School 3 (M3=3.55) and School 4 (M4=3.82) had higher means than other schools, and Canadian schools had also higher means than did the Korean schools. The higher mean meant that students seemed to agree to the ideas that scientific observations are more or less theory-laden.
EMP1, 2, and 3 measured students’ understanding of scientific knowledge that is based on empirical evidence. Among the three variables of EMP, EMP3 ($M_{EMP3}=2.39$) was the lowest mean in comparison with EMP1 ($M_{EMP1}=2.66$) and EMP2 ($M_{EMP2}=2.67$). School 4 ($M_{4}=2.95$) and School 7 ($M_{7}=2.77$) had relatively high average means across these 3 variables while School 5 had the lowest average ($M_{5}=2.32$). This empirical basis of NOS had a relatively low mean ($M_{EMP}=2.57$) among the 5 constructs ($M_{TEN}=3.62$, $M_{SUB}=2.86$, $M_{CUL}=2.54$, $M_{METH}=3.50$). The responses to these three items were noticeably consistent, so that students strongly held the notion that scientific knowledge should be supported by evidence.

Measuring the construct of sociocultural-embedded scientific knowledge involved posing questions about the effects of culture, politics and religion 1) scientists, 2) on their research methods including what and how they study and 3) on what scientific results are deemed acceptable. The histograms (Figure 5.13 (d) CUL1, CUL2, CUL3) show different distributions among the three items. The students chose more “Disagree” or “Strongly disagree” in CUL1 and CUL2 while they chose more “Agree” or “Strongly agree” in CUL3. Students thought that scientists were also members of a society and their ways of research and their selections of research topics were affected by society, culture and politics. However, students’ responses to CUL3 that asked the judgment of scientific knowledge were inclined to agreement; i.e., slightly positively skewed. These distributions reflected the means of CUL1 ($M_{CUL1}=3.24$) and CUL2 ($M_{CUL2}=3.51$) while the mean of CUL3 was 2.91.

Regarding diversity in scientific research methods, students agreed that scientists conduct their scientific research in diverse ways. As the histograms show, most students had relativistic perspectives on scientific methods. They (61%) rejected the notion of stereotyped step-by-step methods ($M_{METH2}$) and 58% agreed that methods differ according to the research
pursued (*METH1*). School 4 had the highest mean ($M_4=4.01$) while other school ranged from 3.25–3.76. However, with respect to the results more students (32%) agreed the idea that if scientists use correct methods, they could get correct results. Figure 5.3 (a–e) and Table 5.4 include histograms on each variable distribution, frequencies in each variables, and school means and standard deviations.

<table>
<thead>
<tr>
<th>TEN1</th>
<th>TEN2</th>
<th>TEN3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>TEN1</td>
<td>18</td>
<td>37</td>
</tr>
<tr>
<td>TEN2</td>
<td>17</td>
<td>56</td>
</tr>
<tr>
<td>TEN3</td>
<td>60</td>
<td>70</td>
</tr>
</tbody>
</table>

*Figure 5.13 (a) Tentativeness*

<table>
<thead>
<tr>
<th>SUB1</th>
<th>SUB2</th>
<th>SUB3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>SUB1</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>SUB2</td>
<td>113</td>
<td>134</td>
</tr>
<tr>
<td>SUB3</td>
<td>129</td>
<td>115</td>
</tr>
</tbody>
</table>

*Figure 5.13 (b) Subjectivity*
Table 5.13 (c) shows the empirical evidence based scientific knowledge. The table compares the means and standard deviations for EMP1, EMP2, and EMP3 for measures 1 to 5.

<table>
<thead>
<tr>
<th></th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMP1</td>
<td>2.64</td>
<td>1.30</td>
</tr>
<tr>
<td>EMP2</td>
<td>2.65</td>
<td>1.41</td>
</tr>
<tr>
<td>EMP3</td>
<td>2.39</td>
<td>1.36</td>
</tr>
</tbody>
</table>

Table 5.13 (d) presents the sociocultural embeddedness. The table compares the means and standard deviations for CUL1, CUL2, and CUL3 for measures 1 to 5.

<table>
<thead>
<tr>
<th></th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL1</td>
<td>3.24</td>
<td>1.35</td>
</tr>
<tr>
<td>CUL2</td>
<td>3.51</td>
<td>1.19</td>
</tr>
<tr>
<td>CUL3</td>
<td>2.91</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Table 5.13 (e) shows the diverse scientific research methods. The table compares the means and standard deviations for METH1, METH2, and METH3 for measures 1 to 5.

<table>
<thead>
<tr>
<th></th>
<th>MEAN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>METH1</td>
<td>3.68</td>
<td>1.15</td>
</tr>
<tr>
<td>METH2</td>
<td>3.66</td>
<td>1.28</td>
</tr>
<tr>
<td>METH3</td>
<td>3.16</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Figure 5.13. Distributions of NOS Concepts
Table 5-4

*Descriptive Statistics for NOS Concepts*

<table>
<thead>
<tr>
<th>Construct</th>
<th>School</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T1</td>
<td>T1</td>
<td>T2</td>
<td>T2</td>
<td>T3</td>
<td>T3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tentative</td>
<td>1</td>
<td>3.77</td>
<td>1.04</td>
<td>3.58</td>
<td>1.23</td>
<td>3.4</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.62</td>
<td>1.18</td>
<td>3.77</td>
<td>1.09</td>
<td>3.57</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOS</td>
<td>3</td>
<td>4.03</td>
<td>1.02</td>
<td>3.86</td>
<td>1.12</td>
<td>3.66</td>
<td>1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.22</td>
<td>1.06</td>
<td>3.84</td>
<td>1.13</td>
<td>3.96</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.62</td>
<td>1.04</td>
<td>3.66</td>
<td>1.11</td>
<td>3.3</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.6</td>
<td>1.1</td>
<td>3.83</td>
<td>1.01</td>
<td>3.62</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3.88</td>
<td>1.08</td>
<td>4.08</td>
<td>0.97</td>
<td>3.56</td>
<td>1.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3.74</td>
<td>1.08</td>
<td>3.58</td>
<td>1.25</td>
<td>3.55</td>
<td>1.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjectivity</td>
<td>1</td>
<td>3.24</td>
<td>1.33</td>
<td>2.81</td>
<td>1.12</td>
<td>2.85</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.02</td>
<td>1.22</td>
<td>2.67</td>
<td>1.4</td>
<td>2.67</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.68</td>
<td>1.25</td>
<td>3.38</td>
<td>1.4</td>
<td>3.6</td>
<td>1.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.98</td>
<td>1.32</td>
<td>3.53</td>
<td>1.08</td>
<td>3.96</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.67</td>
<td>1.36</td>
<td>2.6</td>
<td>1.34</td>
<td>2.41</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.56</td>
<td>1.38</td>
<td>2.52</td>
<td>1.26</td>
<td>2.18</td>
<td>1.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2.74</td>
<td>1.38</td>
<td>2.59</td>
<td>1.28</td>
<td>2.8</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2.99</td>
<td>1.41</td>
<td>2.78</td>
<td>1.33</td>
<td>2.78</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>School</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence Based Nature of Science</td>
<td>1</td>
<td>2.69</td>
<td>1.16</td>
<td>2.44</td>
<td>1.15</td>
<td>2.69</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.65</td>
<td>1.35</td>
<td>3.22</td>
<td>1.36</td>
<td>2.48</td>
<td>1.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.66</td>
<td>1.4</td>
<td>2.38</td>
<td>1.43</td>
<td>2.6</td>
<td>1.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.84</td>
<td>1.31</td>
<td>3.07</td>
<td>1.44</td>
<td>2.93</td>
<td>1.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.43</td>
<td>1.21</td>
<td>2.46</td>
<td>1.37</td>
<td>2.13</td>
<td>1.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.6</td>
<td>1.35</td>
<td>2.63</td>
<td>1.45</td>
<td>2.01</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2.94</td>
<td>1.35</td>
<td>2.84</td>
<td>1.47</td>
<td>2.53</td>
<td>1.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2.66</td>
<td>1.31</td>
<td>2.67</td>
<td>1.41</td>
<td>2.4</td>
<td>1.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociocultural Embeddedness</td>
<td>School</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3.12</td>
<td>1.4</td>
<td>3.46</td>
<td>1.02</td>
<td>2.98</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.22</td>
<td>1.4</td>
<td>3.26</td>
<td>1.07</td>
<td>2.74</td>
<td>1.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.23</td>
<td>1.36</td>
<td>3.62</td>
<td>1.09</td>
<td>3.09</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.42</td>
<td>1.53</td>
<td>4.11</td>
<td>0.71</td>
<td>3.27</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Students’ perceptions of their learning activities and tasks during science classes showed that almost all students considered listening to teachers’ explanation and copying what the teachers wrote on the blackboard were the dominant class activities. No difference was found across schools in both countries. The item, *PLATlecture*, lacked capacity to discern why students have different understandings of NOS; therefore, it was excluded from further analysis. The least commonly perceived class activities were student presentations and class discussions about scientific issues.

Noticeable differences between Canadian and Korean schools were existed on the items, *PLATquestion* and *PLATlab*. Canadian students responded that they were doing more solving questions by themselves than did the Korean students. On the contrary, Korean students thought they did more teacher-guided or demonstrated experiments. The mean differences across the items between countries as well as among schools were found. Further
analyses on differences (e.g., MANOVA) will confirm whether these differences are statistically significant or not.

Turning to assessment formats, students thought that traditional paper-and-pencil typed close-ended and short-ended formats significantly affected their science scores. Generally, Canadian students weighted more on applying what they learned during classes into diverse situations while Korean students weighted on scientific laws, theories and factual knowledge in their tests. Knowledge about science (historical, social and political aspects of science) was not regarded as important in science tests in either country. It was heavily positively skewed. About 95% of the Korean students responded to the question, whether these contents appeared in science tests, with seldom or hardly. It meant the contents were not included in science tests.

Tentativeness and Diversity of Scientific Research Methods showed similar trends; the answers were more on relativistic views while Empirical Evidence Based Scientific Knowledge inclined more on realistic views. Students felt that social and cultural values motivated scientists what to study and how to study, however, they felt that, to be validated and accepted by other scientists and the public, scientific knowledge should be supported by experiments. Meanwhile, no obvious trends were noticed in Subjectivity.

These descriptive statistics results showed that inferential statistics were inappropriate to apply to the two items, due to the uniformity of answers. So, the following examinations of validity and reliability, the effects of the country and school differences on NOS understandings and associations did not include them.

5.5. Validity and Reliability

It was necessary next to examine whether or not the items fully reflect the concepts that this research intended to measure and whether the items were internally consistent. An analysis
was made both of construct validity (convergent and discriminant validity) and of internal consistency reliability. The starting point for providing validity should be the content validity for the questionnaire (Benson, 1998). The theoretical framework for this research was the basis for its instrument questions, and the sections explaining how the instrument was developed and what the questions were about would provide the content validity (Refer to the section 4.5.3 Measuring instrument).

The instrument for this research was designed upon formative and reflective measurement models. Different procedures were employed to examine the validity and reliability. The Average Variance Extracted (AVE) was used for convergent validity, and Fornell-Larcker criterion and cross-loadings were used for discriminant validity in examining reflective measurement models of understanding NOS concepts. The significances of outer weights and multicollinearity of indicators (VIF: variance inflation factor) were applied in examining formative measurement models (PLAT, PAFORMAT and PACONT). To test internal consistency reliability, composite reliability was applied to reflective models. Later, in examining association, the structural models were assessed by \( R^2 \) of endogenous latent variables, inner path coefficients (\( \gamma \)), effect size (\( f^2 \)) of \( R^2 \), and predictive relevance (\( Q^2 \)) and the effect size (\( q^2 \)) of \( Q^2 \).

Figure 5.14 depicts the theoretical model for this research (The same with Figure 3-1). The blue circles represent latent variables and the yellow rectangles represent observed variables (indicators). The observed variables for the independent latent variables (PLAT, PAFORMAT and PACONT) are on the left hand side and the indicators of the latent dependent variables (NOS concepts) are on the right hand side. As depicted in the figure, the arrows of formative models directed from indicators to their latent variables (PLAT, PAFORMAT and
and the arrows of reflective models directed from latent variables to their indicators.

The Canadian, Korean and Combined (both countries) samples tested the theoretical model separately.

![Figure 5-14. Theoretical Model](image)

### 5.5.1 Reflective Measurement Model: Understanding of NOS concepts

Convergent validity was assessed by applying three criteria: (1) The AVE (average variance extracted) of a construct should be at least 0.5, 2) The standardized item loadings should be at least 0.70 and 3) the correlations within constructs should exceed the correlation with other constructs (items should load more highly on constructs they are intended to measure than on other constructs) (Agarwal & Karahanna, 2000).

#### 5.5.1.1 Canadian Sample

Concerning internal consistency reliability, the composite reliability was examined and all values for 5 concepts of NOS were larger than 0.70. Particularly for *Empirical evidence based scientific knowledge* and *Diverse Scientific Methods* were over 0.8 while *Tentativeness* and *Sociocultural Embeddedness* were 0.78. All five constructs’ composite reliabilities ($\rho_c$) are in Table 5.5.
Concerning the observed variable reliability, most outer loadings were over 0.7, the optimal value (0.7=square root 0.5 (AVE)), except CUL1 ($\lambda_{CUL1}=0.63$), EMP1 ($\lambda_{EMP1}=0.63$), and TEN1 ($\lambda_{TEN1}=0.68$). Henseler et al. (2009) recommend that “[E]liminating reflective indicators [observed variables] from measurement models if their outer standardized loadings are smaller than 0.4. … Only if an indicator’s reliability is low and eliminating this indicator goes along with a substantial increase of composite reliability” (p.299, italic original). These three observed variables’ loadings were greater than 0.4. Thus, rather than elimination of the variables, I decided to retain them.

With regard to convergent validity, all 5 NOS concepts’ AVE values were larger than the cutoff value, 0.5. Subjectivity had the greatest value (0.62) while Tentativeness had the smallest, 0.54. In other words, more than 50% of the variances of the NOS concepts were explained by their observed variables; thus, the convergent validity of NOS was well established for the Canadian sample.

Meanwhile, the discriminant validity was examined using Fornell-Larcker criterion (1981). Any other squared correlation coefficients were not greater than the construct’s AVE values. For Tentativeness, the AVE value was 0.54 (the smallest AVE) and it was highly correlated with Sociocultural Embeddedness ($r=0.45$). The squared correlation value was $r^2=0.45^2=0.20$, which was smaller than the AVE value, 0.54. AVE for Subjectivity (theory-laden scientific observation) was 0.62 and it was also highly correlated with Diverse Methods. The correlation coefficient was .34 and squared value was 0.12. The AVE value of Empirical Evidence Based Scientific Knowledge construct was 0.57 and the construct was highly correlated with Diverse Methods ($r=0.32$). As examined, the AVE was larger than square of the correlation coefficient ($r^2=0.10$). Regarding Cultural Embeddedness, the AVE value was
0.59, and the construct was highly correlated with *Diverse Methods* (r = 0.45 and $r^2 = 0.21$). The AVE value of *Diverse Methods* was 0.62, which was higher than any other squared correlation coefficient with other constructs. The results meant the observed variables discriminated the targeting construct from other constructs.

Table 5.5 indicates the loadings of the reflective models. All of the loadings within (highlighted with yellow) were greater than all of the loadings between (not highlighted). In other words, the observed variables measured the targeting constructs but not other constructs. The discriminant validity for the NOS concepts was well established for Canadian data. The table also includes composite reliability and AVE values and correlations.

Table 5-5

*Cross Loadings, Composite Reliability, AVE, Correlations (Canadian)*

$N_i = 217$

<table>
<thead>
<tr>
<th></th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
<th>$\rho_c$</th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL1</td>
<td>0.63</td>
<td>0.16</td>
<td>0.35</td>
<td>0.19</td>
<td>0.34</td>
<td>0.78</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUL2</td>
<td>0.80</td>
<td>0.00</td>
<td>0.22</td>
<td>0.32</td>
<td>0.36</td>
<td></td>
<td></td>
<td>0.03</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUL3</td>
<td>0.77</td>
<td>-0.08</td>
<td>0.24</td>
<td>0.25</td>
<td>0.30</td>
<td></td>
<td></td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP1</td>
<td>0.18</td>
<td>0.63</td>
<td>0.27</td>
<td>0.26</td>
<td>0.22</td>
<td></td>
<td></td>
<td>0.01</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP2</td>
<td>-0.03</td>
<td>0.87</td>
<td>0.29</td>
<td>0.08</td>
<td>0.11</td>
<td></td>
<td></td>
<td>0.79</td>
<td>0.03</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>EMP3</td>
<td>0.01</td>
<td>0.74</td>
<td>0.16</td>
<td>0.11</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METH1</td>
<td>0.25</td>
<td>0.26</td>
<td>0.74</td>
<td>0.27</td>
<td>0.35</td>
<td></td>
<td></td>
<td>0.83</td>
<td>0.35</td>
<td>0.32</td>
<td>0.61</td>
</tr>
<tr>
<td>METH2</td>
<td>0.35</td>
<td>0.21</td>
<td>0.80</td>
<td>0.25</td>
<td>0.35</td>
<td></td>
<td></td>
<td>0.35</td>
<td>0.35</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>METH3</td>
<td>0.22</td>
<td>0.29</td>
<td>0.81</td>
<td>0.29</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUB1</td>
<td>0.40</td>
<td>0.15</td>
<td>0.30</td>
<td>0.79</td>
<td>0.37</td>
<td></td>
<td></td>
<td>0.83</td>
<td>0.35</td>
<td>0.16</td>
<td>0.34</td>
</tr>
<tr>
<td>SUB2</td>
<td>0.24</td>
<td>0.12</td>
<td>0.24</td>
<td>0.77</td>
<td>0.29</td>
<td></td>
<td></td>
<td>0.83</td>
<td>0.35</td>
<td>0.16</td>
<td>0.34</td>
</tr>
<tr>
<td>SUB3</td>
<td>0.17</td>
<td>0.12</td>
<td>0.26</td>
<td>0.80</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEN1</td>
<td>0.42</td>
<td>0.06</td>
<td>0.29</td>
<td>0.30</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEN2</td>
<td>0.26</td>
<td>0.14</td>
<td>0.33</td>
<td>0.30</td>
<td>0.71</td>
<td></td>
<td></td>
<td>0.78</td>
<td>0.45</td>
<td>0.19</td>
<td>0.45</td>
</tr>
<tr>
<td>TEN3</td>
<td>0.30</td>
<td>0.22</td>
<td>0.37</td>
<td>0.28</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The bold and italic numbers in the diagonal are AVE.

* The numbers in off diagonal are correlation coefficients.
5.5.1.2 Korean Sample

The values of the composite consistency reliability for 5 constructs were satisfactory; the smallest value was $\rho_c=0.776$ for Subjectivity, while the greatest was Empirical Evidence Based Scientific Knowledge ($\rho_c=0.82$). Table 5.6 includes the constructs’ internal reliability consistency, composite reliabilities ($\rho_c$). For the convergent validity, the results were similar to Canadian sample. The AVE values were larger than 0.5; particularly, Empirical Evidence Based was the highest, 0.61. More than 50% of the variances of the five concepts of NOS were explained by their observed variables.

For the observed variable reliability, the 12 absolute standardized outer loadings were greater than 0.7 while 3 loadings were not: METH1 was $\lambda_{METH1}=0.50$; SUB2 was $\lambda_{SUB2}=0.67$; and TEN1 was $\lambda_{TEN1}=0.65$. However, the composite reliability for the three constructs were not severely low, and the loadings of these three variables were greater than 0.4, so they were retained.

With regard to the discriminant validity of Fornell-Larcker criterion, the highest correlation was between Tentativeness and Empirical Evidence Based Scientific Knowledge ($r=0.38$) and the squared value was 0.14 ($r^2=0.38^2=0.14$). The AVEs were greater than the squared correlations. In addition, the cross loadings (Table 5.6) showed the loadings of within the observed variables and their targeting constructs were much greater than the between cross-loadings. Based on these results, the construct validity of the reflective models was established for Koran samples.
### Table 5-6

**Cross Loadings, Composite Reliability, AVE and Correlations (Korean)**

\[ N_3 = 319 \]

<table>
<thead>
<tr>
<th></th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
<th>( \rho_c )</th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL1</td>
<td>0.72</td>
<td>0.29</td>
<td>0.17</td>
<td>0.30</td>
<td>0.15</td>
<td>0.82</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUL2</td>
<td>0.80</td>
<td>0.26</td>
<td>0.21</td>
<td>0.31</td>
<td>0.12</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUL3</td>
<td>0.80</td>
<td>0.16</td>
<td>0.29</td>
<td>0.22</td>
<td>0.06</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP1</td>
<td>0.19</td>
<td>0.76</td>
<td>0.17</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP2</td>
<td>0.24</td>
<td>0.86</td>
<td>0.28</td>
<td>0.32</td>
<td>0.41</td>
<td>0.82</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMP3</td>
<td>0.28</td>
<td>0.72</td>
<td>0.18</td>
<td>0.34</td>
<td>0.24</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METH1</td>
<td>0.08</td>
<td>0.11</td>
<td>0.50</td>
<td>0.17</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METH2</td>
<td>0.30</td>
<td>0.26</td>
<td>0.85</td>
<td>0.34</td>
<td>0.16</td>
<td>0.78</td>
<td>0.29</td>
<td>0.28</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>METH3</td>
<td>0.21</td>
<td>0.22</td>
<td>0.83</td>
<td>0.29</td>
<td>0.22</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUB1</td>
<td>0.23</td>
<td>0.32</td>
<td>0.22</td>
<td>0.77</td>
<td>0.13</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUB2</td>
<td>0.25</td>
<td>0.21</td>
<td>0.41</td>
<td>0.67</td>
<td>0.20</td>
<td>0.78</td>
<td>0.35</td>
<td>0.37</td>
<td>0.38</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>SUB3</td>
<td>0.29</td>
<td>0.28</td>
<td>0.21</td>
<td>0.76</td>
<td>0.27</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEN1</td>
<td>0.03</td>
<td>0.24</td>
<td>0.00</td>
<td>0.20</td>
<td>0.65</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEN2</td>
<td>0.17</td>
<td>0.33</td>
<td>0.21</td>
<td>0.24</td>
<td>0.78</td>
<td>0.78</td>
<td>0.14</td>
<td>0.37</td>
<td>0.20</td>
<td>0.28</td>
<td>0.54</td>
</tr>
<tr>
<td>TEN3</td>
<td>0.09</td>
<td>0.24</td>
<td>0.19</td>
<td>0.18</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The bold and italic numbers in the diagonal are AVE.
* The numbers in off diagonal are correlation coefficients.

#### 5.5.1.3 Combined Sample

Canadian and Korean data were combined into one sample (Combined sample size were \( N_3 = 536 \)). This process can be another cross validation for the instrument. All composite reliabilities of the constructs (\( \rho_c \)) were greater than 0.7. The smallest one was Tentativeness (\( \rho_c = 0.77 \)), and the greatest one was Subjectivity (\( \rho_c = 0.82 \)). Other composite values are in Table 5.7. The outer loading of METHI was \( \lambda_{\text{METHI}} = 0.64 \), smaller than the recommended standard. But it was greater than the elimination standard, 0.4. All other observed variables’ outer loadings were greater than 0.7.
All AVE values were greater than 0.5. Regarding Fornell and Larcker criterion (1981), the AVE values for each construct were greater than any other squared correlation coefficients with other constructs; the greatest correlation coefficient was $r=0.35$ between Sociocultural Embeddedness and Subjectivity); the squared value was 0.14 ($r=0.35$, $r^2=0.12$) and their AVEs were 0.58 and 0.60, respectively. Cross-loadings also supported that the observed variables measured the targeting constructs rather than other constructs. Table 5.7 represents the cross-loadings. The highlighted coefficients (within) were greater than the other coefficients (between).

Table 5-7

Cross Loadings, Composite Reliability, AVE, Correlations (Combined Sample)

$N=536$

<table>
<thead>
<tr>
<th></th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
<th>$\rho_c$</th>
<th>CUL</th>
<th>EMP</th>
<th>METH</th>
<th>SUB</th>
<th>TEN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CUL1</strong></td>
<td>0.70</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CUL2</strong></td>
<td>0.80</td>
<td>0.19</td>
<td>0.21</td>
<td>0.31</td>
<td>0.21</td>
<td>0.81</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CUL3</strong></td>
<td>0.78</td>
<td>0.09</td>
<td>0.27</td>
<td>0.24</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EMP1</strong></td>
<td>0.19</td>
<td>0.71</td>
<td>0.21</td>
<td>0.23</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EMP2</strong></td>
<td>0.15</td>
<td>0.80</td>
<td>0.28</td>
<td>0.22</td>
<td>0.20</td>
<td>0.82</td>
<td>0.23</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EMP3</strong></td>
<td>0.19</td>
<td>0.81</td>
<td>0.19</td>
<td>0.28</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>METH1</strong></td>
<td>0.15</td>
<td>0.18</td>
<td>0.64</td>
<td>0.23</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>METH2</strong></td>
<td>0.32</td>
<td>0.25</td>
<td>0.83</td>
<td>0.35</td>
<td>0.21</td>
<td>0.80</td>
<td>0.32</td>
<td>0.30</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>METH3</strong></td>
<td>0.22</td>
<td>0.24</td>
<td>0.80</td>
<td>0.28</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUB1</strong></td>
<td>0.30</td>
<td>0.28</td>
<td>0.29</td>
<td>0.80</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUB2</strong></td>
<td>0.25</td>
<td>0.19</td>
<td>0.35</td>
<td>0.72</td>
<td>0.21</td>
<td>0.82</td>
<td>0.34</td>
<td>0.32</td>
<td>0.38</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td><strong>SUB3</strong></td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.79</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TEN1</strong></td>
<td>0.17</td>
<td>0.09</td>
<td>0.13</td>
<td>0.24</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TEN2</strong></td>
<td>0.17</td>
<td>0.18</td>
<td>0.23</td>
<td>0.28</td>
<td>0.73</td>
<td>0.77</td>
<td>0.24</td>
<td>0.22</td>
<td>0.25</td>
<td>0.32</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>TEN3</strong></td>
<td>0.18</td>
<td>0.20</td>
<td>0.18</td>
<td>0.17</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The bold and italic numbers in the diagonal are AVE.

* The numbers in off diagonal are correlation coefficients.
5.5.2 Formative Measurement Model: PLAT, PAFORMAT and PACONT

Multicollinearity (Reliability): Many scholars insist that classical test theory does not apply to formative measurement models (Bagozzi, 1994; Bollen & Lenox, 1991; Chin, 1998; Diamantopoulos, 2006; Diamantopoulos & Winklhofer, 2001). Diamantopoulos and Winklhofer (2001) and Diamantopoulos and Siguaw (2006) suggest that the proper reliability evaluation for formative constructs is to assess the assumption of no multicollinearity. Variance Inflation Factor (VIF) is evaluated for the measurement models of the latent variables (PLAT, PAFORMAT and PACONT).

Validity: The fundamental idea in examining the quality of individual observed variables is that each observed variable could be correlated with another variables, and the observed variables that are significantly correlated to a desirable construct should be retained (Bollen & Lennox, 1991; Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003). That is, observed variable weights (an index of correlation between an observed variable and its construct) should be significant. Observed variable weights (outer weight), t-statistics of each weight and the significance were tested using bootstrapping (cases 217 for Canadian, cases 319 for Korean, and 536 for combined data and sample 1000) using SmartPLS software (Ringle et al., 2005). Two-tailed t-tests are considered with 1.645, 1.96, and 2.576 critical t values at significant level (p-value) 0.1, 0.05, and 0.01, respectively (Warner, 2008).

5.5.2.1 Formative measurement models: Canadian Sample

The multicollinearity of the formative measurement models was examined using the scores of latent variables obtained from the PLS outcome. Table 5.8 includes means, standard deviations, condition index (k) and VIF. No item had greater than 10 in VIF value. In addition, any other condition indexes were greater than 30, the cutoff value. These results implied no
concern for multicollinearity problems in the formative measurement models for the Canadian sample.

With respect to the observed variables’ validity test for Canadian sample, the outer loadings ranged from $\lambda_{PACONT\text{application}} = 0.20$ to $\lambda_{PLAT\text{question}} = 0.57$. Most $t$-values were significant at the significant level ($p < .05$) while $PACONT\text{application}$ (assessment content on application of scientific knowledge into diverse situations) was $\lambda_{PACONT\text{application}} = 0.20$, $t(216) = 1.93$, $p = .055$, which was valid within the significant level .1. However, $PLAT\text{presentation}$ (learning activities and task of presentation and discussions on scientific issues) were not significant: $\lambda_{PLAT\text{presentation}} = 0.22$, $t(216) = 1.47$ and $p = .14)$. The lack of contribution to its construct raised a problem that whether the variable should be retained or removed. Bollen and Lennox (1991) and Henseler et al. (2009) recommend that eliminating a formative indicator should be done when the meaning of the targeting construct does not change. If the variable, $PLAT\text{presentation}$, were dropped, the construct meant science class activities and tasks do not include the activities of science projects, presentation and discussion on scientific topics at all and the meaning of the construct has changed. Furthermore retaining the variables is convenient when the results of Canadian and Korean data were compared. Thus, the variable was retained.
Table 5-8

Formative Measurement Models: Multicollinearity (Canadian)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Condition index</th>
<th>Observed variables (Indicator)</th>
<th>Mean</th>
<th>SD</th>
<th>VIF</th>
<th>Outer weight</th>
<th>t-values</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAT</td>
<td>9.61</td>
<td>PLATquestion</td>
<td>2.69</td>
<td>1.29</td>
<td>1.13</td>
<td>0.57</td>
<td>5.37</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATlab</td>
<td>2.80</td>
<td>1.11</td>
<td>1.12</td>
<td>0.42</td>
<td>3.79</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATmemo</td>
<td>2.40</td>
<td>1.05</td>
<td>1.10</td>
<td>0.43</td>
<td>4.33</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATpresentation</td>
<td>2.70</td>
<td>1.21</td>
<td>1.09</td>
<td>0.22</td>
<td>1.47</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>6.37</td>
<td>PAFORMATclose</td>
<td>2.17</td>
<td>1.11</td>
<td>1.33</td>
<td>0.36</td>
<td>3.50</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAFORMATshort</td>
<td>2.25</td>
<td>1.09</td>
<td>1.16</td>
<td>0.35</td>
<td>3.62</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAFORMATalter</td>
<td>2.85</td>
<td>1.41</td>
<td>1.36</td>
<td>0.56</td>
<td>5.05</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>PACONT</td>
<td>8.08</td>
<td>PACONTknowledge</td>
<td>3.18</td>
<td>1.42</td>
<td>2.04</td>
<td>0.20</td>
<td>1.93</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PACONTapplication</td>
<td>2.84</td>
<td>1.41</td>
<td>1.99</td>
<td>0.42</td>
<td>4.18</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

5.5.2.2. Formative measurement model: Korean Sample

As with the Canadian sample, Multicollinearity was not a problem for the Korean sample. The observed variables’ VIF were less than 3, and condition indexes were also less than 30. All outer weights were significant at the alpha level .05 except PACONTknowledge ($\lambda_{PACONTknowledge}=0.08$, $t(318)=0.093$, $p=.353$). Although PACONTknowledge did not significantly contribute to its construct PACONT to compare other observed variables, with the same reason above the variable, PLATpresentation, of the Canadian sample, this variable was retained.

---

4 Condition number (k) equals the square root of the largest eigenvalue ($\lambda_{max}$) divided by the smallest eigenvalue ($\lambda_{min}$):

$$k = \sqrt{\frac{\lambda_{Max}}{\lambda_{Min}}}$$

If there is no collinearity at all, k will be 1. An informal rule of thumb is that if the condition number is 15, multicollinearity is a concern; if it is greater than 30, multicollinearity is a very serious concern (Hensler et al., 2009).
Table 5-9

Formative Measurement Models: Multicollinearity (Korean)  

<table>
<thead>
<tr>
<th>Construct</th>
<th>Condition index</th>
<th>Observed variables (Indicator)</th>
<th>Mean</th>
<th>SD</th>
<th>VIF</th>
<th>Outer weight</th>
<th>t-values</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAT</td>
<td>8.04</td>
<td>PLATquestion</td>
<td>2.39</td>
<td>1.08</td>
<td>1.46</td>
<td>0.35</td>
<td>5.27</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATlab</td>
<td>1.97</td>
<td>1.05</td>
<td>1.43</td>
<td>0.38</td>
<td>6.59</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATmemo</td>
<td>2.01</td>
<td>1.20</td>
<td>1.07</td>
<td>0.34</td>
<td>6.36</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLATpresentation</td>
<td>2.42</td>
<td>1.22</td>
<td>1.14</td>
<td>0.52</td>
<td>9.47</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>8.03</td>
<td>PAFORMATclose</td>
<td>1.85</td>
<td>0.92</td>
<td>1.75</td>
<td>0.45</td>
<td>4.38</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAFORMATshort</td>
<td>2.09</td>
<td>1.05</td>
<td>1.57</td>
<td>0.32</td>
<td>2.94</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAFORMATalter</td>
<td>2.78</td>
<td>1.11</td>
<td>1.15</td>
<td>0.51</td>
<td>5.53</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PACONT</td>
<td>8.08</td>
<td>PACONTknowledge</td>
<td>2.06</td>
<td>1.04</td>
<td>1.01</td>
<td>0.08</td>
<td>0.93</td>
<td>=0.353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PACONTapplication</td>
<td>3.09</td>
<td>1.16</td>
<td>1.12</td>
<td>0.73</td>
<td>9.24</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PACONTLab</td>
<td>1.73</td>
<td>0.89</td>
<td>1.11</td>
<td>0.47</td>
<td>4.84</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

5.5.3. Formative measurement model: Combined Sample

For the combined sample’s formative measurement models, there was no variable over 3 in their VIF. In addition, all conditional indexes were less than 30. The unstable observed variables in Canadian (PLATpresentation) and Korean samples (PACONTLknowledge) were stabilized in the combined sample. This meant that the sample sizes were large enough, all the observed variables significantly contributed to their construct. Table 5.10 includes the detailed information.
Table 5-10

Formative Measurement Models: Multicollinearity (Combined Sample)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Condition index</th>
<th>Observed variables (Indicator)</th>
<th>Mean</th>
<th>SD</th>
<th>VIF</th>
<th>Outer weight</th>
<th>t-values</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLAT</strong></td>
<td>(Mean: 2.33, SD: 0.74)</td>
<td><strong>PLATquestion</strong></td>
<td>2.43</td>
<td>1.21</td>
<td>1.17</td>
<td>0.47</td>
<td>6.57</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PLATlab</strong></td>
<td>2.30</td>
<td>1.17</td>
<td>1.22</td>
<td>0.39</td>
<td>6.13</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PLATmemo</strong></td>
<td>2.17</td>
<td>1.03</td>
<td>1.12</td>
<td>0.35</td>
<td>6.32</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PLATpresentation</strong></td>
<td>2.43</td>
<td>1.23</td>
<td>1.10</td>
<td>0.42</td>
<td>7.89</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td>(Mean: 2.24, SD: 0.89)</td>
<td><strong>PAFORMATclose</strong></td>
<td>2.07</td>
<td>1.06</td>
<td>1.37</td>
<td>0.44</td>
<td>6.42</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PAFORMATshort</strong></td>
<td>2.20</td>
<td>1.08</td>
<td>1.19</td>
<td>0.32</td>
<td>4.79</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PAFORMATalter</strong></td>
<td>2.83</td>
<td>1.33</td>
<td>1.27</td>
<td>0.51</td>
<td>7.55</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td>(Mean: 2.55, SD: 1.01)</td>
<td><strong>PACONTknowledge</strong></td>
<td>2.69</td>
<td>1.27</td>
<td>1.42</td>
<td>0.44</td>
<td>5.32</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PACONTapplication</strong></td>
<td>3.15</td>
<td>1.34</td>
<td>1.54</td>
<td>0.30</td>
<td>3.33</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>PACONTlab</strong></td>
<td>2.49</td>
<td>1.37</td>
<td>1.69</td>
<td>0.53</td>
<td>6.26</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

5.5.4 Summary of Reliability and Validity of Measuring Instrument

The reflective measurement models of NOS concepts were examined for construct validity (convergent and discrimination validity) and internal consistency reliability. All AVE values of the five constructs had greater than 0.5. Also, the AVE values of the constructs were greater than any other squared correlation coefficient (Fornell-Larcker criterion for discriminant validity) in the Canadian, Korean and combined samples. The loadings of the observed variables and targeting constructs were greater than the cross-loadings with other constructs. Regarding internal consistency reliability for the reflective model, the composite reliabilities were examined and all values were greater than 0.70.

The quality of the formative measurement models of the perceptions of learning activities and assessment were assessed focusing on multicollinearity and observed variables’ outer weights. VIF indexes of the observed variables in formative models were much less than the cut off value, 10 and the condition indexes were less than 15; thus, there was no concern
about multicollinearity. The outer loading examination showed that one observed variable from the Canadian sample and one observed variable from the Korean sample did not significantly contribute to their constructs. Although the contributions of the two observed variables to their constructs were not statistically significant, their loadings for the targeting constructs (i.e., PLATpresentation is to PLAT and PACONTapplication is to PACONT) were greater than other constructs. Considering the meaning of constructs, these three observed variables were retained. Most of all, as a cross validation, the examination of the validity and reliability for the combined data showed more solid than the separate models, which had greater sample sizes. This might mean that the weak outer weights could be strong when the sample sizes were large enough.

Overall validity and reliability of the survey were satisfactory since the observed variables measured the intended constructs and the composite reliabilities were greater than 0.7 over the reflective measurement models, and no multicollinearity problem was found for the formative models.

The structural model validity would be examined through the analysis of associations between the independent latent variables (exogenous variables) and dependent latent variables (endogenous variables). This is because one of the aims of this research is to identify the factors, which affect students’ understanding of NOS. As well, the research seeks to establish which factors can predict students’ understanding of NOS. By answering these questions, the validity of the structural models would be diagnosed.

The following sections deal with the differences between two countries and among the schools under consideration by using two-way nested MANOVA.
5.6 Differences

In the descriptive analysis section, the differences existed in the item means and frequencies; further inferential statistics were required to identify if the differences were statistically significant or not. More than one dependent variable was involved in the analyses, and multiple comparisons could increase the risk of committing Type I errors. Therefore, MANOVA was adopted. The factors were country and school, and schools were nested within country. The dependent variables were the 4 observed variables in PLAT, 3 variables in PAFORMAT and 3 variables in PACONT and the NOS concepts.

5.6.1 Multivariate Analysis of Variance (MANOVA): PLAT

Differences of means in PLAT items were observed in the descriptive statistics analysis between the two countries and among schools. It was important to establish whether or not these differences originated from the schools rather than the countries. Confirming the origin of the differences was important to clarify teacher effects on students’ understanding of NOS because not only what to teach and how to teach a subject does strongly depends on a teacher’s decisions, but also the effects of school differences might be the effects of teacher differences. If there were school differences on learning activities and tasks, and no difference in students’ understanding of NOS, then, the present study could provide further evidence that teacher influence on students’ understanding of NOS is minor, which Lederman (1986) and Lederman (2007) argued. On the contrary, if the school effect on students’ understanding of NOS is significant, the study could be contradicting the findings of those researchers.

The MANOVA assumptions of linearity, homogeneity, independence among participants and normality were checked. The three variables distributed approximately normally; following the conventional rules on non-normal distributions, over absolute value of
3 for kurtosis and of 2 for skewness, there were no serious violations of the distributions. The limitation of the analyses was that the variables of PLAT were not actually continuous. Instead they consisted of a 5-point-Likert scale, so scatter plots did not show any clear linear relationships between paired variables. However, most of the Pearson correlation coefficients between dependent variables were statistically significant, which meant that the linearity indexes, $R^2$, were significant. Box’s $M$ tests (using $\alpha=.001$ as the criterion for significance\(^5\)) showed the homogeneity assumption of variance/covariance matrices across groups was not violated (Box’s $M=84.70$, $F(60,217343)=1.38$, $p=.030$). As it was not significant, Wilks’ lambda\(^6\) was adopted as the multivariate test statistics to test the null hypotheses. For the independence of the variables was established: 1) the students were from different schools, and 2) this study was not designed repeated measures.

The overall $F$ tests were illustrated in Table 5.11. It provided the information whether to reject or retain the null hypothesis that no difference existed in means of the observed variables of PLAT for the schools from Canada and Korea. All $F$ values across four variables were significant, which meant that there existed the mean differences of the four dependent variables ($PLATquestion$, $PLATlab$, $PLATmemo$ and $PLATpresentation$) among groups. Thus the null hypothesis of no mean difference was rejected and the alternative was accepted. Wilks’ lambda was used as the multivariate test that focused on the effects of the factors on the dependent variables. The main effects of country and school differences affected the means of the dependent variables. Wilks’ lambdas of country and school were .85 ($F(4,522)=23.05$, 0.001.

\(^5\) Refer to the rational for lowering down the significance level in the section, 4.7.1.2 in the chapter 4

\(^6\) Wilks’ Lambda=$\frac{|\text{SCP}_{\text{error}}|}{|\text{SCP}_{\text{effect}}+\text{SCP}_{\text{error}}|}$
An index of badness-of-fit: the lambda is close to 1, the model does not account for the variances of the sample.
were .15 (medium) and .03 (small), respectively. Follow-up tests (Univariate ANOVAs) were needed to understand the nature of the country and school effects on which variables.

Table 5-11

**MANOVA Results for PLAT**

\[N=536\]

<table>
<thead>
<tr>
<th>Effect</th>
<th>Wilks' Lambda</th>
<th>Hypothesis</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial η²</th>
<th>Observed Power²</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>0.85</td>
<td>23.05⁺</td>
<td>4</td>
<td>522</td>
<td>0</td>
<td>.15</td>
</tr>
<tr>
<td>school (country)</td>
<td>0.939</td>
<td>1.662</td>
<td>20</td>
<td>1732</td>
<td>.03</td>
<td>.012</td>
</tr>
</tbody>
</table>

⁺: the significance level at .05 (two tailed)

- Factor 1 = country, Factor 2 = school,
- Y1 = PLATquestion, Y2 = PLATlab, Y3 = PLATmemo, and Y4 = PLATpresentation

For the univariate homogeneity test of Levene’s tests were checked on the 4 variables and they did not violate the homogenous assumption. The country difference significantly affected the differences of the four variables. The F value of PLATquestion was significant \((F(1,525)=19.46, p<.001)\), partial \(\eta²=.04\) (a small effect) and the power was .93. F test results for PLATlab, PLATmemo and PLATpresentation were 74.27 \((p<.001)\), 17.99 \((p<.001)\) and 19.72 \((p<.001)\), respectively. Corresponding to these F tests, their effect sizes were \(\eta²_{PLATlab}=.12\) (a medium effect), \(\eta²_{PLATmemo}=.03\) (a small effect), and \(\eta²_{PLATpresentation}=.04\), respectively. All observed powers were greater than .95, which met the conventional standard, .80. The nested factor, school, was effective on PLATquestion; \(F(5,525)=2.56, p=.03\), partial \(\eta²=.02\) (small) and power=.79. Eta-squared means the proportion of the total variability in the dependent variable was accounted for by the variation in the independent variables.

Thus, \(\eta²_{PLATlab}=.12\) meant that the country differences accounted for 12% of the variance of
the dependent variables.

Figure 5.15 illustrates the means of each school. In accordance with the MANOVA results, generally Korean schools had lower means than Canadian schools. Particularly, Schools 2, 3 and 4 were obviously distinctive from Korean schools across the four variables while School 1 located in between.

Table 5.12 includes the detail information of the results. Based on MANOVA, the null hypotheses of no mean differences of the observed variables of PLAT were rejected and the alternative hypotheses were accepted. The follow-up tests (Univariate analyses) revealed that
the country effect was significant on the mean differences of the four observed variables while the school effect was significant on the PLATquestion whereas the effect was not significant on the other three variables.

Table 5-12

Tests of Between Subject Effects (PLAT)

\[ N=536 \]

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Powerb</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>PLATquestion</td>
<td>1</td>
<td>26.826</td>
<td>19.455</td>
<td>0</td>
<td>0.036</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>PLATlab</td>
<td>1</td>
<td>87.732</td>
<td>74.274</td>
<td>0</td>
<td>0.124</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>PLATmemo</td>
<td>1</td>
<td>18.45</td>
<td>17.987</td>
<td>0</td>
<td>0.033</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>PLATpresentation</td>
<td>1</td>
<td>27.961</td>
<td>19.717</td>
<td>0</td>
<td>0.036</td>
<td>0.993</td>
</tr>
<tr>
<td>School (country)</td>
<td>PLATquestion</td>
<td>5</td>
<td>3.522</td>
<td>2.555</td>
<td>0.027</td>
<td>0.024</td>
<td>0.793</td>
</tr>
<tr>
<td></td>
<td>PLATlab</td>
<td>5</td>
<td>2.382</td>
<td>2.017</td>
<td>0.075</td>
<td>0.019</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>PLATmemo</td>
<td>5</td>
<td>1.932</td>
<td>1.883</td>
<td>0.096</td>
<td>0.018</td>
<td>0.641</td>
</tr>
<tr>
<td></td>
<td>PLATpresentation</td>
<td>5</td>
<td>1.33</td>
<td>0.938</td>
<td>0.456</td>
<td>0.009</td>
<td>0.337</td>
</tr>
<tr>
<td>Error</td>
<td>PLATquestion</td>
<td>525</td>
<td>1.379</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLATlab</td>
<td>525</td>
<td>1.181</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLATmemo</td>
<td>525</td>
<td>1.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLATpresentation</td>
<td>525</td>
<td>1.418</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.6.2 MANOVA: Differences between Canadian and Korean Students’ Perceptions on Science Assessment

This analysis provided answers to the research question “Are there any differences of means between Canadian and Korean students’ perceptions on their science assessment?” The mean differences were examined in item levels across countries and schools (nested) using MANOVA. Country and school were fix variables and the 3 variables of PAFORMAT and 3 variables of PACONT were the dependent variables. The null and alternative hypotheses for
country difference were set up as follow:

\[
H_0: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} \text{ and } H_1 = \not H_0 \text{ for } PAFORMAT
\]

\[
H_0: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} \text{ and } H_1 = \not H_0 \text{ for } PACONT
\]

The null hypothesis is that the means for the set of 3 variables do not differ across any of the population that correspond to the groups taking intercorrelations among the outcome variables into account. Here, \( \mu_{1i} \) means the population mean of group 1 (Canada) and \( i^{th} \) variable and \( \mu_{2i} \) does the population mean of group 2 (Korea) with \( i^{th} \) variable.

The null and alternative hypotheses for school difference were:

\[
H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{61} \\ \mu_{71} \\ \mu_{76} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{16} \\ \mu_{26} \\ \vdots \\ \mu_{66} \\ \mu_{76} \end{bmatrix} \text{ and } H_1 = \not H_0
\]

These null hypotheses for \( PAFORMAT \) and \( PACONT \) were tested separately: firstly, the three observed variables of \( PAFORMAT \) and secondly the three observed variables of \( PACONT \) were performed.

Firstly, MANOVA was performed on \( PAFORMAT \). One of the assumptions of MANOVA is homogeneity of covariances. Box’s \( M \) tested the null hypothesis that (the
observed covariance matrices of the dependent variables are equal across schools) and the result showed significant; $M=88.85$, $F(36, 262033)=2.41$ and $p<.001$. As Box’s $M$ was significant, groups differed in their covariance matrices of all dependent observed variables of PAFORMAT. Although the homogeneous assumption of equal covariance was violated, the $F$ test is quite robust about the violation. For the Multivariate test, Pillai’s Trace was used to test the null hypothesis of MANOVA, which is more conservative than Wilks’s Lambda. If Pillai’s Trace were significant, follow-up tests would be performed for identifying which factors (country or school) are effectively differentiated the groups and which groups are most clearly differentiated.

The MANOVA results showed that Pillai’s Traces for country and school (country) were .04 ($F(3,525)=7.99$, $p<.001$) and .15 ($F(15, 1581)=5.63$, $p<.001$), respectively. The corresponding eta squared values were .04 (small) and .05 (small), respectively. Both country and school (nested within country) affect significantly the variances of the dependent variables PAFORMATclose, PAFORMATshort and PAFORMATalternative. Unlike Wilks’s Lambda, Pillai’s Trace is the index of goodness-of-fit. The larger Pillai’s Trace is the more significant effect that the factors affect to the dependent variables. In this MANOVA, it was significant for the both factors; therefore, the null hypothesis on country effect on assessment was rejected and the alternative hypothesis was accepted.

The results of MANOVA revealed that significant difference in the main factors means that the residual covariance among multiple dependent variables include at least one significant difference among the variables and hence follow-up tests to assess the nature of significant main effects. Levene’s tests of equality of error variances showed that PAFORMATclose and PAFORMATshort were not significant while PAFORMATalternative
was significant. Due to the violation of homogeneity assumption in PAFORMATalternative, the power may be decreased on those variables. However, $F$ test is quite robust on the violation of the equality assumption, especially the cases per school is sufficiently large (over 15), the analysis yields reasonably accurate $p$ values. Thus, further univariate analysis was performed.

The country difference affected significantly the mean difference of PAFORMATclose and PAFORMATalternative differed significantly; the $F$ values for the two variables were 19.33 ($p<.001$) and 13.36 ($p<.001$), respectively. However, the country effect did not significantly affect PAFORMATshort $F(1,527)=2.50, p=.12$. In addition, the nested factor, school, affected significantly all three variables. The results indicated that Canadian students were likely to have the perceptions that their science test formats were more flexible and close-ended questions were less important than what the Korean students perceived. Figure 5.16 (a~c) and Table 5.13 includes detail information on the analysis result.
Table 5-13

Tests of Between Subject Effects on PAFORMAT

\(N=536\)

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>df</th>
<th>Mean Square</th>
<th>(F)</th>
<th>Sig.</th>
<th>Partial (\eta^2)</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>PAFORMATclose</td>
<td>1</td>
<td>17.95</td>
<td>19.33</td>
<td>0</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>PAFORMATshort</td>
<td>1</td>
<td>2.85</td>
<td>2.50</td>
<td>0.12</td>
<td>0.01</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>PAFORMATalternative</td>
<td>1</td>
<td>18.06</td>
<td>13.36</td>
<td>0</td>
<td>0.03</td>
<td>0.95</td>
</tr>
<tr>
<td>School (country)</td>
<td>PAFORMATclose</td>
<td>5</td>
<td>9.38</td>
<td>10.10</td>
<td>0</td>
<td>0.09</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>PAFORMATshort</td>
<td>5</td>
<td>3.47</td>
<td>3.0</td>
<td>0.01</td>
<td>0.03</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>PAFORMATalternative</td>
<td>5</td>
<td>14.86</td>
<td>10.99</td>
<td>0</td>
<td>0.09</td>
<td>1</td>
</tr>
<tr>
<td>Error</td>
<td>PAFORMATclose</td>
<td>527</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAFORMATshort</td>
<td>527</td>
<td>1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAFORMATalternative</td>
<td>527</td>
<td>1.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>PAFORMATclose</td>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAFORMATshort</td>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAFORMATalternative</td>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Secondly, MANOVA was performed for PAFORMAT. Since the homogeneity assumption was violated, Pillai’s Trace was used for multivariate tests. The values for country and school effects were significant; Pillai’s Trace=.24, \(F(3, 527)=56.70, p<.001\) and Pillai’s Trace=.11, \(F(15,1587)=4.03, p<.001\), respectively. The corresponding effect sizes were \(\eta^2=.24\) (medium).
for country and $\eta^2=.04$ (small) for school; that is, 24% and 4% of variances were predictable from the difference of country and school.

Among three observed variables, the country effects were significant for $PACONT_{knowledge}$ ($F(1,529)=67.97, p<.001$) and $PACONT_{lab}$ ($F(1,529)=84.56, p<.001$). As Figure 5.17 presents the means of each school, the Korean students ($M_{Ko}=2.13$) perceived that their test content was more on knowledge emphasized content than the Canadian students ($M_{Ca}=2.99$) did. However, the country effects on $PACONT_{application}$ was not significant ($p=.22$). The school effects on the three variables were statistically significant at alpha level .05. This result meant that individual schools had different emphasis in their test content.

Figure 5.17 (b) shows the Canadian school 2 had the lowest mean while School 3 had the highest. In other words, for the observed variable of $PACONT_{application}$ school differences significantly affected students’ perceptions on their assessment rather than the country effect. Due to these two-school differences, the country effect may not be significant.

MANOVAs were performed with the schools and countries as factors (schools were nested within the country) and 6 assessment variables as dependent variables. Both country and school differences were revealed as the significant factors across almost all the dependent
variables. The country effect was not significant of the mean differences on \textit{PAFORMATshort} and \textit{PACONTapplication}, otherwise the country difference was predictive of the overall variances ranged from 4\% to 24\%.

As the plots show (Figure 5.17 (a–f), Schools 3 and 4 had generally higher means than other 5 schools across all 6 variables. Regarding assessment formats, students from these lower mean schools were more likely to perceive that the traditional paper-and-pencil based test formats (close-ended and short-answering test formats) were more important than the alternative test formats (science lab reports, portfolios, projects, performances tasks). In other words, Lock (1990) says that the close-ended and short answering test formats do not allow alternative answers, and teachers expect their students to answer “correct answer.” On the other hand, teachers expect diverse answers from the alternative formats. Also, the low means in the variables of assessment content (\textit{PACONTknowledge}, \textit{PACONTapplication} and \textit{PACONTlab}) imply that students weighted scientific laws and theories (knowledge of science) as the important test content.

The answer to the research question, “Are there any differences of means between Canadian and Korean students’ perceptions on their science assessment?” was that there existed both country and school effects on the variance differences of the observed variables of PAFORMAT and PACONT. Thus, the null hypotheses of no difference were rejected and the alternative hypotheses were accepted.

\textit{5.6.3 MANOVA: Differences in Students’ NOS Understanding}

MANOVAs were performed over the five NOS constructs with the factors of country and school (nested within country). The analyses will answer the research question, “Are there any differences of means between Canadian and Korean students’ NOS understanding?” and
“Are there any differences of means among different school students’ NOS understanding?”

The comparisons will go construct-by-construct. The null hypotheses are that the means for the observed dependent variables (NOS items) do not differ across any of the population that correspond to the groups taking intercorrelations among the dependent variables into account. An example of the null and alternative hypotheses for Tentativeness of NOS for country effect were set up as follow:

$$H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

$$H_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \neq \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

Here, $\mu_{11}$ means the population mean of group 1 (Canada) and the first variable ($TEN1$) and $\mu_{21}$ does the population mean of group 2 (Korea) with $TEN1$.

The analysis procedures would be similar to the previous PLAT, PAFORMAT and PACONT analyses: 1) the MANOVA assumption test (Box’s M), 2) Multivariate tests, and 3) univariate tests.

**Tentativeness**: Box’s $M$ was 38.44 and the value was not significant: $F=1.05, p=.39$. Thus, Wilks’s Lambda was used to test the multivariate analysis. The effect of country was not significant; Wilks’s Lambda=.99, $F(3,524)=2.51, p=.06$. The school difference significantly affected students’ understanding of Tentativeness; Wilks’s Lambda=.95, $F(15, 1447)=1.70, p=.04, \eta^2=.016$ (a small effect size) and power=.90. These indices implied that at least one school was different at least one out of the three observed dependent variable ($TEN1$, $TEN2$, and $TEN3$).
The univariate analysis results showed which school and on which observed variables were different. As a preliminary ANOVA assumption test, Levene’s test of equality of error variance was checked. TEN1 and TEN2 were not significant while TEN3 was significant. As known, ANOVA is robust in the violation of the homogeneity assumption; the examination on the difference was continued. The $F$ values showed that the school effect was significant in TEN1, but TEN2 and TEN3 did not differ among schools; $F_{TEN1}(5,526)=3.03$ and $p=.02$ and the effect size of partial $\eta^2=.03$, which was small effect. That is, the school difference accounted for about 3% of the variance in TEN1. The observed power was .82.

Figure 5.18 present the detail information on MANOVA analysis results on Tentativeness. The means varied school by school rather than country. Generally, School 4 had the highest means of TEN1 and TEN3 while School 5 had small means.

Subjectivity: The effects of school and country differences on Subjectivity were examined. Both country and school effects on the variances of dependent variables were significant: Wilks’s Lambdas were .88 at $F(3, 524)=24.68$, $p<.001$ for the country and .91 at $F(15,1447)=3.55$, $p<.001$ for the school effect. The results corresponded to the partial $\eta^2=.12$ (a medium effect size) and $\eta^2=.03$ (small). Country and school effects accounted for 12% and 3%
of the overall variance of Subjectivity, respectively. The multivariate analysis results indicated that at least one variable of Subjectivity would be different in 2 countries and from 7 schools.

Table 5.14 includes detail information about the country and school effects on the three dependent variables. The variances of the three variables of Subjectivity were significantly affected by the country and school effects. Relatively the country difference on the variance differences of the dependent variables affected more than the school difference. For instance, the country and school effects accounted for 8% and 3% of the variance of SUB1.

Table 5-14
Tests of Between Subjects Effects on Subjectivity

<table>
<thead>
<tr>
<th>Source</th>
<th>DV</th>
<th>$df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>SUB1</td>
<td>1</td>
<td>85.974</td>
<td>48.52</td>
<td>0.00</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SUB2</td>
<td>1</td>
<td>36.27</td>
<td>21.87</td>
<td>0.00</td>
<td>0.04</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>SUB3</td>
<td>1</td>
<td>80.416</td>
<td>48.23</td>
<td>0.00</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>school (country)</td>
<td>SUB1</td>
<td>5</td>
<td>5.914</td>
<td>3.34</td>
<td>0.01</td>
<td>0.03</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>SUB2</td>
<td>5</td>
<td>5.715</td>
<td>3.45</td>
<td>0.01</td>
<td>0.03</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>SUB3</td>
<td>5</td>
<td>15.003</td>
<td>9.00</td>
<td>0.00</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Error</td>
<td>SUB1</td>
<td>526</td>
<td>1.772</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SUB2</td>
<td>526</td>
<td>1.659</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SUB3</td>
<td>526</td>
<td>1.667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Figure 5.19 shows, the Korean schools had lower means ($M_{SUB1}=2.65$, $M_{SUB2}=2.56$, $M_{SUB3}=2.47$) than the Canadian schools ($M_{SUB1}=3.48$, $M_{SUB2}=3.10$, $M_{SUB3}=3.27$) in all the three variables. As the dependent variables attempted to measure students’ views on theory-laden observations in scientific research, Korean students were likely to view that scientific observations were less biased by individual scientists’ opinions and background knowledge than Canadian students were. Particularly, Schools 4 had the highest means ($M_{SUB1}=3.98$,
\( M_{\text{SUB2}} = 3.53, M_{\text{SUB3}} = 3.96 \) in all the three variables while School 6 (\( M_{\text{SUB1}} = 2.54, M_{\text{SUB2}} = 2.58, M_{\text{SUB3}} = 2.18 \)) had the lowest means; that is, the majority of the students from School 4 accepted that scientific observations were not objective whereas the majority of the students from School 6 regarded scientific observations as factual knowledge.

![Estimated Marginal Means](image)

(a) SUB1 (b) SUB2 (c) SUB3

**Figure 5-19. Means of Schools in Subjectivity**

**Empirical Evidence Based Scientific Knowledge:** Like the previous constructs of NOS, country and school differences affected the differences of variances of three observed dependent variables (\( EMP1, EMP2 \) and \( EMP3 \)); Wilks’ Lambdas for country and school were .973 and .93 and their \( F \) statistics were \( F(3,524) = 4.82, p = .003, \) partial \( \eta^2 = .03 \) and \( F(15,1447) = 2.50, p = .001, \) partial \( \eta^2 = .02 \) (small), respectively. Both factors had the observed power = 1.00.

The follow-up univariate tests showed the country difference significantly affected the variance differences of \( EMP3 \) \( (F(1,526) = 14.00, p < .001, \) partial \( \eta^2 = .03 \) while it did not affect the variances of \( EMP1 \) and \( EMP3 \). On the other hand, the school differences were significant

---

7 Estimated marginal means
in $EMP2$ ($F(5,526)=7.77, p=.001$ and partial $\eta^2=.04$). School differences accounted for 4% of the variances of $EMP2$. Table 5.15 includes the test results of between subject effects.

Table 5-15

*Tests of Between Subjects Effects on Empirical Evidence Based Science*

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Observed Power$^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>$EMP1$</td>
<td>0.73</td>
<td>0.43</td>
<td>0.51</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>$EMP2$</td>
<td>3.09</td>
<td>1.59</td>
<td>0.21</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>$EMP3$</td>
<td>25.22</td>
<td>14.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.96</td>
</tr>
<tr>
<td>school (country)</td>
<td>$EMP1$</td>
<td>3.35</td>
<td>1.99</td>
<td>0.08</td>
<td>0.02</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>$EMP2$</td>
<td>7.77</td>
<td>4.01</td>
<td>0.00</td>
<td>0.04</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>$EMP3$</td>
<td>3.74</td>
<td>2.08</td>
<td>0.07</td>
<td>0.02</td>
<td>0.69</td>
</tr>
<tr>
<td>Error</td>
<td>$EMP1$</td>
<td>526</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$EMP2$</td>
<td>526</td>
<td>1.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$EMP3$</td>
<td>526</td>
<td>1.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{a}$ Computed using alpha = .05

Figure 5.20 showed that no pattern could be found within countries. $EMP2$ fluctuated school-by-school. School 4 had high means among Canadian schools and School 7 had high means in comparison with other Korean schools.

![Figure 5-20. Means of Schools in Empirical Evidence Based Science](image)
**Sociocultural Embeddedness of Scientific Knowledge**: MANOVA was performed to examine the effects of country and school differences on understanding of the concepts of culturally embedded scientific knowledge (*Sociocultural Embeddedness*). First, the homogeneous assumption was checked using Box’s *M*. It was not significant; *M* = 48.01, *F*(36, 268062) = 1.31, *p* = .10). MANOVA results showed the country and school effects were not significant. Wilks’s Lambda for country was .99, which meant that the Sum of Cross Product of country (*SCP*<sub>country</sub>) was about 0 while the Sum of Cross Product of error (*SCP*<sub>error</sub>) was about 1. *SCP*<sub>school</sub> was not zero but it was not large enough to make Wilks’ lambda small; Wilk’s Lambda was .96, *F*(15, 1441) = 1.42, *p* = .131. The null hypothesis of no effect of country and school on students’ understanding of NOS was retained and the alternative hypothesis was rejected.

**Diverse Methods of Scientific Research**: The homogenous tests of Box’s *M* were significant; *M* = 75.70, *F*(36, 267989) = 2.06, *p* < .001; thus, a conservative multivariate test would be good to avoid biased results. For the country factor, Pillai’s Trace was .05, *F*(3, 524) = 9.90, *p* < .001, partial η² = .05 and for the school factor, it was .10, *F*(15, 1578) = 3.57, *p* < .001, partial η² = .03.

The country factor significantly affected on the variance of *METH2*; *F*(1, 526) = 25.83, *p* < .001 while *METH1* and *METH3* were not. On the other hand, there existed school differences across all three observed variables. For *METH1*, *METH2* and *METH3* the *F* tests were 4.23 (*p* = .001), 4.14 (*p* = .001) and 4.98 (*p* < .001), respectively. Corresponding effect sizes were .04, .04 and .05, respectively. In other words, school differences accounted for 4% of variances of *METH1* and *METH2*, and 5% of variances of *METH3*.

Figure 5.21 shows the means of each school. Consisted with the MANOVA test, three means of Korean schools in *METH2* had lower than the Canadian schools. Overall School 4 had the highest means and the students from that school were likely to admit diverse research
scientific research methods. Therefore, the null hypotheses on country and school differences in means of diverse research methods of science were rejected and the alternative hypotheses were accepted.

Figure 5-21. Means of Schools in Diverse Scientific Methods

Summary of Differences between Countries and among Schools: MANOVAs examined the effects of country and school on the variance of observed variables of the NOS concepts and the effects varied depending on the construct. For Subjectivity, all three observed variables were significantly affected by both country and school differences while Sociocultural Embeddedness was not. Although the country and school difference affected the variance differences of the variables, the effect sizes were mostly small. The variance of SUB2 was highly predictive by country and school differences while others ranged from 1% to 8%. There existed differences, but not as much as the perceptions of learning activities/tasks or assessment.

The following section concerns associations, and addresses how students’ perceptions of their learning activities and tasks and their perceptions of their science assessments accounted for the variances in their understandings of NOS.
5.7 Associations

5.7.1 PLS Analysis Overview

Previous sections in this chapter have analyzed participants’ responses of the three domains: 1) perceptions of learning activities and tasks during science classes (PLAT); 2) perceptions of science assessment formats (PAFORMAT) and content (PACONT) and their understanding of NOS and the differences of countries and schools in these three domains. The present section focuses on the associations among these variables.

One aim of the research was to identify potential factors involved in science education (learning activities and assessment), which influence students’ understanding of the concepts of NOS. To ascertain what factors affect students’ understanding of NOS, several research assumptions were postulated based on the previous studies on students’ perceptions on their learning environment and attitudes (Dorman et al., 2005), perceptions on assessment formats and study strategies (Biggs, 1987, 2003; Brown & Hirschfeld, 2008; Sambell et al., 1997; Scouller, 1998) and NOS studies (Duschl, 1990; Lederman, 2007; Roth & Roychoudhury, 1994). They were 1) how students learn science (major learning activities and tasks during science classes) affects students’ understanding of NOS; 2) how their learning is assessed (science test formats) affects their understanding of NOS and 3) what kind of science content is assessed affects their understanding of NOS. Since many observed variables (indicators) and latent variables corresponding the observed variables were involved, both independent and dependent variables, Partial Least Squares (PLS) (software, SmartPLS, developed by Ringle, Wende and Will (2005)) were used to analyze the data.

Since the measurement models (the formative and the reflective models) were examined in the validity and reliability secection (5.4), the following section devotes to
examining the structure models. The PLS analyses were made in the following order: 1) testing the hypothesized model, 2) modifying the model by removing constructs of which inner coefficients ($\gamma$s) were not significant at alpha level .05, and 3) examining overall accountability of the independent latent variables (exogenous variables: here, Learning Activities and Task ($PLAT$) and Assessment Format ($PAFORMAT$) and Assessment Content ($PACONT$) for the dependent latent variables (endogenous variables: here, understanding of the 5 NOS concepts).

The hypothesized model was tested in 5 NOS concepts. For parsimony (minimum number of variables accounts for the maximum variances of NOS concepts), the models were modified based on the inner coefficient’s ($\gamma$) significance.

With respect to mediator effects, this research assumed that assessment formats and content decide learning activities and tasks. Here, the mediator effects mean that $PAFORMAT$ and $PACONT$ affect directly affect the understanding of NOS, and they affect it indirectly via $PLAT$. Of course, students’ learning activities and assessment formats and content do not move in one direction but involve mutual interaction. Learning activities can affect either/both assessment formats or/and assessment content, and vice versa. In an analysis, however, both directions cannot be examined at the same time since technically the specification problems do not allow such models. Therefore, it is necessary to select one direction over the others. Either the perceptions of learning activities and tasks affect the perceptions on the assessment or vice versa. Previous literature (Biggs, 1987, 2003; Scouller, 1998) has revealed that assessment formats and content are decisive to how students approach their studies and prepare for tests. Ideally, an assessment should be congruent with what students learned and how they learned in a subject, and take a direct path from activity to assessment. However, from the perspective of
the students, these paths did not work ideally. This research focused on the students’ perspectives, how they perceived their learning activities and assessment. This calls for the study of variables to move from assessment formats and content to learning activities.

### 5.7.2 Canadian Sample

Not all inner coefficients (starting from the independent latent variables, PLAT, PAFORMAT and PACONT to the dependent latent variables (NOS concepts)) were significant. A nonparametric bootstrapping was performed to identify which inner coefficients were significant and which were not.

Insignificant coefficients were excluded in the modified model, so Table 5.16 reports those effective coefficients, their t values and significances. The three inner coefficients were not significant at alpha level 0.05; the inner coefficients from PLAT to Subjectivity ($\gamma_{PLAT Subjectivity} = .003, (t_{216})=0.06, p=.95$), from PAFORMAT to Sociocultural Embeddedness ($\gamma_{PAFORMAT Culture} = .043, (t_{216})=0.87, p=.385$), and PACONT to EMP ($\gamma_{PACONT EMP} = -0.16, (t_{216})=1.78, p=.08$). The mediator effect from PACONT to PLAT ($\gamma_{PACONT PLAT} = -.038, (t_{216})=0.50, p=.618$) was not significant. Figure 5.22 depicts the theoretical and the modified models. The results of PLS analyses for Canadian were in Table 5.18. It includes the measurement models’ outer weights, t-values and p-values, $R^2$, the effect sizes ($f^2$) and the predictive relevance ($Q^2$) and their effect sizes ($q^2$).

\[
\begin{align*}
\text{t}_{\text{emp}} &= \frac{W}{se(W)} \\
\text{As the equation shows, t-values were decided by the coefficients and their standard errors in bootstrapping.}
\end{align*}
\]
Diversity of scientific research methods (METHOD) had the highest $R^2$ while the Empirical evidence based scientific knowledge (EMP) had the smallest $R^2$. The highest effect size ($f^2$) was 0.65 (the path from PLAT to METHOD) and the lowest was 0.01 (the path from PAFORMAT to EMP). The consequence of small $R^2$ of EMP, the effect sizes were also small and the predictive relevance ($Q^2$) was zero, but not a negative sign. The values of $q^2$ (the index of the impacts of structure models to the indicators of the endogenous variables) of PLAT to METH, and of PACONT to CULTURE were high. Since the values of $q^2$ were greater than zero, the observed variables were well reconstructed and relevant by the model. The reports of each understanding of the NOS concepts were followed in the next section.

Table 5-16

Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Canadian)

<table>
<thead>
<tr>
<th>Paths</th>
<th>$\gamma$</th>
<th>$t$-value</th>
<th>$p$-Value</th>
<th>$R^2$</th>
<th>$f^2$</th>
<th>$Q^2$</th>
<th>$q^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAT TENTATIVENESS</td>
<td>0.29</td>
<td>4.88</td>
<td>$p &lt; .001$</td>
<td>0.417</td>
<td>0.02</td>
<td>0.21</td>
<td>0.01</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>0.13</td>
<td>2.21</td>
<td>$p = .028$</td>
<td>0.13</td>
<td>0.01</td>
<td>0.26</td>
<td>0.06</td>
</tr>
<tr>
<td>PACONT</td>
<td>0.5</td>
<td>10.25</td>
<td>$p &lt; .001$</td>
<td>0.36</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAT SUBJECTIVITY</td>
<td>0.32</td>
<td>6.21</td>
<td>$p &lt; .001$</td>
<td>0.445</td>
<td>0.13</td>
<td>0.26</td>
<td>0.06</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>0.48</td>
<td>8.78</td>
<td>$p &lt; .001$</td>
<td>0.30</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACONT</td>
<td>-0.162</td>
<td>1.78</td>
<td>$p = .076$</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAT EMP</td>
<td>0.26</td>
<td>2.15</td>
<td>$p = .033$</td>
<td>0.192</td>
<td>0.08</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>0.33</td>
<td>4.32</td>
<td>$p &lt; .001$</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PACONT</td>
<td>-0.162</td>
<td>1.78</td>
<td>$p = .076$</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAT CULTURE</td>
<td>0.18</td>
<td>2.70</td>
<td>$p = .007$</td>
<td>0.409</td>
<td>0.05</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>0.61</td>
<td>12.47</td>
<td>$p &lt; .001$</td>
<td>0.56</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACONT</td>
<td>0.18</td>
<td>3.57</td>
<td>$p &lt; .001$</td>
<td>0.07</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$N_i=217$
5.7.2.1 Tentativeness

The inner coefficients to Tentativeness were $\gamma_{PLAT}=0.29$ ($t(216)=4.88, p<.001$), $\gamma_{PAFORMAT}=0.13$, ($t(216)=2.21, p=.007$) and $\gamma_{PACONT}=0.50$ ($t(216)=10.25, p<.001$). These three
independent latent variables (exogenous variables), PLAT, PAFORMAT and PACONT, accounted for 41.7% of the variance of Tentativeness. Additionally, PACONT also indirectly affected Tentativeness through PAFORMAT. The effect sizes of PLAT, PAFORMAT and PACONT were 0.13 (medium), 0.02 (small) and 0.36 (a large effect), respectively.

Before introducing the predictive equations for the observed variables of the endogenous variables (the relationships between the independent latent variables (exogenous variables) and the dependent latent variable (endogenous) (Tentativeness), the notations of the equation were in Table 5.17; x represents one of the exogenous variables’ observed variables (indicator) and y represents one of the endogenous variables’ observed variables (indicator). For instance, \( x_i \) is for the exogenous variables’ observed variables (ith variable, in the group j), and \( y_{ij} \) were for the endogenous variables’ observed variables.

Table 5-17

<table>
<thead>
<tr>
<th>notation</th>
<th>variable</th>
<th>notation</th>
<th>variable</th>
<th>notation</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_{11} )</td>
<td>PLATlab</td>
<td>( X_{32} )</td>
<td>PAFORMATclose</td>
<td>( X_{83} )</td>
<td>PACONTapplication</td>
</tr>
<tr>
<td>( X_{21} )</td>
<td>PLATmemo</td>
<td>( X_{62} )</td>
<td>PAFORMATshort</td>
<td>( X_{93} )</td>
<td>PACONTlab</td>
</tr>
<tr>
<td>( X_{31} )</td>
<td>PLATpresent</td>
<td>( X_{72} )</td>
<td>PAFORMATalter</td>
<td>( X_{103} )</td>
<td>PACONTknowledge</td>
</tr>
<tr>
<td>( X_{41} )</td>
<td>PLATquestion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{11} )</td>
<td>TEN1</td>
<td>( Y_{73} )</td>
<td>EMP1</td>
<td>( Y_{135} )</td>
<td>METH1</td>
</tr>
<tr>
<td>( Y_{21} )</td>
<td>TEN2</td>
<td>( Y_{83} )</td>
<td>EMP2</td>
<td>( Y_{145} )</td>
<td>METH2</td>
</tr>
<tr>
<td>( Y_{31} )</td>
<td>TEN3</td>
<td>( Y_{93} )</td>
<td>EMP3</td>
<td>( Y_{155} )</td>
<td>METH3</td>
</tr>
<tr>
<td>( Y_{42} )</td>
<td>SUB1</td>
<td>( Y_{104} )</td>
<td>CUL1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{52} )</td>
<td>SUB2</td>
<td>( Y_{114} )</td>
<td>CUL2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{62} )</td>
<td>SUB3</td>
<td>( Y_{124} )</td>
<td>CUL3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[^9\text{Effect size } f^2 = \frac{R^2_{\text{included}} - R^2_{\text{excluded}}}{1 - R^2_{\text{included}}}\]

According to Chin (1998) values of 0.02, 0.15, and 0.35 viewed as a standard for whether a exogenous variable has a weak, medium or large effect at the structure level.
Not considering the disturbance ($\zeta$s), the prediction of $TEN1$, $TEN2$ and $TEN3$ would take the following steps:

The calculation of the independent Latent variables: the equations from (1) to (3),

The calculation of the dependent Latent variables: the equation (4)

The predictions of $TEN1$, $TEN2$ and $TEN3$: the equations from (5) to (7)

If a student responded to the survey like Figure 5.23, the variables of Tentativeness could be predicted using the equations:

\[
\xi_{PLAT} = \lambda_{x11} x_{11} + \lambda_{x21} x_{21} + \lambda_{x31} x_{31} + \lambda_{x41} x_{41} + \zeta_1 \quad \cdots \quad (1)
\]
\[
\xi_{PAFORMAT} = \lambda_{x12} x_{12} + \lambda_{x62} x_{62} + \lambda_{x72} x_{72} + \zeta_2 \quad \cdots \quad (2)
\]
\[
\xi_{PACONT} = \lambda_{x83} x_{83} + \lambda_{x93} x_{93} + \lambda_{x103} x_{103} + \zeta_3 \quad \cdots \quad (3)
\]
\[
\eta_i = \xi_1 \gamma_{i1} + \xi_2 \gamma_{i2} + \xi_3 \gamma_{i3} + \zeta_i \quad \cdots \quad (4)
\]
\[
y_{11} = \eta_1 \lambda_{y11} + \epsilon_1 \quad \cdots \quad (5)
\]
\[
y_{21} = \eta_2 \lambda_{y21} + \epsilon_2 \quad \cdots \quad (6)
\]
\[
y_{31} = \eta_3 \lambda_{y31} + \epsilon_3 \quad \cdots \quad (7)
\]

Figure 5-23. An Example of Student's Response

<table>
<thead>
<tr>
<th>PLAT problem</th>
<th>PLAT lab</th>
<th>PLAT memo</th>
<th>PLAT present</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PAFOEMAT close</th>
<th>PAFOEMAT short</th>
<th>PAFOEMAT alternative</th>
<th>PACONT knowledge</th>
<th>PACONT application</th>
<th>PACONT lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

\[
\eta_{Tentativeness} = \gamma_{PLAT} \xi_{PLAT} + \gamma_{PAFORMAT} \xi_{PAFORMAT} + \gamma_{PACONT} \xi_{PACONT} + \zeta_4
\]
\[
\eta_{Tentativeness} = 0.29(2.83 + \zeta_1) + 0.13(2.67 + \zeta_2) + 0.50(2.54 + \zeta_3) + \zeta_4
\]
\[ \eta_{\text{Tentativeness}} = 2.44 + \zeta_5 \]

\[ TEN1 = \eta_{\text{Tentativeness}} \times \lambda_{\text{TEN1}} + \epsilon_1;\]

\[ TEN2 = \eta_{\text{Tentativeness}} \times \lambda_{\text{TEN1}} + \epsilon_2;\]

\[ TEN3 = \eta_{\text{Tentativeness}} \times \lambda_{\text{TEN1}} + \epsilon_3 \]

\[ TEN1 = 2.44 \times 0.68 + \epsilon_1 = 1.66 + \epsilon_1 \quad \ldots \ldots \ldots \text{(a)};\]

\[ TEN2 = 2.44 \times 0.71 + \epsilon_2 = 1.73 + \epsilon_2 \quad \ldots \ldots \ldots \text{(b)}; \text{ and} \]

\[ TEN3 = 2.44 \times 0.81 + \epsilon_3 = 1.98 + \epsilon_3 \quad \ldots \ldots \ldots \text{(c)} \]

Here, all \( \zeta \)s are disturbance, which were not measured through the observed variables. Nor are they dependent from other disturbances. The \( \epsilon \)s were the error variances of the observed variables of the endogenous variables. The equations, (a), (b) and (c) provided a prediction of how the student view on Tentativeness.

The predictive relevance \( (Q^2) \) was calculated based on the blindfolding procedures of SmartPLS and it was 0.21, which is greater than the cutoff value, 0. The observed variables of Tentativeness (i.e., TEN1, TEN2 and TEN3) were well reconstructed by the structure model. The impacts \( (q^2) \) of PLAT, PAFORMAT and PACONT were 0.05, 0.01 and 0.14. That is, students’ perceptions on their assessment content importantly impact their understanding of tentative nature of science.

Related to Tentativeness, all inner coefficients were positive across three samples. That is, if independent variables’ values increase, the dependent variables also increase. All indicators were arranged to the order that the perceptions of teacher directed learning activities and tasks, knowledge of science emphasized assessment and traditional paper-and-pencil based test focused scored low. Also, the concepts of NOS also rearranged so that realistic views of nature of science had low scores. Therefore, the positive signs of all outer loadings and inner coefficients meant that if a student perceive his/her learning activities as teacher-
directed, knowledge of science emphasized assessment, and science tests required them only one correct answers, his/her understanding of tentative nature of science would likely be a realistic view.

Regarding the research question that whether learning activities and tasks and assessment of science affect students’ understanding of tentative nature of science, the answer would be that how students perceive their learning activities/tasks and assessment could affect significantly their understandings of tentative nature of science.

5.7.2.2 Subjectivity (Theory-laden observations)

The hypothesized model with Canadian data accounted for 44.5% of the variance of Subjectivity. The effect sizes were 0.13 (small) and 0.30 (medium) for PAFORMAT and PACONT, respectively. Since the inner coefficients ($\gamma$s), PLAT was not statistically significant to Subjectivity ($\gamma_{PLAT \rightarrow Subjectivity} = .003, (t(216)=0.06, p= .95)$); thus, PLAT was excluded in the modified model. The inner coefficients of PAFORMAT and PACONT were $\gamma_{PAFORMAT} = 0.32$ ($t(216)=6.21, p<.001$) and $\gamma_{PACONT} = 0.48$ ($t=8.78, p<0.001$). The predictive relevance ($Q^2$) was 0.26 and corresponding effect sizes ($q^2$) for PAFORMAT and PACONT were 0.06 and 0.15, respectively. As the predictive relevance is positive, the observed variables of Subjectivity were well reconstructed by the structural model. The equations below (d, e and f) are the reconstructed based on the model.

$$\eta_{Subjectivity} = 0.32 \xi_{PAFORMAT} + 0.48 \xi_{PACONT} + \xi_5$$

$$SUB1 = 0.79 \eta_{Subjectivity} + \varepsilon_4 \cdots \cdots (d)$$

$$SUB2 = 0.77 \eta_{Subjectivity} + \varepsilon_5 \cdots \cdots (e)$$

$$SUB3 = 0.80 \eta_{Subjectivity} + \varepsilon_6 \cdots \cdots (f)$$
All relevant coefficients had positive signs, which meant that students who perceived their assessment content weighed more on knowledge of science were likely to have epistemologically logical positivistic views on science. On the contrary, students who thought their tests required applying the knowledge into diverse situations or allow different views in their tests (alternative assessment methods) were likely to have the idea that scientists’ prior knowledge can affect their observations and observations are not objective from the observers.

Looking into the effect sizes of the two independent latent variables, students’ perceptions on their science assessment content were an effective predictor on understanding of Subjectivity.

5.7.2.3 Empirical Evidence Based Scientific Knowledge

The three exogenous variables accounted for 19.2% of the variance of Empirical Evidence Based Scientific Knowledge (EMP). The effect sizes were weak since the total accounted variance was small (19.2%); for PLAT it was $f^2=0.08$, for PAFORMAT, $f^2=0.01$ and for PACONT, $f^2=0.08$. Two inner coefficients from PLAT and PAFORMAT were positive: \( \gamma_{PLAT} = 0.26, t(216)=2.15, p=.033 \) and \( \gamma_{PAFORMAT} = 0.33, t=4.32, p<.001 \). However, the path from PACONT to EMP had a negative sign: \( \gamma_{PACONT} = -0.16, t=1.78, p=.076 \) and the significance level was over .05. Although the significance of a path is not significant, the path was retained since the sign of coefficient was negative and conventionally accepted in an alpha level .10 (Karim, 2009). It was unique across other constructs and other samples, Korean and combined.

Concerning the prediction of students’ understanding of EMP, the Stone-Gessier’s criterion, \( Q^2 \), was zero, but not negative sign. The model had the ability to predict the observed variables (EMP1, EMP2 and EMP3) for the Canadian sample. The equations were derived from the modified model:

\[
\eta_{EMP} = 0.26\xi_{PLAT} + 0.33\xi_{PAFORMAT} - 0.16\xi_{PACONT} + \zeta_6
\]
\[
EMP1 = 0.64\eta_{EMP} + \varepsilon_7
\]

\[
EMP\ 2 = 0.87\eta_{EMP} + \varepsilon_8
\]

\[
EMP\ 3 = 0.74\eta_{EMP} + \varepsilon_9
\]

As the coefficients’ signs were not consistent, the model accounted for the smallest variance of \(EMP\). As this consequence, the predictive relevance was 0, but not negative sign.

As with the previous interpretations of the relationships between the exogenous variables \((PLAT, PAFORMAT, \text{ and } PACONT)\) and the NOS concepts, students who favoured traditional paper-and-pencil based test formats over alternative assessments were likely to understand scientific knowledge to be supported by empirical evidence and experimentally testable. However, the assessment content predicted the opposite way; students who weighed scientific laws and theories in their tests were likely to understand experimental results could be different and empirical evidence may not be an absolute condition for a theory to be scientific.

To identify reasons for this difference, the descriptive statistics (means, \(SD\) and frequencies) and the correlations at the item level were examined. The means of \(EMP2\) fluctuated among schools (refer to Figure 5.24). The correlation between the observed variables of \(PACONT\) and \(EMP1–3\) showed that \(EMP2\) correlated negatively with the variables of \(PACONT\). \(EMP2\) (“The evidence for scientific knowledge does not have to be repeatable”) was stated negatively, so it was inversely arranged before the analysis (The red line is \(EMP2\) in the Figure 5.24).

Most Canadian students who thought that knowledge of science (scientific laws, theories and factual knowledge) was important in their science tests were likely to view that consistency in an experiment and evidence of scientific research were important. Conversely,
students who thought understanding experiments and applying scientific knowledge to diverse situations were most important to the content of tests were also likely to accept realistic views of scientific knowledge. That is, regardless of students’ perceptions of their test contents, they were likely to think that science requires evidence, which was consistent with the descriptive statistics that students held realistic views on EMP.

![Figure 5-24. Means of Observed Variables (EMP1, EMP2, EMP3) (Canadian)](image)

5.7.2.4 Sociocultural Embeddedness

PLAT and PACONT accounted for 40.9% of the variance on Soci-Cultural Embeddedness of Scientific Knowledge. The significant inner coefficients were PLAT and PACONT: $\gamma_{PLAT}=0.18 \ (t(216)=2.70, \ p=.007$, $\gamma_{PACONT}=0.61 \ (t(216)=12.47, \ p<.001$. The effect sizes ($f^2$'s) were 0.05 for PLAT and 0.56 for PACONT. The predictive relevance was 0.21 and the effect size ($q^2$) of PACONT was large 0.23. The predictive relevance was positive, so the observed variables were well reconstructed by the modified model. This means that students who perceive that learning science as transmission of a body of knowledge from teachers to
students, and science tests as knowledge and correct answers were likely to have universal views of science. On the contrary, student-directed learning activities and tasks, and most science tests were allowed alternative views were likely to have multiculturalistic views on science.

Based on the modified model, the prediction equations for \(CUL1\), \(CUL2\) and \(CUL3\) were below:

\[
\eta_{\text{Culture}} = 0.18 \xi_{\text{PLAT}} + 0.61 \xi_{\text{PACONT}} + \xi_7
\]

\[
CUL1 = 0.63\eta_{\text{Culture}} + \epsilon_{10}
\]

\[
CUL2 = 0.80\eta_{\text{Culture}} + \epsilon_{11}
\]

\[
CUL3 = 0.77\eta_{\text{Culture}} + \epsilon_{12}
\]

### 5.7.2.5 Diverse Scientific Research Methods

*Diverse Scientific Research Methods* had the greatest \(R^2\) value in this model. The three exogenous variables accounted for 62.9% of the variance, which was substantial (Chin, 1998). The effect sizes \((f^2)\) of \(\text{PLAT}, \text{PAFORMAT}\) and \(\text{PACONT}\) were 0.65, 0.43 and 0.07, respectively. The inner coefficient for \(\text{PLAT}\) was \(\gamma_{\text{PLAT}} = 0.51, t(216) = 11.04, p < .001\), and the inner coefficient of \(\text{PAFORMAT}\) was \(\gamma_{\text{PAFORMAT}} = 0.44, t = 8.73, p < .001\). \(\text{PACONT}\) had the least effect among the three \((f^2 = 0.07, \text{small effects})\) in accordance with its coefficient, \(\gamma_{\text{PACONT}} = 0.18, t = 3.57, p < .001\). \(q^2\) was 0.37 and the observed variables of *Diverse Scientific Research Methods* \((\text{METH1, METH2 and METH3})\) were well reconstructed by the model. Concerning the prediction of students’ understanding of *Diverse Scientific Research Methods*, the derived equations are followings:

\[
\eta_{\text{METH}} = 0.51\xi_{\text{PLAT}} + 0.44\xi_{\text{PAFORMAT}} + 0.18\xi_{\text{PACONT}} + \xi_8
\]

\[
\text{METH1} = 0.74\eta_{\text{METH}} + \epsilon_{13}
\]
\[ \text{METH}2 = 0.80 \eta_{\text{METH}} + \varepsilon_{14} \]

\[ \text{METH}3 = 0.81 \eta_{\text{METH}} + \varepsilon_{15} \]

To summarize the PLS analysis for the Canadian sample, the research question that whether PLAT, PAFORMAT and PACONT could predict students’ understanding of NOS, the \( R^2 \) ranged from 0.19 to 0.63. The observed variables of the EMP could be weakly accounted for by the variables while other concepts of NOS were relatively well accounted for by the students’ perceptions of their learning activities and assessment. One noticeable point was the opposite coefficient sign of PACONT to EMP. The patterns of answering EMP1, EMP2 and EMP3 were not consistent. Because of this, the consequence was low predictive relevance, 0. Otherwise, students’ perceptions on their learning activities and assessment predict students understanding of NOS.

**5.7.3 Korean Sample**

Through a bootstrapping procedure, \( t \)-statistics were performed for the Korean sample with 1000 sample size and the cases were 319. The paths from PLAT to Tentativeness

\( (\gamma_{\text{PLAT} \rightarrow \text{Tentativeness}} = 0.09, (t(318) = 1.28, p = 0.201) \)

and from PACONT to PAFORMAT

\( (\gamma_{\text{PACONT} \rightarrow \text{PAFORMAT}} = -0.03, (t(318) = 0.60, p = 0.549) \)

were insignificant. Figure 5.25 depicts the theoretical and the modified models. Table 5.18 includes the results of PLS analyses of the modified model such as the inner coefficients, \( t \)-values and \( p \)-values, \( R^2 \), the effect sizes (\( f^2 \)) and the predictive relevance (\( Q^2 \)) and their effect sizes (\( q^2 \)). The highest \( R^2 \) was 0.47 for the Empirical Evidence Based Scientific Knowledge (EMP) while Diversity of Scientific Research Methods (METHOD) were the lowest \( R^2 \) (0.27). All the \( Q^2 \)'s were over zero, so the model well reconstructs the observed variables of the endogenous variables.
### Table 5-18

Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Korean)

\( N = 319 \)

<table>
<thead>
<tr>
<th>Paths</th>
<th>( \gamma )</th>
<th>( t )-value</th>
<th>( p )-value</th>
<th>( R^2 )</th>
<th>( f^2 )</th>
<th>( Q^2 )</th>
<th>( q^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAT</td>
<td>TENTATIVENESS</td>
<td>0.31</td>
<td>5.9</td>
<td>( p &lt; .001 )</td>
<td>0.13</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PACONT</td>
<td>0.49</td>
<td>11.19</td>
<td>( p &lt; .001 )</td>
<td>0.32</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>PLAT</td>
<td>SUBJECTIVITY</td>
<td>0.37</td>
<td>6.36</td>
<td>( p &lt; .001 )</td>
<td>0.14</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PACONT</td>
<td>0.30</td>
<td>5.62</td>
<td>( p &lt; .001 )</td>
<td>0.08</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EMP</td>
<td>0.12</td>
<td>1.84</td>
<td>( p = .067 )</td>
<td>0.02</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>PLAT</td>
<td>METHOD</td>
<td>0.40</td>
<td>6.89</td>
<td>( p &lt; .001 )</td>
<td>0.19</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PACONT</td>
<td>0.13</td>
<td>2.21</td>
<td>( p = .028 )</td>
<td>0.27</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>CULTURE</td>
<td>0.10</td>
<td>1.92</td>
<td>( p = .056 )</td>
<td>0.28</td>
<td>0.27</td>
<td>0.01</td>
</tr>
<tr>
<td>PLAT</td>
<td>METHOD</td>
<td>0.35</td>
<td>4.83</td>
<td>( p &lt; .001 )</td>
<td>0.22</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PACONT</td>
<td>0.19</td>
<td>3.03</td>
<td>( p = .003 )</td>
<td>0.27</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>EMP</td>
<td>0.14</td>
<td>2.62</td>
<td>( p = .009 )</td>
<td>0.29</td>
<td>0.28</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Figure 5-25. Theoretical (a) and Modified Model (b) (Korean)
5.7.3.1 Tentative Nature of Science

The hypothesized model for Korean sample showed that the three exogenous variables accounted for 32.1% of the variance of Tentativeness, which was a weak range (Chin, 1998) albeit the index of convergent validity, AVE, was 0.52 which was over the cut-off value 0.5 in the measurement model of Tentativeness. The composite reliability was 0.76.

The inner coefficients of PAFORMAT and PACONT were significant ($\gamma_{PAFORMAT}$=0.31, $t(318)=5.90$, $p<.001$ and $\gamma_{PACONT}$=0.49, $t(318)=11.19$, $p<.001$) while PLAT coefficient was not. The effect sizes ($f^2$) of PAFORMAT and PACONT were 0.13 (a medium effect) and 0.34 (a large effect). All of the Stone-Geisser’s criterion ($Q^2$) were over 0, which meant the model was able to provide a prediction of the observed variables of Tentativeness (TEN1~TEN3). The predictive equations are the below:

$$\xi_{PAFORMAT} = 0.45x_{52} + 0.31x_{62} + 0.52x_{72} + \zeta_2$$
$$\xi_{PACONT} = 0.74x_{83} + 0.47x_{93} + 0.08x_{103} + \zeta_3$$
$$\eta_{Tentativeness} = 0.31\xi_{FORMAT} + 0.49\xi_{PACONT} + \zeta_4$$

$$TEN1 = 0.64* \eta_{Tentativeness} + \epsilon_1 \ldots \ldots (g)$$
$$TEN2 = 0.78* \eta_{Tentativeness} + \epsilon_2 \ldots \ldots (h)$$
$$TEN3 = 0.78* \eta_{Tentativeness} + \epsilon_3 \ldots \ldots (i)$$

As the equations of (g)~ (i), show, the predicted value of TEN1~ TEN3 depended on $\eta_{Tentativeness}$ also $\eta_{Tentativeness}$ depended on both $\xi_{FORMAT}$ and $\xi_{PACONT}$. That is, for the Korean students, how to assess and what are assessed affected their understanding of tentative nature of science. The research hypotheses, PLAT, PAFORMAT and PACONT can affect/predict students’ understanding of NOS were partially rejected (PLAT) and partially accepted (PAFORMAT and PACONT).
5.7.3.2 Subjectivity

The hypothesized model accounted for 40.6% of the total variance of Subjectivity. The paths from the three exogenous variables were significant (γ_{PLAT}=0.37, t(318)=6.35, p<0.001; γ_{PAFORMAT}=0.30, t=5.62, p<.001; γ_{PACONT}=0.16, t=2.98, p=.003). The mediator path from PAFORMAT to PLAT and PACONT to PLAT were also significant (γ_{PAFORMAT,PLAT}=0.58, t=14.63, p<0.001 and (γ_{PACONT,PLAT}=0.34, t=7.02, p<0.001). Therefore, PAFORMAT and PACONT affected Subjectivity directly and indirectly. The effect sizes of PLAT, PAFORMAT and PACONT were $f^2=0.14$ (a medium effect), $f^2=0.08$ (weak) and 0.03 (weak), respectively. Based on the modified model, the predictions of the observed variables of Subjectivity would follow the equations:

$$\xi_{PLAT}=0.34x_{11}+0.34x_{21}+0.53x_{34}+0.38x_{44}+0.58\xi_{PAFORMAT}+0.34 \xi_{PACONT} + \xi_{7}$$

$$\xi_{PAFORMAT}=0.45x_{52}+0.31x_{62}+0.52x_{72} + \xi_{2}$$

$$\xi_{PACONT}=0.74x_{83}+0.47x_{93}+0.08x_{103}+\xi_{3}$$

$$\eta_{Subjectivity}=0.37\xi_{PLAT} + 0.30 \xi_{PAFORMAT} + 0.16\xi_{PACONT} + \xi_{5}$$

$$SUB1=0.77\eta_{Subjectivity} + \varepsilon_4------(j)$$

$$SUB2=0.67\eta_{Subjectivity} + \varepsilon_5------(k)$$

$$SUB3=0.76\eta_{Subjectivity} + \varepsilon_6------(l)$$

As the values of Stone-Geisser’s criterion, $Q^2$, were positive, the equations (j, k and l) were well reconstructed to predict the endogenous observed variables (SUB1, SUB2 and SUB3).

5.7.3.3 Empirical Evidence Based Scientific Knowledge

The three exogenous variables effectively accounted for the variance of Empirical Evidence Based Scientific Knowledge. About 47.0% of the variance was accounted for the
variables and which was medium effects (Chin, 1998). The path coefficients for PLAT, PAFORMAT and PACONT were \( \gamma_{PLAT}=0.12 \) \((t(318)=1.84, p=.067)\), \( \gamma_{PAFORMAT}=0.28 \) \((t(318)=2.98, p=.003)\) and \( \gamma_{PACONT}=0.55 \) \((t(318)=11.87, p<.001)\), respectively. As the coefficients showed, PACONT had a large effect \( (f^2=0.43) \) while PLAT and PAFORMAT were weak effects \( (f^2=0.02 \text{ for PLAT and } f^2=0.09 \text{ for PAFORMAT}) \). The prediction for students’ understanding of Empirical Evidence Based Scientific Knowledge could be based on following equations:

\[
\eta_{EMP} = 0.12 \xi_{PLAT} + 0.28 \xi_{PAFORMAT} + 0.55 \xi_{PACONT} + \zeta_6
\]

\[
EMP1 = 0.76 \eta_{Subjectivity} + \epsilon_7
\]

\[
EMP2 = 0.86 \eta_{Subjectivity} + \epsilon_8
\]

\[
EMP3 = 0.72 \eta_{Subjectivity} + \epsilon_9
\]

5.7.3.4 Sociocultural Embeddedness

The three exogenous variables accounted for 28.0% of the variance of Socio-Cultural Embeddedness. The coefficients of PLAT, PAFORMAT and PACONT were significant \( \gamma_{PLAT}=0.40 \) \((t(318)=6.89, p<.001)\) and \( \gamma_{PAFORMAT}=0.13 \) \((t(318)=2.21, p=.028)\) \((r_{PACONT}=0.10, t=1.92, p=.056)\). As the model did not account for the variance with a high percentage, the effect sizes \( (f^2) \) were not substantial; they were 0.13 and 0.01 and 0.01 for PLAT, PAFORMAT and PACONT, respectively. The Stone-Geisser’s criterion was 0.16, and the corresponding effect sizes \( (q^2) \) were very small \( (0~0.06) \). The model could be reconstructed the observed variables of Cultural Embeddedness. The equations were the followings:

\[
\eta_{Culture} = 0.40 \xi_{PLAT} + 0.13 \xi_{PAFORMAT} + 0.10 \xi_{PACONT} + \zeta_7
\]

\[
CUL1 = 0.72 \eta_{Culture} + \epsilon_{10}
\]
\[ \text{CUL2} = 0.80 \eta_{\text{Culture}} + \varepsilon_{11} \]
\[ \text{CUL3} = 0.80 \eta_{\text{Culture}} + \varepsilon_{12} \]

In light of the research hypotheses on the relationships between \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT} and understanding of \textit{Cultural Embeddedness}, the hypotheses were accepted. The three variables can be effect factors to affect/predict students understanding of NOS.

5.7.3.5 Diverse Scientific Research Methods

Twenty nine percent of the total variance of \textit{Diverse Scientific Research Methods} (\textit{METHOD}) was accounted for by the three exogenous variables. The inner coefficients of \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT} were \( \gamma_{\text{PLAT}} = 0.35, (t(318)=4.83, p<.001), \gamma_{\text{PAFORMAT}} = 0.19, (t=3.03, p=.003) \) and \( \gamma_{\text{PACONT}} = 0.14, (t=2.62, p=0.009) \), respectively. The effect sizes for each exogenous variable were 0.10, 0.03, 0.01. The predictive relevance (\( Q^2 \)) was 0.15, and the effect sizes of the predictive relevance ranged from 0.01 to 0.05. Although the predictive relevance was positive, the model did not account for the variance with a high percentage. As when the outer measurement model was reviewed, the Korean sample model was not solid since \textit{METH1} (\( \lambda = 0.50 \)) did not have a high outer weight. The prediction equations for the Korean sample can reconstruct the observed variables of \textit{METHOD} as follows:

\[ \eta_{\text{METH}} = 0.35 \xi_{\text{PLAT}} + 0.19 \xi_{\text{PACONT}} + 0.14 \xi_{\text{PACONT}} + \zeta_{8} \]
\[ \text{METH1} = 0.50 \eta_{\text{METH}} + \varepsilon_{13} \]
\[ \text{METH2} = 0.85 \eta_{\text{METH}} + \varepsilon_{14} \]
\[ \text{METH3} = 0.84 \eta_{\text{METH}} + \varepsilon_{15} \]

Summing the Korean sample analyses, the three exogenous variables accounted for the variables of students’ understanding of NOS from 28%~47%. The corresponding effect sizes (\( f^2 \)) ranged from 0.01 (very weak) to 0.43 (substantial). All Stone-Geisser’s criterion values
were positive and the model could well reconstruct the observed variables of the endogenous variables (NOS understanding). Therefore, concerning this research questions, students’ perceptions of how students were taught, how their learning was assessed and what were assessed could be factors for their understanding of NOS.

5.7.4 Combined Sample

As seen in the measurement models’ validity and reliability assessment, the larger sample size was the stable the models became. The combined sample size was 536, and all inner coefficients were significant. Through a bootstrapping procedure, $t$-statistics were preformed for with 1000 sample size and the cases were 536. Figure 5.26 depicts the theoretical model. Table 5.19 includes the results of PLS analyses. The highest $R^2$ was 0.46 for the Subejctiveness while Cultural Embeddedness was the lowest $R^2$ (0.27). The effect sizes ($f^2$) were ranges from 0.48 for the path from PLAT to Subjectivity to 0.03 (a weak size) for the paths from PLAT to Tentativeness, from PACONT to EMP, from PFORMAT to Cultural Embeddedness, from PACONT to Methods. All the $Q^2$s were much greater than 0 (ranged from 0.14 to 0.23), so the model can well reconstruct the observed variables of the endogenous variables.
### Table 5-19

*Path Coefficients, Accountability, Predictive Relevance and Effect Sizes (Combined Sample)*

\(N=536\)

<table>
<thead>
<tr>
<th>from</th>
<th>to</th>
<th>(\gamma)</th>
<th>t-value</th>
<th>(p)-value</th>
<th>(R^2)</th>
<th>(\beta)</th>
<th>(Q^2)</th>
<th>(q^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLAT</strong></td>
<td></td>
<td>0.16</td>
<td>4.08</td>
<td>(p&lt;.001)</td>
<td>0.03</td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td><strong>TENTATIVE</strong></td>
<td>0.19</td>
<td>4.47</td>
<td>(p&lt;.001)</td>
<td>0.3</td>
<td>0.04</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td><strong>NESS</strong></td>
<td>0.39</td>
<td>10.68</td>
<td>(p&lt;.001)</td>
<td>0.20</td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td><strong>PLAT</strong></td>
<td></td>
<td>0.24</td>
<td>6.65</td>
<td>(p&lt;.001)</td>
<td>0.48</td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td><strong>SUBJECTIVE</strong></td>
<td>0.34</td>
<td>10.06</td>
<td>(p&lt;.001)</td>
<td>0.46</td>
<td>0.17</td>
<td>0.27</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td><strong>NESS</strong></td>
<td>0.36</td>
<td>9.89</td>
<td>(p&lt;.001)</td>
<td>0.20</td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td><strong>PLAT</strong></td>
<td></td>
<td>0.31</td>
<td>6.9</td>
<td>(p&lt;.001)</td>
<td>0.09</td>
<td></td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td></td>
<td>0.2</td>
<td>4.66</td>
<td>(p&lt;.001)</td>
<td>0.24</td>
<td>0.04</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td><strong>EMP</strong></td>
<td>0.15</td>
<td>3.03</td>
<td>(p&lt;.001)</td>
<td>0.03</td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td><strong>PLAT</strong></td>
<td></td>
<td>0.25</td>
<td>6</td>
<td>(p&lt;.001)</td>
<td>0.07</td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td></td>
<td>0.16</td>
<td>4.08</td>
<td>(p&lt;.001)</td>
<td>0.27</td>
<td>0.03</td>
<td>0.15</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td><strong>CULTURE</strong></td>
<td>0.3</td>
<td>7.44</td>
<td>(p&lt;.001)</td>
<td>0.09</td>
<td></td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td><strong>PLAT</strong></td>
<td></td>
<td>0.37</td>
<td>10.55</td>
<td>(p&lt;.001)</td>
<td>0.20</td>
<td></td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td><strong>PAFORMAT</strong></td>
<td></td>
<td>0.31</td>
<td>8.87</td>
<td>(p&lt;.001)</td>
<td>0.4</td>
<td>0.13</td>
<td>0.23</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>PACONT</strong></td>
<td><strong>METHOD</strong></td>
<td>0.16</td>
<td>4.48</td>
<td>(p&lt;.001)</td>
<td>0.03</td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
</tbody>
</table>

*Figure 5-26. Theoretical Model (Combined Sample)*
5.7.4.1 Tentative Nature of Science

The three exogenous variables accounted for 30% of the variance of Tentativeness. The inner coefficients of PLAT, PAFORMAT and PACONT were $\gamma_{PLAT}=0.16, (t(535)=4.08, p<.001)$, $\gamma_{PAFORMAT}=0.19, (t=4.47, p<.001)$ and $\gamma_{PACONT}=0.39, (t=10.68, p<.001)$, respectively. The effect sizes for each exogenous variable were 0.03, 0.04, and 0.20. The predictive relevance ($Q^2$) was 0.16, which meant that the model could well predict the observed variables of Tentativeness.

The equations based on the model are the below:

$$\xi_{PLAT} = 0.40x_{11} + 0.35x_{21} + 0.43x_{31} + 0.40x_{41} + 0.38\xi_{PAFORMAT} + 0.16 \xi_{PACONT} + \xi_I$$

$$\xi_{PAFORMAT} = 0.51x_{52} + 0.42x_{62} + 0.34x_{72} + 0.21\xi_{PACONT} + \xi_2$$

$$\xi_{PACONT} = 0.40x_{83} + 0.51x_{93} + 0.37x_{103} + \xi_3$$

$$\eta_{Tentativeness} = 0.16\xi_{PLAT} + 0.19\xi_{FORMAT} + 0.39\xi_{PACONT} + \xi_4$$

$$TEN1 = 0.71* \eta_{Tentativeness} + \varepsilon_1$$

$$TEN2 = 0.71* \eta_{Tentativeness} + \varepsilon_2$$

$$TEN3 = 0.78* \eta_{Tentativeness} + \varepsilon_3$$

5.7.4.2 Subjectivity

From the model, about 46% of the variance of Subjectivity was accounted for by the three exogenous variables. All inner coefficients were significant and the effect size ($f^2$) of the path PLAT was 0.48, which was large. The other two paths’ effect sizes were medium; $f^2=0.17$ for PAFORMAT and $f^2=0.20$ for PACONT. The Stone-Geisser’s criterion was 0.27, so the model had predictive relevance. The equations to predict the observed variables of Subjectivity are the followings:

$$\eta_{Subjectivity} = 0.24\xi_{PLAT} + 0.34\xi_{FORMAT} + 0.36\xi_{PACONT} + \xi_5$$
\[ SUB1 = 0.80 \times \eta_{\text{Subjectivity}} + \varepsilon_4 \]
\[ SUB2 = 0.73 \times \eta_{\text{Subjectivity}} + \varepsilon_5 \]
\[ SUB3 = 0.80 \times \eta_{\text{Subjectivity}} + \varepsilon_6 \]

5.7.4.3 Empirical Evidence Based Scientific Knowledge (EMP)

The combined model was actually influenced by both Canadian and Korean samples. As Korean sample size is larger than Canadian sample size, the results were more close to the results of Korean sample. In EMP analysis, in the Canadian sample the path coefficient from PAFORMAT was negative, but in the combined sample the path was positive (0.16); Probably, the sign and the magnitude resulted in the Korean data. Figure 5.27 shows the means of Korean schools in EMP1, 2, and 3 changed with similar trends, but not Canadian schools (1, 2, 3, and 4).

![Figure 5-27. Means of EMP](image)

Twenty-seven percent of the variance of EMP was accounted for by the three exogenous variables, and the corresponding effect sizes ($f^2$) were 0.09, 0.04 and 0.03 (small
effects). The predictive relevance was 0.14 greater than 0. The observed variables of the endogenous variables could be well reconstructed by the model.

\[ \eta_{EMP} = 0.31 \xi_{PLAT} + 0.2 \xi_{FORMAT} + 0.15 \xi_{PACONT} + \xi_6 \]

\[ EMP1 = 0.80 \eta_{EMP} + \epsilon_7 \]

\[ EMP2 = 0.73 \eta_{EMP} + \epsilon_8 \]

\[ EMP3 = 0.80 \eta_{EMP} + \epsilon_9 \]

### 5.7.4.3 Cultural Embeddedness

The three exogenous variables accounted for 27\% of the variance of Cultural Embeddedness of the combined sample. The corresponding effect sizes (\(f^2\)s) were weak. The predictive relevance (\(Q^2\)) was 0.15 and the model could reconstruct the observed variables of Cultural Embeddedness. The equations are the below:

\[ \eta_{Culture} = 0.25 \xi_{PLAT} + 0.16 \xi_{FORMAT} + 0.30 \xi_{PACONT} + \xi_7 \]

\[ CUL1 = 0.71 \eta_{Culture} + \epsilon_{10} \]

\[ CUL2 = 0.80 \eta_{Culture} + \epsilon_{11} \]

\[ CUL3 = 0.78 \eta_{Culture} + \epsilon_{12} \]

### 5.7.4.3 Diverse Scientific Research Methods (METHOD)

R2 was 0.40 and the effect sizes were 0.03 (\(f^2_{PACONT}\)), 0.20 (\(f^2_{PLAT}\)), and 0.13 (\(f^2_{PFORMAT}\)). The Stone-Geisser’s criterion was 0.23, which was much greater than 0. The model had predictive relevance. The following equations were to predict for the observed variables of the endogenous variable based on the model.

\[ \eta_{Method} = 0.25 \xi_{PLAT} + 0.16 \xi_{FORMAT} + 0.30 \xi_{PACONT} + \xi_8 \]

\[ METH1 = 0.64 \eta_{Method} + \epsilon_{13} \]
\[ METH 2 = 0.83 \ast \eta_{Method} + \varepsilon_{14} \]

\[ METH 3 = 0.80 \ast \eta_{Method} + \varepsilon_{1} \]

As the sample size was larger, the measurement and structural model were stabilized. All paths were significant, and all \( R^2 \) were medium. The results implied that students’ perceptions of learning activities and tasks, assessment formats and content could be effective factors on students’ understanding of NOS.

**5.7.7 Section Summary**

Generally, the Canadian model had a high explicability across 4 NOS concepts. Especially the concepts of \textit{METHOD} for the Canadian model were accounted for 63\% of the variance. However, for the Korean model two concepts of NOS (\textit{Subjectivity} and \textit{EMP}) had moderate values while three concepts were weak (Chin 1998, p.323).

While it was postulated that the factors, \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT}, would affect students understanding of NOS, not all three were effective for all 5 concepts of NOS. The decisions whether the hypotheses should be rejected or retained are presented in Table 5.20. The null and alternative hypotheses were:

\[ H_0: \text{There is no effect of } PLAT/PAFORMAT/PACONT \text{ on students’ understanding of Nature of Science.} \]

\[ H_1: \text{not } H_0 \]
Table 5-20

*Effects of PLAT/PAFORMAT/PACONT on Understanding of NOS*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Sample</th>
<th>Tentativeness</th>
<th>Subjectivity</th>
<th>EMP</th>
<th>Socio-Cultural Embeddedness</th>
<th>Diverse Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAT</td>
<td>Canadian</td>
<td>Reject $H_0$</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Korean</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>PAFORMAT</td>
<td>Canadian</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Korean</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>PACONT</td>
<td>Canadian</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Korean</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Retain $H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
<td>Reject $H_0$</td>
</tr>
</tbody>
</table>

Almost all the significant paths were positive except the path from PACONT to *Empirical Evidence Based Scientific Knowledge* in the Canadian sample, which meant that if the scores of the independent latent variables increased, the scores of the dependent variables also increased. That is, the data were rearranged to be conceptually consistent that low scores were for teacher directed learning and assessment emphasizing the knowledge of science and realistic views of NOS, while high scores were for student-directed learning and assessments stressing knowledge about science and understanding of science, and relativists’ views. Therefore, the more students perceived their learning as student-directed, with diverse forms of assessment and content the more they were likely to have relativistic views of science, and *vice versa*.

**5.8 Chapter Summary**

The quantitative data collected by a survey were analyzed using statistics software, SPSS and SmartPLS. The results could answer the research questions if there were any differences in the three domains between Canadian and Korean students and the associations
among variables in students’ perceptions of their learning activities and tasks (\textit{PLAT}), the perceptions of assessment formats (\textit{PAFORMAT}) and content (\textit{PACONT}) and their understanding of NOS.

The quality of the measuring instruments was checked in terms of the validity and reliability and it was acceptable overall. A few observed variables did not meet the optimal recommendations; however, the exogenous variables of the measurement models were formatively designed which cannot be applied for the standards of reflective measurement models. In the reflective models, no observed variables were so extreme that they could not be accepted. When the sample size was large (the combined sample), the unstable variables met the standards.

Multivariate Analysis Of Variance (MANOVA) examined the effects of country and school differences on students’ \textit{PLAT}, \textit{PAFORMAT}, \textit{PACONT} and understanding of NOS. Both country and school differences were predictive of the variances of the variables of \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT}. Relatively the two factors were less strong to explain the variances of understanding of NOS than the variances of \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT}.

Regarding the accountability of \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT} to NOS understanding, $R^2$s ranged from 0.19 (\textit{Empirical Evidence Based Scientific Knowledge} for the Canadian sample) to 0.63 (\textit{Diverse Methods of Scientific Research} for the Canadian sample). The exogenous constructs’ values of the predictive relevance ($Q^2$) were higher than 0, so the models could predict the understanding of NOS concepts; however, most effect sizes were relatively weak according to Chin’s (1998) standards.

After the quantitative data analyses, three issues were emerged; firstly, as to school effects on the constructs, for the time being, it would appear that there exist teacher effects on
students’ understanding of NOS, which would be evidence contrary to some previous literature; secondly, the kinds of schools and the constructs differences were emerged. School 1 was a religion-based school while the other 3 Canadian schools were public schools. This school difference may be a significant factor that influences on the understanding of NOS. Thirdly, if gender could be a factor for PLAT, PAFORMAT, PACONT and understanding of NOS. School 5 was a boys’ school while the other two Korean schools were for both boys and girls’ schools. These emerging interests will be further discussed once the open-ended NOS responses and the semi-structured interview analyses are completed in Chapter 7.
6. RESULTS OF THE QUALITATIVE DATA ANALYSIS

6.1 Introduction and Overview

This chapter reports the analysis of responses to the five open-ended questions about NOS, and to the semi-structured interviews. This qualitative data provided in-depth understanding of the four domains (PLAT, PAFORMAT, PACONT and Understanding of NOS) as well as clarification and compensation for the vagueness of close-ended answers.

This chapter consists of four sections. Section 6.2 describes the features of responses to the open-ended questions of NOS, with subsections for each NOS concept: Tentativeness, Subjectivity, Empirical Evidence Based Scientific Knowledge, Sociocultural Embeddedness and Diversity of Scientific Research Methods. Section 6.3 sets out the semi-structured interviews. Common themes are grouped together, and atypical opinions are reported. The crucial point of the interviews was to identify the results of the PLS analysis; i.e., whether PLAT, PAFORMAT and PACONT could be good predictors of students’ understanding of NOS. The last section, 6.4 includes a summary of the qualitative data analysis and reiterates the main findings relating to the research questions.

6.2 Responses to the Open-Ended NOS Questions

The open-ended questions for this study were derived from two different instruments. Five questions were based on VNOS-B and C (Lederman et al., 2002) and one was adapted from SUSSI (Liang et al., 2008). Recently, VNOS-C has been frequently used in NOS studies (Lederman et al., 2002, 2007; Martin-Dunlop, 2004; Schwartz, 2004). Most of these studies used a small number of samples because of the amount of writing required and the heavy demand upon researchers for interpretation. The developers of these instruments have strongly
recommended conducting follow-up interviews to discern whether or not participants understand the questions correctly, and so semi-structured interviews were conducted with 9 volunteers for follow-up.

In analyzing qualitative data, a number of blank sheets and overlapping responses required attention. First, blank or sketchy responses were problematic. Since Liang et al. (2009) had noticed several blank responses to open-ended questions when students found writing a burden, this study tried to reduce that burden by splitting the open-ended questions into two types. Nevertheless, Canadian students left 10.6% and Korean students left 17.5% of their responses blank. Second, in transforming qualitative data into numeric data, accuracy is crucial; however, some responses included more than one concept, so that these were sometimes counted twice or three times. These multiple counts meant that the responses exceeded 100%.

The results and interpretations of the major concept of NOS are reported in the following order: 1) common themes through coding and reconstructions of the responses, 2) comparisons between the results of close-ended and open-ended answers using data transformations; 3) identifications of discrepancies between answers to close-ended and open-ended questions, 4) similarities and differences between Canadian and Korean students’ understanding of each construct, and 5) emergent themes or atypical points. Excerpts from both representative and atypical answers were provided as evidence for the claims and assertions made throughout the narrative and multi-dimensional layers of writing. Findings are classified into sections corresponding to each open-ended question as follows:

Section 6.2.1- Tentativeness (Tentative nature of science and scientific knowledge)

Section 6.2.2- Subjectivity (Theory-laden scientific observation)
6.2.1 Tentativeness

Science and scientific knowledge are tentative. That does not mean that scientific knowledge is wrong, but it does mean that it can be modified or replaced when new interpretations of existing data and new evidence become available. At the same time, scientific knowledge is also durable since it rests upon the ongoing support of evidence (NRC, 1996).

Question: After scientists have developed a theory (e.g. atomic theory or cell theory), does the theory ever change? If you believe that theories do change, explain why we bother to teach scientific theories. Defend you answer with examples.

Abd-El-Khalick (1998) illuminates the intention of this question: “to assess understandings of the tentative nature of scientific claims, why these claims change—students mostly attribute such change solely to the accumulation of new facts, and the role that scientific theories play in science (p.534).”

The primary reference for the interpretation of this question was the conceptually clustered matrix, Figue 4-9, which was based on the authors’ intention and guidelines. Two distinct categories of positions were briefly described: naïve realist views and informed views. Another category listed responses as “not classifiable”, in cases where no reason was provided or answers could not be assigned to either distinct position. Abd-El-Khalick and Lederman (2002), Lederman et al. (2002) and Liang et al. (2009) stress that the idea that scientific knowledge does not change, or that discoveries make scientific knowledge change, is a naïve
Naïve realist views hold that scientific theories or laws exist in nature independent of human cognition; scientists are devoted to *discovering* these existing laws and entities. Since the advancement of the observatory and the growth of detective technology has expanded the scope of people’s sensory recognition, scientists can discover these existing laws and principles both in macro- and micro-worlds. In contrast, they emphasize that in an informed view, scientists are properly seen as *construing* laws and theories to understand nature better. They reconceptualize existing data, just as the history of science documents gradual changes in scientific knowledge. Such construed knowledge need not be absolutely ‘true’; it can be changed both by new discoveries and by the re-working of previous interpretations.

### 6.2.1.1 Common Themes

In total 264 answer sheets were collected on this question. Of these, 17 cases were blank, and 32 cases did not provide any reasons for “changes” or “does not change.” Almost all the students agreed with the notion that scientific theories and laws change as time goes on due to 1) corrections of errors of scientific theories and laws (25%), 2) discoveries of new evidence through advanced technology (16%), and 3) changes of human ideas (3%). Some responses indicated there is no change (12%) either because the knowledge at issue was ‘proven’ or because it dealt with the ‘unchanging’ physical world. Few provided in-depth descriptions of what errors in science are corrected, or how; it was accordingly hard to categorize these comments as reflecting “naïve realistic” or “informed” views. Turning to errors in scientific theory, some students viewed errors as the main cause for change in scientific knowledge, while a few felt major theories would not change regardless of such errors. The major theories become more accurate. Table 6.1 indicates the sub-codes and distributions of students’ responses.
Table 6-1

Summary of Common Themes on Tentativeness

\[ N_{Ca}=103, N_{Ko}=161 \]

<table>
<thead>
<tr>
<th>Change/No</th>
<th>Sub-code</th>
<th>Canadian</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Descriptions of Reasons)</td>
<td>No. of responses</td>
<td>%</td>
</tr>
<tr>
<td>No change</td>
<td>Errors of scientific theories made scientific knowledge change.</td>
<td>27</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>New evidence discovered with the help of advanced technology.</td>
<td>20</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>New interpretation of existing data and natural phenomena or human’s ideas changes</td>
<td>4</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>No reason was provided.</td>
<td>10</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>14</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>(Nature changes or multiple reasons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No change</td>
<td>Fundamental theories do not change, but they have become accurate.</td>
<td>5</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>Empirical evidence supported theories do not change.</td>
<td>5</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>No reason was provided</td>
<td>2</td>
<td>1.9</td>
</tr>
<tr>
<td>No Response</td>
<td></td>
<td>3</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Error corrections make scientific knowledge change: A majority of students thought that the scientific laws and theories are not perfect and include errors because, in particular, some of
these were developed in ancient or medieval times. Many of their errors reflect a less
developed technology or a limited understanding of nature.

[T]hey [Theories] do change because a theory is a prediction and may not be fully
correct so you can apply minor or major changes to your theory. (2-14)

Theories are not perfect......the atomic theory is a good example. In ancient
times the Greek philosopher, Tales, suggested the atomic theory. Dalton’s
atomic theory was much improved over Tales’s theory. Of course Dalton’s four
hypotheses were wrong. Current physics has revealed that atoms are not shaped
as hardballs, which cannot be broken down into smaller pieces; there exist
isotopes and a nucleus can also be broken into corpuscles. 이론들은 완벽하지
아니기 때문에 변한다. ....예를 들자면 고대 그리스의 탈레스는 원자설을 주장했다. 돌턴의
원자설은 탈레스 것보다 발전되었지만, 돌턴의 원자설도 틀렸음을 밝혀졌다. 현대
물리학은 원자는 단단한 공같이 생긴것이 아니라라는 것을 밝혔고, 동위원소도 있다는
것을 밝혔다. 원자핵도 쪼개지는 것을 알아냈다. (5-5)

Among the views that error correction made scientific theories change, some responses
showed a hierarchical concept of scientific knowledge which contrasts scientific theories and
laws. They were also confused a scientific theory with a hypothesis. To summarize those
students’ views theories are mere opinions and educated guesses, lacking evidence or accuracy
to qualify as scientific laws. Upon accumulating sufficient evidence, a theory becomes a law.
Some scholars (Lederman et al., 2002; McComas and Olson, 1998) clearly distinguish
scientific laws from theories but some do not agree with the concept (e.g., Hickey, 2005; Trefil
& Harzen, 2010) since the descriptive perspectives of scientific laws have become vague when
scientific observations are not seen as ‘objective’ but theory-laden. Therefore, the borderline
between the description of a natural phenomenon (scientific laws) and the explanation of it
(scientific theories) is not clear. Although the distinction is arguable, the hierarchical
relationship assumed cannot be an informed view, and the confusion cannot be an informed
view either.
Over time, theories change and become facts and laws. Scientific theories are educated guess so that eventually someone can make it a fact or a false statement. (2-21)

Theories do change because they aren’t 100% fully proven. But we learn them [be]cause there are theories that are 90-99% proven and are reliable like the cell theory or the string theory which hasn’t been proven yet. But it does provide us with some answers. (3-14)

Theories are taught because although they are not proven laws (hence the word “theory”) they are still important to know as they are mostly views as correct, with no proof against it or any that entirely disproves it (emphasis original). (3-28)

**Discoveries change scientific knowledge:** Another dominant reason for changing theories and laws advanced in the responses was the discovery of new phenomena or evidence, whether micro-particles in matter or macro-objects in space. Modern scientists can now expand the limitation of our sensory organs through technology. These discoveries become discrepant with existing theories, which thus turn out to be false. Philosophers of science have argued over whether we “discover” or “construe” scientific knowledge. Perspectives which contend that we “discover” something ‘out-there’ imply that scientific knowledge is less changeable and diverse in understanding nature, while perspectives suggesting that we “construe” are more flexible and leave more room to accept diverse ideas about nature (Godfrey-Smith, 2003; Hickey, 2005). Against this background, the idea that scientists “discover” something ‘out there’ is a realistic view (Hickey, 2005; Lederman et al., 2002).

They [scientific theories] do change because I believe science is infinite. There’s always something new and waiting to be discovered. We are learning scientific theories because it helps us understand why our world is this way. Maybe in the future, the theory taught now will be only a fraction of the new and improved theory, but we are always learning. (2-16)

When a scientist developed a theory, it is according to what she/he found to prove its right evidence. But as technology gets better, scientists are able to
find more impossible results and that is why theories would be changed for sure. (3-9)

Yes, I think the theories do change over time, because technology is continuously becoming more advanced. The improvement in technology may lead to further discoveries. For example, particles were discovered. For example, particles were considered to be the smallest unit in life until atoms were discovered. (2-11)

**Recognition changes engender changes in scientific theories:** A minority agreed to the view that the theories change because science is a human enterprise, which studies nature. As both nature and humanity are changeable, our understanding of nature continuously changes, and so does science. Theories may also change because scientists observe natural phenomena differently or because they change their ontological views of scientific theory. This is the view that Lederman et al. (2002) and Liang et al (2008) classified as informed. A few excerpts of students’ open-ended answers to this question follow below.

Theories are changing because they are people’s ideas and ideas can be changed because people can think differently over the generations. It is always good to know before a hand than to wait years for it to be proven correct. I cannot remember the scientist’s name. He thought that everything has natural chemicals inside (when I learned first about chemistry), so the chemicals made contract and shrink automatically, but it was proven incorrect. Particle theory replaced it. (School 4 HJ)

I believe that scientific theories can change after a period of time. I don’t believe that the entire idea of the theory will change within a short period of time but over the years technology will become more advanced and help us find new discoveries. One example is near the 1400s and 1500s. People “knew” that the earth was flat until Christopher Columbus proved his idea to the church. Another example is how people “knew” that the earth was the centre of the universe until Galileo proved that wrong as well (underline original). (3-23)

**No change of scientific knowledge:** Relatively few students (12%) held that once the major concept of a theory was proven, the theory does not change. Because the theory was supported
by evidence, replicated and applied to diverse areas, it is thought to be proven true. Such a
theory may be modified to correct minor errors rather than completely changed.

Scientific theories do not change because they have been proven to be correct
by experimental results (과학이론들은 변하지 않습니다. 왜냐하면 실험결과를 통해 확인된 것들이기 때문입니다.) (5-30)

I don’t think scientific knowledge changes. If a scientific theory has been
applied to other areas and helped develop these areas, it is not easy to change.
If the theory helped develop knowledge, then it was proved to be a fact. A
proven fact needs not to change and it has become a fixed fact. In science, only
true facts can survive, and if there is something wrong or untrue, it will
definitely change. (변화하지 않다고 생각한다. 어떤 과학 이론이 이미 적용 되어 발전되었다면 바꾸기가 매우 어려운데다 발전되었다는 점으로 그 이론이 사실이라는 것을 입증하게 된다. 사실을 입증되면 바꿀 이유도 없고 이론이 고정되어 고정 풀 수가 없다. 과학에는 사실만이 존재한다. 거짓이 있다면 그것은 언제든지 바뀐다. (7-23)

Scientific knowledge is absolute truth and cannot be changed, if it meets
specific conditions…Scientific discussions seek to find absolute truths in
nature. For this reason, some theories are ‘admitted’ to the standing of
scientific knowledge, while others are not. (quotation original)절대 불변의 진리이자 사실이며 바뀔 수 없다. 까다로운 조건들을 만족한다면, 과학적 사실은 불변하다…과학의 토론은 절대 불변의 진실을 찾아내기 위할 뿐이다. 따라서 과학은 ‘인정’이 필수적이지만 다른 과목들은 인정이 필수적이지 않다. (6-5)

Some students had a clear and distinct idea of progress in the accumulation of scientific
knowledge. As our fundamental theories about nature were developed, scientists corrected
errors, aided by current technology and accumulating evidence. In this way theories grew
more precise and accurate. This view actually resembles the stance taken by Albert Michelson,
who won the Nobel Prize in Physics in 1907. At that time, contemporary physicists felt they
understood most of the basic laws of nature; future physicists were left to refine and calibrate
the laws and theories of physics more precisely. Michelson remarked that future scientists
would fight “in six decimal points.” This outlook relied on the accumulation of new evidence
to make scientific knowledge more precise. This view was similar to Lakatos’s Scientific
Research Programs (1978). The core theories like Newton’s gravitation theory rarely change but theories around the core theories (theories in the protective belt) change. From the perspective of paradigm shift (Kuhn, 1962), the previous and the current theories are incommensurable. Many philosophers attack his view of incommensurability, arguing that the reference of nature is the same and scientists in different paradigms can communicate with little or no effort (Curd & Cover, 1998). Therefore, this view of progress and accumulation could be acceptable.

Specifcs of a theory may change due to new discoveries, but the principle remains, e.g., Dalton had an atomic theory-how the atom was structured. Bohr and Rutherford had other more accurate theories later in history that provides us with a better understanding. These theories were built upon the principles of Dalton’s theory or in order to understand Bohr and Rutherford’s theories, students must first have an understanding of some basic specifics from Dalton’s theory. (2-13)

Scientific theories give a base from which further ideas and theories can be expanded from. ..... An example of such a theory would be the Big-Bang theory. Its validity is often debated but accepting it as the most-likely theory for the way to world began has opened idea in various areas of research such as space science or geology, etc. Therefore, theories are crucial in providing headway to other areas and in time even they become out-dated continue to support new ideas. Science depends heavily on evolving ideas and progress without a base, it would be utterly impossible for any form of science to exist and prove itself. (2-6)

*Different views on the same scientific theory:* Two excerpts cited above (5-5 and 2-13) offered the same example of the atomic theory and explained it similarly. However, they differed in their interpretations of the tentative nature of science. One student (5-5) quoted the development of atomic theory as evidence that scientific theories change; the other (2-13) held that the basic ideas of atomic theory did not change, but were made more accurate. Both students probably knew equally well about how the atomic theories have been developed. To
classify the one standpoint as “informed” and the other as “naïve realistic” oversimplifies the matter. Such a judgment also violates the intent of NOS education, which does not look for a student’s understanding of a specific theory, but for what a student focuses on. Such differences of view on the same event could also apply to the study’s questions about scientists’ interpretations of the extinction of dinosaurs. These two students would draw attention in their other open-ended responses.

6.2.1.2. Clarification of closed-ended answers

Generally, responses to open-ended questions were consistent with responses to close-ended questions. In the close-ended question (TEN1), 314 students (59%) believed that fundamental theories do change; 56 students (10%) believed that fundamental theories do not change. In the open-ended responses, 167 students (63%) spoke for “changes”; while 32 students (12%) spoke for “does not change”.

Looking into the reasons for changes offered in detail, the study was still left with some vagueness and discrepancies. While the close-ended questions used the word “interpretation”, very few students answering the open-ended questions mentioned concepts like “re-interpretation of data” or “re-conceptualization of existing ideas from a different perspective.” This result suggests that students did feel scientific theories change mainly from the correction of errors and the addition of new evidence. Re-conceptualization or differing views of natural phenomena simply did not figure in their thinking.

Many students who felt that current scientific theories were more accurate than those previous claimed that scientific theories constantly change and give way to better theories: “Learning and knowing them [scientific theories] is very important to improve and to get a
better theory although they change. Without old theories, we cannot achieve a better theory or improve the wrong theory (5-90).” Did this mean, then, that students saw “better/more accurate/specific/improvement” as “approaching to absolute truth”? (the close-ended question, TEN3) What were the opinions of students who held to the idea of error-free scientific knowledge? Information in Table 6.2 could provide partial answers to these questions. Out of 299 students who believed that scientific knowledge changes, only 49 students (16.4%) held that scientific knowledge approaches to absolute truth (responding to TEN3 with “strongly agree” or “agree”). Another 211 students (70.6%) felt that scientific knowledge does not approach absolute truth. Out of 53 students who claimed scientific knowledge does not change, 23 students (43%) thought such knowledge approaches absolute truth, while 15 students (28.3%) did not agree with this idea. The parallel concepts of error-free scientific knowledge and scientific knowledge approaching absolute truth had similar support. Of 107 students who believed in ‘error-free’ scientific theory, 46 students (43%) regarded scientific knowledge as approaching absolute truth. Of 316 students who admitted the presence of error in scientific knowledge, only 49 students (15.5%) thought scientific knowledge approaches absolute truth.

In short, in answering both open- and close-ended questions, most students declared that scientific theories change; that changes correct errors and provide better explanations; and that such changes do not bring scientific theories closer to absolute truth. Similarly, in answering both classes of questions, other students who believed in “error-less” scientific knowledge were more likely to think that changing scientific theories approach absolute truth. The open- and close-ended answers were therefore consistent with each other.
Table 6-2

*Scientific Theory Changes and Error Inclusion and Approach to Absolute Truths*

<table>
<thead>
<tr>
<th>TEN3</th>
<th>TEN1</th>
<th>TEN2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolute</td>
<td>no change</td>
<td>neutral</td>
<td>change</td>
</tr>
<tr>
<td>truth</td>
<td>23</td>
<td>52</td>
<td>49</td>
</tr>
<tr>
<td>neutral</td>
<td>15</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>disagree</td>
<td>15</td>
<td>91</td>
<td>211</td>
</tr>
<tr>
<td>sum</td>
<td>53</td>
<td>182</td>
<td>299</td>
</tr>
</tbody>
</table>

*1 and 2: due to the missing cases, the total number is different.

6.2.1.3 Comparisons Between Canadian and Korean Students

In giving examples for their responses, Korean students used the atomic theory most frequently while Canadian students used the particle theory most often. According to the Ontario Science Curriculum and the interviewees’ comments in this research, Grade 7 science classes covered the particle theory in fluids. The theory is also repeated in Grade 8 when the students learned about the volume and pressure of gases. On the other hand, Korean middle school science curriculum covers Dalton’s atomic theory in chemical reactions (the law of mass conservation and the law of constant composition) and the basic atomic structures in its introduction to electricity. Another example frequently cited by Korean students was the change consequent to Darwin’s evolutionary theory. The theory of evolution is actually introduced in Grade 9, but a good portion of the Korean students cited it to show how scientific knowledge has changed. It may be common knowledge not from formal science curriculum but gained from news or magazines in everyday life.

Addressing the issue of learning current scientific theories, which could change in the future, almost all students from both countries agreed that they should learn current theories better to understand nature; “Knowing that information can help us understand the world better...
(3-7).” Another noticeable point is that students admitted the knowledge they learned during science classes was the best available: “I think we bother to teach scientific theories because that’s what we know at the moment, but the theory will change later on (2-3).” Students also maintain that older ideas are a foundation for new knowledge that enrich the next generation and enable them to develop better theories.

[T]heories are crucial in providing headway to other areas and in time even they become out-dated continue to support new ideas. Science depends on heavily on evolving ideas and progress without a base, it would be utterly impossible for any form of science to exist and prove itself. (2-6)

Summarizing Tentativeness, students’ responses to the open- and close-ended questions were consistent. The majority thought that new knowledge discovered by advancing technology has corrected error and led to change in existing scientific knowledge. In marked contrast to Kuhn’s perspectives on paradigm shifts in scientific theory, students in this study acknowledged the gradual change and improvement of scientific knowledge without approaching to absolute truth.

6.2.2 Subjectivity (Theory-laden observations)

Subjectivity has become a central topic in the formation of scientific knowledge (Hansen, 1958, cited in Hickey, 2005; Kuhn, 1962). It recognizes that observations are not completely objective, but are affected by related scientific theories. In addition, when scientists analyze and interpret the data, which they have gathered, results can be biased and limited by scientists’ prior knowledge (Bell et al., 2003; Lederman et al., 2002).
**Question:** It is believed that about 65 million years ago the dinosaurs became extinct. Of the hypotheses formulated by scientists to explain the extinction, two enjoy wide support. The first, formulated by one group of scientists, suggests that a huge meteorite hit the earth 65 million years ago and led to a series of events that caused the extinction. The second hypothesis, formulated by another group of scientists, suggested massive and violent volcanic eruptions were responsible for the extinction. How these different conclusions are possible if scientist in both groups have access to and use the same set of data to derive their conclusions? (From VNOS-C)

**Intention:** To assess students’ understandings of “reasons for controversy in science when scientists use the same available data. Ideas of subjectivity, inference, creativity, social and cultural influences, and tentativeness are often elicited. The question aims to assess respondents’ beliefs about what influences data interpretation including personal preferences and bias (personal subjectivity) to differing theoretical commitments and impacts of social and cultural values” (Abd-El-Khalick, 1998, p.539)

As the developers of the question explained, this question includes theory-laden observations but also cultural influences on science, inferences and creative works of science. The advantage of this question is that it uses a concrete example to show different interpretations from different groups of scientists.

**6.2.2.1 Common Themes**

Five common themes were extracted from student responses: 1) scientists’ prior knowledge affects their interpretations, 2) when data are not clear, scientists use imagination and guesses, 3) similar effects may come from two different disasters (e.g. the extinction of dinosaurs happened from the similar effects of either volcano activity or meteorite collisions), 4) the views of scientists are wrong if contradicted by the Bible (e.g. because dinosaurs must have perished in the Great Flood, other theories on their extinction are mistaken) 5) scientists hold different interpretations when no direct observation is possible. The distribution of
student responses is represented in Table 6.3. The fourth theme was most dominant among the students from a Christian school involved in the study.

As with the analysis of *Tenativeness*, some students expressed more than one idea at the same time (e.g. themes 1 and 2), so that the total number of responses exceeded the actual number of students.

Table 6-3

*Summary of Common Themes on Subjectivity (Theory-Laden Scientific Research)*

<table>
<thead>
<tr>
<th>Subcode (reasons)</th>
<th>Canadian</th>
<th>Korean</th>
<th>(N_{Ca}=103)</th>
<th>(N_{Ko}=161)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal preference/ background</td>
<td>36</td>
<td></td>
<td>84</td>
<td>52</td>
</tr>
<tr>
<td>knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data ambiguity</td>
<td>24</td>
<td></td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>Causing effect similarity</td>
<td>16</td>
<td></td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>Biblical explanation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Bible is correct (because of The</td>
<td>16</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Great Flood, dinosaurs were extinct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>others</td>
<td>10</td>
<td></td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>No direct observation/off topic</td>
<td>10</td>
<td></td>
<td>21</td>
<td>12</td>
</tr>
</tbody>
</table>

In seeking to account for scientists’ differing conclusions from the same evidence, a majority of students (35% Canadian, 52% Korean) admitted that individual scientists’ interpretations depended on their backgrounds, levels of knowledge and personal inclinations. Students stated that scientists were human, so that they prefer their own methods of research; that even great scientists cannot know everything, and cannot interpret data beyond their knowledge; and that, for these reasons, scientists’ interpretations could differ. According to
Lederman et al. (2002), these responses are “informed views.” The following excerpt sums up these students’ opinions:

Scientists use the facts they already know and make an inference based on prior knowledge and different scientists may have different prior knowledge even if there facts on the same so their conclusion or inference may be different. (3-16)

It’s possible because scientists could add their own ideas to the data that they’ve collected. Besides, the data they’ve collected are inaccurate. Therefore, they must have put their own assumptions into their conclusions. (1-16)

I believe this is because all scientists see and think slightly differently about everything. It is almost impossible for 2 different people to draw the same conclusion from the same type of data without interacting with each other. Considering the fossilized remains do[does] not show concrete proof of either−. It doesn’t mean that it isn’t true. (3-27)

Individual scientists are different in background knowledge, and unique in their knowledge and creativity. People are inclined to interpret what they want.

A second group of students reasoned that different interpretations arose from vagueness of the data rather than from differences among scientists. In their mind, if the data were clear, then scientists would reach the same conclusions. Lacking evidence, scientists make guesses, and their prior knowledge influences the resulting interpretations. This theme accords with the views of scientists themselves examined in Wong and Hodson’s (2009) study. Nine out of thirteen scientists saw differences of interpretation as the consequence of insufficient relevant data. As an example, an experimental particle physicist commented that insufficient data can make scientists infer the value of $x$ and $y$ in the equation “$x + y = 3$” (p.13). Many different pairs of $(x, y)$ are possible if the condition is not clear. Akerson, Morrison and McDuffie (2006) also showed that quite a portion of the students in their study held this view: the more
data scientists have, the more their conclusions will converge toward agreement. The following are student responses:

Different results/interpretation was not due to scientists but to the information. The information scientists have allowed to many interpretations. (4-9)

Perhaps the data that both they have is not sufficient enough to draw a definite conclusion. Then they could interpret what happened based on their opinion of what disaster was more probable which is how there are two opposing theories. (2-12)

The third group of responses was classified as holding the “objectivity of scientific observation” but their point of view was different from the second group of students. They tried to analyze the reason of the extinction of dinosaurs by themselves and then concluded that scientists reached different conclusion due to the similarities between possible causes (eruption of volcanoes and collisions of meteorites). A few of them were very knowledgeable and expressed their ideas logically. Rather than simply classifying them as a naïve view, they were very analytic and tried to understand both scientists’ arguments.

Two theories are possible because an abnormally large amount of iridium was found in the strata of the end of Cretaceous period, and there was also evidence of active volcanic explosions during the same period. The two different theories were compatible because two natural events coincidently occurred neither because scientists’ background knowledge was different nor because the data were not clear. 두 가지 가설 모두 가능하다. 백악기 말에 발견되는 지층에 다량의 이리듐이 발견되었고, 또 활발한 화산 활동이 있었다는 증거가 많이 발견되었기 때문이다. 이는 증거가 불충분하다거나 과학자들의 개인적인 생각이 달라서가 아니라 우연하게도 비슷한 시기에 두 가지 사건이 발생했기 때문이다. (5-11)

I think the scientist drew two different conclusions because the information and clues they might have gotten be very similar. …the bottom line is, scientists can get different ideas on one thing that happened but there is usually some common ground or a similarity with both ideas. (3-23)

---

10 During semi-structured interview, the participants’ ideas were proved.
We can think about this in two ways. First, two things might happen at the same time; meteorites can collide with the earth when a volcano erupted. So then, the trace elements can be discovered along with evidence of a volcano eruption. Second, these two cases may closely be related to each other. The tremendous power of the collision may have led to the volcano eruption where the collision occurred. The plate of earth became weak and the magma nearby exploded on a large scale. The shockwave of the collision may have moved around the surface of the Earth, and caused sequential reactions.

A few students reasoned that conclusions differed because no direct observation was possible in the case:

Firstly scientists were not there, when the dinosaur were extinct so their information must come from background information, evidence and personal opinions. … Therefore, groups of scientists may believe things that support their perspectives along with the limited evidence they have. (3-3)

Because none of the scientists were there to see what actually happen so based on there scientists had a different image to what has background knowledge they do a prediction to what has happened maybe on group of happened to another group of scientists. (3-21)

Of course, different scientists have different knowledge and how much a scientist knows is also different. However, in this case, scientists could not directly perform an experiment, and nobody knows what happened 65million years ago. So they heavily relied on guesswork. Due to vague data, different theories are possible. (7-35)

Sixteen students out of 53 from School 1, a Christian school, disagreed with scientists’ interpretations and stated that God brought about the extinction of dinosaurs. This Bible-based view was not found in the other schools.
My personal opinion is that God, destroyed all the dinosaurs. I’m not sure how but I think both the groups of scientist are wrong. In fact the world isn’t even 65 million years old. It’s only a couple thousands. That is just my scientific opinion. Well to answer your question I think that through the scientists are working with the some data they could make mistakes. They’re only human. (1-17)

Neither are [is] possible according the word of God. The world was created by him and will be destroyed by him thinking that the world and all its inhabitants were created by either of these is senseless. (1-51)

6.2.2.2 Comparison between Close- and Open-Ended Answers

The close-ended questions asked whether scientists’ prior knowledge affects their observations, and whether the results of different scientists’ observations would not be the same. The open-ended question on Subjectivity probed for aspects of the issue different from those posed in the close-ended questions. It focused on scientific interpretations rather than scientific observations. No dominant answering pattern was found in the answers to the three close-ended questions posed (Figure 5.13 (b)). Responses were evenly distributed. In SUB1, 39% of student answers agreed that scientists can interpret experimental data differently but 40% disagreed. As to scientific observation, 46% of the students disagreed that scientists’ background knowledge affects their observations, whereas 33% agreed that their observations are influenced by their education. More students (45.5%) stated that observations are not opinions, while 31% said they are. In general, close-ended responses indicated that more students seemed to hold objective views on scientific observations. This is reflected in the student excerpt cited below which asserts that observations are facts but interpretations are opinions.

Scientists use the facts [the gather data] they already know and make an inference based on prior knowledge and different scientists may have different prior knowledge
even if there facts [the data] on the same so their conclusion or inference may be different. (3-16)

Comparing these results with the summary of open-ended answers in Table 6.3, it is clear that responses to open- and closed-ended questions matched approximately but not exactly. Subjectivity traces the influences on individual scientists, which create differences in scientific research, examining the effects of their cultural, religious, social and political background. Open-ended questions on Sociocultural Embeddedness explored students’ attitudes to this area. Interestingly, most students admitted that scientific research has been affected by differences in individual scientists. The gap between open- and close-ended responses on Subjectivity probably arose from their differences of focus. The close-ended questions dealt more with scientific observation, while the open-ended questions emphasized scientific interpretation. Another possible factor could be the misclassification of open-ended responses arising from their ambiguity. A considerable number of responses ran: “Different scientists can interpret differently.” With no indication as to why scientists interpret differently, these responses were counted into the first of the three groups described here.

6.2.2.3 Comparisons between Canadian and Korean Students

On theory-laden observation and interpretation, Canadian and Korean students have similar opinions. However, a discrepancy was found between quantitative and qualitative data. In the quantitative results, the Canadian students were likely to hold a more relativist position than did the Korean students. However, in the qualitative data, the proportion was reversed; that is, 35% of the Canadian students described subjectivity of data interpretation while 52% of the Korean students did. This reversed result between open- and close-ended responses may arise from the Biblical interpretation of some Canadian students. About 16% of Canadian students offered a Biblical explanation for the extinction of dinosaurs while no Korean
students did. It is possible that some participating students from public schools in Canada and Korea might hold evangelical Christian beliefs, but no student from these settings voiced ideas of creationism or intelligent design. Faced with these unique responses, semi-structured interviews were used in an attempt further to clarify this issue.

6.2.2.4. Different Views of the Same Example

In discussing Tentativeness, two students cited the history of the atomic theory as an example to support their opposing ideas. Student 2-13 used the atomic theory as the evidence that major theories in science do not change while Student 5-5 used it as the evidence that such theories do change. Again, in considering Subjectivity, Student 2-13 student said that different scientists interpret the same data differently because of various outside influences, besides the lack of direct observation:

...because there could have been many factors that influence the data. Based on the circumstances surrounding it could be quite possible, for there to be more than one explanation for the facts. Also, and quite important to remember none of the scientists were there at the time to determine exactly how these common facts arose, and despite facts people are entitled to their opinion and people will be sure if they defend them well. (2-13)

On the contrary Student 5-5, saw differences in the individual scientists themselves as the reason for their differences of interpretation.

I think different interpretations were rooted in the different ways people approach a problem. When scientists try to solve a problem, they set a hypothesis and then they examine and find causes that can match the hypothesis. Particularly, before setting up a hypothesis, people's ideas are different from each other. However, it is not only that their ideas are different; there is also the possibility that people can reach different conclusions from the same evidence.
These two cases imply that neither extreme realist views nor extreme relativist views exist in actual science. How one views scientific theories and the relationships between the theories and evidence closely relates to one’s own perspective. In the philosophy of science, empirical positivists interpreted Mendel’s law of genetics as an inductive outcome while deductivists interpreted it as a deduction. In the case of the two students in this research, Student 2-13 focused on the major concepts of atoms. In looking at the practice of science, the student put weight upon scientific methods of direct observations and experimental data. In contrast, Student 5-5 focused on the changes of atomic structure from its most primitive to its most recent model; in considering the practice of science, his thought centered on scientists as a human factor, not on objective scientific methods.

To summarize responses on Subjectivity (theory-laden scientific observation and interpretation), a majority of the students thought that the prior knowledge of individual scientists affected them both in observation (close-ended reposes) and interpretation (open-ended responses). However, a considerable proportion of students also regarded observations as objective facts rather than subjective data in the close-ended questions. A second major response was that the lack of relevant data and ambiguity in the data prompted scientists to guesswork, involving the influence of their background knowledge and biases. Unique answers came from students in Christian schools who used their beliefs about the Great Flood in the Bible to assess scientific discussions on the extinction of dinosaurs.
6.2.3 Empirical Evidence Based Scientific Knowledge

As reviewed and defined in Chapter 2, Empirical Evidence Based Scientific Knowledge (EMP) means that a suggested scientific theory should be testable; the evidence advanced for it should be observed and replicable by other scientists. According to Bell (2009) the term empirical refers to “both quantitative and qualitative data. …. ultimately, all scientific ideas must conform to observational or experimental data to be considered valid” (p.3). Two open-ended questions, one asking for a definition of science, and another, asking what specific features in science distinguish it from other subjects, directly or indirectly touch on the characteristics of empirical evidence based scientific knowledge.

Question 3. A. What is science and what makes science different from other subjects (mathematics, literature, social studies, history etc.?)

Intention: “This question aims to assess respondents’ views regarding science as a discipline to address questions about the natural world, the role of science in providing explanations for natural phenomena, and the role that empirical evidence plays in science that separates science from other “ways of knowing.” Responses to this question often reveal a common misconception regarding the use of the “Scientific Method” as an objective process by which the knowledge is discovered. Such a view is often presented as an explanation for how science differs from other disciplines of inquiry.” (Abd-El-Khalick, 1998, p.532)

Question 3. B. Science textbooks often represent the atom as a central nucleus composed of protons (positively charged particles) and neutrons (neutral particles) with electrons (negatively charged particles) orbiting that nucleus. How certain are scientists about the structure of the atom? What specific evidence, or types of evidence, do you think scientists used to determine what an atom looks like?

Intention: “This question aims to assess understandings of the role of human inference and creativity in science, the role of models in science, and the notion that scientific models are not copies of reality.” (Abd-El-Khalick, 1998, p.536)
6.2.3.1 Common Themes in the Definition of Science

Because the question concerns a general definition of science, students’ responses did not converge into empirical evidence based scientific knowledge. Among their diverse ideas of science, student answers corresponding to the concepts of EMP were extracted as they compared and contrasted science with other subjects: 1) science studies the physical world, so that concrete evidence is required, 2) ideas must be tested and confirmed by experiment; and 3) science changes continuously in response to new evidence. Other unique answers were given.

Applying the standards of the conceptually clustered matrix (Figure 4-9) to ascertain whether responses were “informed” or “naïve”, the common themes extracted were found to be naïve; most answers spoke of the discovery of the physical world (the body, the earth, or the universe). Concerning science experiments, Carey and Smith (1993, cited in Sandoval & Morrison, 2003) say, “[Experiment is seen] as distinct [from ideas], and the purpose of science is seen as testing ideas, but [a naïve view] does not yet necessarily recognize that such ideas are tentative or that theories are socially constructed and constrain the kinds of experimentation scientists might do.”

Many responses included more than one element (sub-codeable words), so the percentages in Table 6.4 were the results of double or triple counts if a response had more than one element. Some answers were not related to empirical evidence based nature of science. Therefore Table 6.4 presents only a rough distribution of students’ views.
Table 6-4

Summary of Common Themes on Empirical Evidence Based Scientific Knowledge

<table>
<thead>
<tr>
<th>Code</th>
<th>Sub-code (reasons)</th>
<th>Canadian</th>
<th></th>
<th>Korean</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>Characteristics of science</td>
<td>Continuous changing due to new evidence (discoveries) and trial-and-errors (seeking for evidence)</td>
<td>31</td>
<td>30</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Objective, logical subject</td>
<td>7</td>
<td>7</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Working on physical world</td>
<td>35</td>
<td>34</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Experiments (verifying theories)</td>
<td>22</td>
<td>21</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>No response</td>
<td>6</td>
<td>6</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>Other responses</td>
<td>Benefits for human life, Technological advancement, etc.</td>
<td>19</td>
<td>18</td>
<td>21</td>
<td>13</td>
</tr>
</tbody>
</table>

When asked in which domains science works, most students answered the material world (e.g. the universe, earth, the body) as opposed to spiritual, societal, political or psychological domains. They see the methods to study the physical world as observation and experiment: “It [science] provokes questions on many aspects of our life that other subjects don’t explain or prove” (2-6); “Science is a subject of proofs and conclusions which seeks to find the causes of natural phenomena in the world and universe, and through its findings, to explain them to others.” (6-17) (과학은 우주 지구상의 모든 현상들에 대하여 그 현상의 원인을 규명하고 그를 통해 다른 현상까지 설명하는 증명과 결론의 과목이라고 한다.)

Students mentioned that science is unique as knowledge, which is tested and proven by experiment. Experiment and observation verify scientific theories, and without the process of empirical proof, scientific knowledge cannot be accepted: “Science is based on theories and
methods; they require proving what is real by conducting experiments” (2-43); “An idea is just what you believe until proven to be true. Scientific experiments help you understand whether a hypothesis is true. You can’t say something is true before it’s tested” (3-51). Korean students said, “Scientific knowledge is the outcome of an experiment and its causes and effects are revealed by experiments” (과학적 지식은 실험의 결과물로서 원인과 결과가 실험에 의해 밝혀진 것이다 (6-15); “Science is a set of objective facts from accurate results of experiments, and the results are truths.” (6-47) 과학은 정확한 실험 결과에서 나온 사실이고 진리이다. As these excerpts show, students regarded science as the result of experiments.

This is mainly due to the fact that when a piece of scientific knowledge is presented you will have to use a form of experiment, one way or another, to prove that knowledge correct. After all if the scientific knowledge cannot be applied upon a real world application, it shouldn’t be scientific knowledge in the first place. This is because science is the subject of learning how they ways things work in our world, and what laws governing it. (3-46)

Without experiments, how can you observe different theories and see the results and come up with a conclusion. Without performing an experiment, there’s no way to tell whether your theory is correct or not. For example; if you wanted to find out whether animals liked a certain kind of cheese you would need to have gotten some rats and different cheese and performed an experiment in order to see whether your theory was right. (3-52)

In accordance with the tentative nature of scientific knowledge, some students differentiated science from other subjects by pointing out how new empirical evidence can lead science to change. The continuous change of scientific knowledge implies that knowledge supported by empirical evidence is not absolute truth; there is always the possibility that it may be modified or disproven through further experiment.

Science is always on the move, more discoveries are made everyday, old theories are proven wrong, and new issues pop up, science is always updated by new experimental results or discoveries. It changes with the era and the people
in that era, unlike religion or history or many other disciplines of inquiry, which will never change. (4-40)

Science is an unstable subject to compare with other subjects. Scientific theories, whether biology or astronomy, always can be changed by adding discoveries or by removing existing theories due to new experimental results or discoveries. …Scientists can make hypotheses and the results of experiments can support the hypotheses, then other scientists will accept the hypotheses as theories. So, science can change with high probability through new results of scientific experiment.

6.2.3.2 Common Themes on Atomic Models

Students’ viewpoints on the basis for atomic models were relatively simple. Two distinct categories emerged: 1) the empirical outcome of direct observations or experiments and 2) the agreement among scientists based on empirical evidence. Table 6.5 indicates students’ responses.

Table 6-5

Summary of Common Themes on Empirical Evidence Based Scientific Knowledge (Atomic Model)

<table>
<thead>
<tr>
<th>Code</th>
<th>Sub-code</th>
<th>Canadian</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(reasons)</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>Atomic</td>
<td>The outcome of direct observation or experiments</td>
<td>55</td>
<td>48</td>
</tr>
<tr>
<td>Model</td>
<td>The outcome of scientists’ inferences</td>
<td>37</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Others (imagination, the solar system, gravitation)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>No response</td>
<td>18</td>
<td>16</td>
</tr>
</tbody>
</table>

N_Ca=114  N_Ko=158
In explaining the development of atomic structure models, many students answered that direct observations using an electronic microscope enable scientists to make models, rather than creative ideas or inferences; “Scientists use special microscopic tools to determine the look of an atom” (3-7); “When materials were magnified, scientists could observe the shapes. The basic shapes were similar to every material (확대하니 그형태는 물질마다 기본적인 형식은 같아서 (6-57))” Also, students answered that the model is scientific and real since it is supported by experiment’ if it were not, it would not have the ability to explain natural phenomena.

Scientists are certain in atomic structure because they have researched about it for many years, the technology of microscope has really advanced so they are more sure about the structure of an atom. I think scientist used very high magnifying microscopes to see what an atom looks like. (3-19)

Students recognized that indirect evidence from other experimental results could also be admissible evidence. For instance, the cathode ray experiment and Rutherford’s experiment do not directly picture atomic structure; rather, they provided scientists with clues to the structure.

In the excerpts below, some students point out how in the history of developing atomic structure various experiments provided empirical evidence.

Scientists cannot confirm the atomic structure because they have not directly observed it. However, several experiments such as a cathode ray experiment support the structure, so it could be real. 과학자들도 원자 구조를 직접 관찰한 것이 아니기 때문에 원자의 구조에 대해 확신할 수 없다. 그러나, 음극선 실험등 여러 실험을 거치면서 원자의 구조에 대한 추측을 뒷받침 하는 것이다. (5-51)

At the beginning, scientists did not know anything about atoms’ structure, but they tried to figure out how an atom looked. They performed experiments for hundreds of years. One of the experiments sent an X-ray into silver foil. The results of the experiment offered a clue for scientists to think about a hard nucleus in the middle of an atom and large empty space is around the nucleus. (처음에는 아무것도 몰랐다가 나중에 과학자들은 핵을 구조에 대해 생각하였으며, 그
A shift in students’ views was noticed. In *Subjectivity*, few students mentioned “no direct observation” but a dominant number of students mentioned “direct observations of atomic models.” It could be hasty to conclude that the students’ views were not solid. Rather, the interpretation reflects a commonsense understanding of the science textbooks. It is obvious that scientists were not on the scene when dinosaurs were extinct, so such a view that “seeing is believing is less likely to appear in students’ responses when science textbooks provide photos of atoms taken by electron-microscopes. The responses are evidence that students’ learning knowledge of science affects their views on knowledge about science, rather than that students do not have consolidated ideas.

6.2.3.3 Clarification and Consistency with the Close-Ended Answers

The major themes of the open-ended responses were the continuous changes in science, the domain of science in the physical world, and the necessity of experiment and repeatable evidence as the method and proof of scientific study. Responses discussing the development of atomic models also stressed that plausible observation as well as indirect evidence from other experimental results were admissible evidence for scientific theory. On the other hand, the three close-ended answers had a relatively lower mean \( M_{EMP} = 2.57 \) than other questions on NOS concepts. Particularly, 62% of the students agreed that scientific knowledge needed experimental testing \( (EMP3) \). The distributions were positively skewed, which meant that most students agreed that scientific evidence should be observable, repeatable and consistent. Therefore, the close-ended and open-ended responses were congruent.
However, an in-depth review of student opinion showed that most answers emphasized direct observation experiments to prove theories, and objective logic, reflecting naïve views of science (Lederman et al., 2002; Liang et al., 2008). The close-ended responses revealed that students seemed well aware of empirical evidence based scientific knowledge, but saw evidence as determined by direct observation rather than by theory-drive assumptions.

6.2.3.4 Comparisons between Canadian and Korean Students

NOS studies are concerned lest students think scientific knowledge does not change across history. Interestingly, a good portion of students in this study expressed the opposite view: scientific knowledge is characteristically alterable. Canadian (38%) and Korean students (19%) asserted that the domain of science is discovery, and therefore it has never been static, but always dynamic. And this view was consistent with the responses to Tentativeness. More Canadian students confined scientific knowledge to the physical world (30% Canadian versus 16% Korean) while more Korean students viewed science as an exercise in objective logic (17% Korean versus 7% Canadian).

Although the developmental history of atomic theories was one of the frequent examples mentioned in the open-ended answers, unfortunately, many Canadian (16%) and Korean (27%) students did not answer this question at all. The Ontario Science Curriculum (2007) shows that Grade 6 students could learn a basic atomic model as they study the generation of electricity. They could also learn a little more about atomic models during their lessons on particle theories. Detailed concepts of atomic structures are introduced at Grade 9 in the unit on chemistry. During her semi-structured interviews, “Teresa” remarked that she learned about the atom in Grade 6, but not about its structures. Korean students are given opportunity to learn the basic concepts and a rough structure of the atom in the Grade 6 unit on
matter, and in a unit on electricity at the end of Grade 8. School 7 gathered its data in June (the middle the school year) while other schools gathered their data in December (the end of the school year). Taking the school year into account, then, students from School 7 had not yet studied atomic structure.

To recapitulate this section, student responses defined the characteristics of science by differentiating it from other subjects, and by discussing the development of atomic models. Overall, students thought science was distinct because it studies the physical world rather than social or psychological domains, and that it requires plausible proofs. Many students believed that evidence is necessary for a scientific theory to be accepted and some argued that a proven theory is absolutely true. Many of the responses to these two questions revealed students’ naïve understanding of science. To their mind, empirical evidence is objective, absolute proof; the models of atomic structures came from directly observed real atoms.

### 6.2.4 Social and Cultural Influences on Science

This concept examines whether students view the scientific enterprise as non-cultural and non-historical, or as multicultural and history-embedded. Historians in the philosophy of science such as Hickey (2005) have agreed on distinguishing the context of discovery from the context of judgment, and some science educators acknowledge this view (Chalmers, 1999; Godfrey-Smith, 2003; Matthews, 1994). Thus, students’ responses about the acultural or multicultural characteristics of science would be classified either in a discovery context or in a judgment context. In addition, as with responses on theory-laden observations (Section 6.2.2), the analytic focus of the responses paid attention to scientists as individuals and to science as a human enterprise embedded in society.
**Question:** Some claim that science is infused with social and cultural values. That is science reflects the social and political values, philosophical assumptions and intellectual norms of the culture in which it is practiced. Others claim that science is universal. That is science transcends national and cultural boundaries and is not affected by social, political, and philosophical values, and intellectual norms of the culture in which it is practiced.

If you believe that science reflected social and cultural values, explain the reasons with examples.

If you believe that science is universal, explain the reasons with examples. (from VNOS-C question 10)

**Intention:** The intention of this question is to examine participants’ views of the influence of social and cultural values and expectations on the scientific endeavor. Naïve views are “Science is about the facts and could not be influenced by cultures and society. … scientific knowledge is universal and does not change from one place to another.” On the contrary the informed view is “All factors in society and the culture influence the acceptance of scientific ideas” (Lederman et al., 2002, p.516)

---

**6.2.4.1 Common Themes**

Most students considered either the discovery or the judgment context, but not both, in their responses. Students who adopted Multicultural views of science focused on the discovery context of science. Under this Multicultural view, four subcategories were extracted: 1) external pressure and motivations which determine what to study, 2) economic restrictions to scientific research, 3) scientists’ place as members of society and 4) the impact of science outcomes on society. In contrast, students who rejected the idea of cultural, political and societal influences upon science focused on the judgmental context of scientific knowledge. The subcategories of this Universalist view were 1) the independence of nature from human cultures and 2) the objectivity of scientific findings, which are consistent across cultures, and superior to political ideology and religious belief. To support the former idea, students cited the world-wide work of stem cell research. To support the latter idea, they pointed to Galileo’s
conflict with the Roman Catholic Church over heliocentrism, and the controversy between Darwin’s evolutionary theory and Creationism or Intelligent Design theory. A third, minority view, suggested a gap between “ideal science” and “real science”. “Ideal science” is universal, while “real science” is multicultural. Table 6.6 indicates these major themes with a general distribution of students’ responses.

Table 6-6

**Summary of Common Themes on Cultural, Political and Societal Embeddedness of Scientific Research**

<table>
<thead>
<tr>
<th>Sub code (Description of Responses)</th>
<th>Canadian</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>No. of response</strong></td>
<td><strong>%</strong></td>
</tr>
<tr>
<td>What to study (External pressure)</td>
<td>29</td>
<td>25.4</td>
</tr>
<tr>
<td>Society (laws), culture (ethics) and politics demand scientists doing something for their benefits or scientists were motivated by societal demands</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>Societal economic supports for research have influenced science.</td>
<td>13</td>
<td>11.4</td>
</tr>
<tr>
<td>(Internal factors)</td>
<td>4</td>
<td>3.5</td>
</tr>
<tr>
<td>Scientists are members of a society, so they cannot avoid the influence (education and ways of thinking)</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>Science has been playing a role of societal development.</td>
<td>9</td>
<td>7.9</td>
</tr>
<tr>
<td>No reasons</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Yes | Ideality and Reality Gaps | 9 | 7.9 | 4 | 2.5 |
| Ideal science is not affected the outside factors but real science is affected political, cultural and societal values | 2 | 1.8 | 11 | 7.0 |
| Yes: What to study and How to study/ No: the results of science (universal) | | | | |
Characteristics of Subject areas
Scientific knowledge, and research is dealing with our nature (physical worlds), scientific truth cannot be affected by culture or political reasons.

<table>
<thead>
<tr>
<th>Response</th>
<th>No</th>
<th>Science is completely objective. (Scientists were educated not to be bias)</th>
<th>No reasons</th>
<th>No responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>18.4</td>
<td>15</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>7.9</td>
<td>11</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.6</td>
<td>5</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>13.2</td>
<td>25</td>
<td>15.8</td>
</tr>
</tbody>
</table>

**Multicultural Views:** Most responses expressed a multicultural view of science. Society, religion and culture restrict and direct scientists in considering what and what not to study, as well as what and what not to disclose to the public for the good of society:

A situation where a scientist lives decides what to study; for instance, during World War I and II, almost all scientific research was involved in developing arms. 과학자가 어떤 상황에 처해 있는지에 따라 과학 연구의 대상 자체가 바뀌게 된다. 예를 들어 세계 대전 때는 정말 많은 연구가 전쟁을 중심으로 돌아갔다 (6-13).

Scientific research and scientists were influenced by society, religion and politics, particularly political power. For instance, the politics of North Korea is focused on preparing for a war against its enemy. So, politicians force scientists to invent new arms. 북한을 예로 볼 때 그들 적대국과의 전쟁을 위해 군사위주의 정책을 하고 있고 또 그로 인해 무기 연구 개발 등의 활동에서 많은 과학자들이 일하고 있다. (6-23)

I think that it [science] is influenced because sometimes a discovery might cause hysteria among the public, so the political government would try not to let it leaked [leak] out to the public to keep them calm. Also some cultures might refuse a finding because if goes against their religion or value. (4-43)

A few students also mentioned that research funding would control what scientists should study, as well as what results they could make known to the public:

Those people who recognize the importance of scientific research support scientists. Some of them want scientists to work a specific area in which they have interests. Otherwise, scientists should buy equipment and spend their
money to continue research by themselves. It can be a great burden for scientists, so science is affected by society. 과학을 인정하는 사람들은 과학 발전에 기여를 하기 위해 여러가지 지원을 해 주고 또 돈을 대 주면서 자기가 원하는 분야에 연구하기를 바라기도 한다. 그렇지 않으면 연구자 스스로가 비용과 기구를 사들여야해서 부담이 크다. 그러므로, 과학은 사회에 영향을 받다고 할 수 있다. (7-14)

Not only science but also every human activity is affected by the principles of politics, culture, and society. Many scientific observations and conclusions great scientists made faced opposition from existing misconceptions and religious ideas. For instance, it took a long time for heliocentric idea to be accepted due to the opposition of religion. 과학뿐만 아니라 인간의 모든 활동은 정치적, 문화적, 사회적 원리에 영향을 받는다. 위대한 과학자들에 의해 이루어진 많은 과학적 관찰과 결과들은 기존의 종교적인 잘못된 관념들과 맞부딪히게 된다. 예를 들자면, 지동설이 받아들여지는 데는 천동설을 믿는 종교적 반대에 부딪혀서 받아들여지는데 꽤 오랜 시간이 걸렸다. (6-5)

Because scientists are also human, they cannot avoid influences from the culture in which they were born and educated, and from the situation of their society:

Scientists are obvious raised from a child to who they are today. This means she or he may be raised most likely to the ideas of their family. (2-41)

I believe that scientists can be heavily impacted and influenced by social/cultural values. For instance, when it comes to the topic of stem cell research it caused much controversy throughout the scientific world. Scientists were questioning themselves things such as morals whether it was right or not to use embryos began to intervene, “human life should not be toyed with” many religious leaders would say stem cell research has banned in several countries due to all the controversy. So, as you can see science can be heavily impacted by social, political and cultural values. (Quotation original) (3-46)

Scientists are human, after all. If you would say that science is discovering knowledge then it is affected. We wouldn’t use experiments and such to find out things of no importance and this is because of all the values. (3-62)

A few students who granted societal and cultural influences on science, pointed out that a society must mature cognitively as a necessary prior condition to accept some theories as authentic. In other words, the values of a society and its degree of maturity are a standard for
acceptance, rather than mere “truth”. Students commonly mentioned how the public responded when Galileo advanced his heliocentric view of earth or when Darwin published his theory of evolution. Lederman et al., (2002) and Liang et al., (2009) classified the view that “all factors in society and the culture influence the acceptance of scientific ideas” as an “informed view”

Science is fully influenced by the values of society, culture and religion. When Darwin published his idea of evolution, the public looked down the idea and some caricatured him as a monkey. … In other words, the society was not ready to accept the theory. For a scientific theory to be accepted, society should mature to accept the theory.

The reason why science has been affected by the societal and religious values is that as time goes on, people’s ways of thinking and people’s standards also change. So people will view the same natural phenomena with the changed standards, and they may accept a scientific theory that their precedents had not accepted. Heliocentric idea is a good example. Of course, there was a religious factor not to accept the idea, but also the public could not accept the idea with their natural senses.

According to the responses of students of multiculturalist views, science has positively influenced society whereas culture, politics and religion have negatively affected science. So, the citations above mention religious opposition to Galileo or Darwin, and the political direction of scientific work in North Korean, while science is credited with providing society cures for disease, advanced technology, and deeper insights into the natural world.

[S]cience might tend to lean towards appealing to the public so as to be accepted socially, for example, there is a science developed to improve our performance
(aeromechanics). This leans towards the public because they would want to buy a car with advancement in technology new additions. (3-43)

**Universal Views of Science:** On the opposite side of the issue, some 24% of students in the study held to the independence of science from cultural, societal and political values. Their reasons were 1) that science deals simply with the physical world, which is less likely to change, 2) that scientific knowledge consists of experimentally proven facts, 3) that scientists work objectively not subjectively, and 4) that the results of science are truth, so even though external factors may try to hide or distort scientific outcomes, science wins out in the end.

Lederman et al. (2002) and Liang et al. (2009) considered these views naïve. Of course, some of these ideas could be classified as a naïve; for example, it is surely naïve to say that science is truth and cannot be different across cultures or societies (as said Student 2-29 below). However, most students described extant scientific knowledge in a *judgment* context rather than in a *discovery* context. This could be quite acceptable as an “informed view.” Even Feyerabend (1975) who has been recognized as a methodological anarchist admitted the rationality of the judgment context. Thus, depending on an individual student’s explanation, the group as a whole cannot be simply judged as holding to naïve views.

I somewhat discourage with the educators claiming that science is influenced by social, political and cultural values because many cultures have many different views and takes on things. However, science is based on facts, and cannot be biased, so although people may have opinions on things, science is based on the facts and they cannot be changed. (2-29)

Science itself cannot be influenced by anything. Science itself is independent. It’s always there…. Science is everywhere from a dropping apple to $E=MC^2$, from a bathtub to the chemical element chart. It’s all over our basic lives. (3-40)

No matter what culture you are from or what values you have, $H_2O$ will still be the formula for water. H may be written in different languages, but they all mean the same. (2-31)
Science is universal and is not affected by social cultural and political influence because truth does not change whatever the surroundings and conditions have changed. Of course, long time ago, due to political and societal reasons, such a wrong theory as the Phlogiston theory was admitted as a scientific theory. Definitely, science was affected by a society. As time passed, however, that theory was proven to be wrong and the truth [oxygen] was accepted. Such cases are very rare through the whole history of science; people cannot change the truth.

Religion and Science- Christian School Students Versus Public School Students: Many students had the impression that when a society or a culture gets involved with science, it creates obstacles to its development rather than supporting its progress. As one of the schools in this research was a religion-based school, its students’ responses on the scientific discussion of reasons for the extinction of dinosaurs were markedly different from those of public school students. About 30% of the Christian school students preferred the religious view that God created the universe to any scientific explanation. However, very few public school students from either Canada or Korea compared the doctrine of divine creation with scientific ideas; most students plainly gave priority to scientific explanations. However, students from School 1 were more likely to admit socio-cultural influences on science than were the students from the public schools. The following excerpts were students from School1 and other schools.

I do not believe scientists conduct research without bias. Scientists can be biased somewhat by their social, political and cultural values, but not so much as religious beliefs would affect them. If I were to conduct research about CREATION VS. EVOLUTION, I would be biased towards creation because of my religious beliefs because I am a Christian. In the same way scientists can be biased towards a certain result (capital letters and underline original). (1-23)
I agree because if you are a religious scientist that could affect the way other scientists think of. As example is how creationist/Christian scientists disagree on evolutionist/atheist on how we were made and about the earth. I would prove as a Christian that other scientists that think that should actually give accurate detail instead of just saying false things. (1-13)

I believed that science isn’t influenced by social, cultural and political values. The way science is portrayed may be affected by the above influences. A great example of this is creation of the universe. The scientific knowledge is the Big Bang theory. The way it is portrayed depends on the religion. Some portray the science of the big bang as chance; others portray it as the creation of god. This should that the science isn’t affected, but the way it is portrayed is affected. (3-59)

I do not agree with the statement that says science is influenced by social, political, cultural values being science is what you see in front of you. Science is the study of many things, and it’s based on hard evidence, experiments etc. Scientific ideas are always proven right or wrong where as god’s creature is what people choose to say exists or doesn’t exist, but there is no proof so you can’t let beliefs that may or may not be correct, decide whether facts and what you see in front of you, is science (a scientific method, new form of life, etc.). (3-51)

Science is independent of society, culture and politics. … An extreme example is Creationism of Christian imagine that humans are creatures of God, but actually humans are the results of evolution from an Australopithecus that resembles monkeys. Science is not the outcome of such creative thinking. Science should be based on proven evidence. … 과학이 사회 문화 정치적 영향을 받지 않고 보편적인 것이라고 생각한다. 극단적인 예를 들어 가톨릭에서 인간은 하느님의 창조물이라고 하지만 실제로는 원숭이의 모습을 한 오스트랄로피테쿠스가 진화해서 지금의 인간이 된 것이다. 단지 인간의 상상력과 창의적으로 과학을 만들어서는 안된다. … 과학은 정확한 근거와 자료가 필요하다고 생각한다. (6,30)

Instead of simply judging this view as naïve or informed, it would be more insightful to focus on the reasons why students adopt a universal and cross-cultural view of science. Most of the knowledge they learned in science classes does not differ from culture to culture.

Student 2-31 mentioned that water is H₂O, and a culture may give it hydrogen with a different name, the composition of water itself does not change from culture to culture. In accepting or
rejecting a scientific theory, there is no room for the influence of cultural values. In a way, these students were actually well trained by a current educational system.

**Both Multicultural and Universal Views**: Some of the students (5%) admitted that some areas of science have been affected by the societal and cultural influence, while other areas of science are not affected. A few expressed gaps of ideal science and real science.

I believe some aspects of science are influenced by social/cultural values, and some sub-studies of science is [are] not influenced. For example, the sciences that are about physical data, and our physics should not be influenced by values; biology, chemistry should not be affected. But science that involves the actions of people and their reactions must be influenced by values if not, everyone would be easy to solve as 2+2, but we are not. (2-38)

I think science should not be affected by these values but they are. I agree that science is influenced by social, political and cultural values. Scientists are humans, after all. If you would say that science is discovering knowledge then it is affected. We wouldn’t use experiments and such to find out things of no importance and this is because of all the values. (3-62)

Very rare cases considered both the discovery context and the judgment context.

Science is universal, which is not affected by social and cultural values. Merely, the values motivate scientists what to do. For instance, presently people are threatened by abnormal climate and by global warming. Scientists are motivated by the social situations, but the results cannot be affected by the situations. 과학적인 연구는 사회, 문화적 가치에 영향을 받지 않는 보편적인 것이다. 다만, 사회문화와 과학적 연구에 그 필요성에 따른 동기부여일 뿐이다. 예를 들자면, 현재 지구는 이상기후와 과거에는 없었던 각종 이상 재해들로부터 위협 받고 있다. 이러한 사회적 상황은 연구를 하자는 동기부여가 되지만 그 과학적인 연구의 결과에는 영향을 미치지 못한다. (6-5)

**6.2.4.2 Consistency between Responses to Open-Ended and Close-Ended Questions**

The close-ended questions *CUL1* and *CUL2* asked whether scientists’ work could be affected by their society and culture; *CUL3* asked whether cultural and societal values were decisive in accepting or rejecting scientific knowledge. About 50% of students agreed that
society and religion affect how scientists work and what scientists work on in CUL1 and CUL2. On the other hand, 40% of students said that societal and cultural values do not decide what scientific knowledge can be accepted (CUL3). Unfortunately, a majority of students chose neutral. It was not clear from these neutral answers whether the students did not have any idea of the cultural aspects of science or whether they had really neutral views.

Direct one-to-one comparisons between open- and close-ended answers were difficult. Since many chose neutral in the close-ended questions, and left the question unanswered, no reason could be ascertained. Double counted cases in the open-ended questions were frequent. However, once neutral answers were excluded from the close-ended answers, and unanswered/vague cases of open-ended responses were discarded from the open-ended answers, a rough comparison was made, and the student answers became very consistent. Further, their views were compatible with those espoused by philosophers of science and educators of science (Chalmers, 1999; Godfrey-Smith, 2002; Hickey, 2005). In other words, in questions CUL1 and CUL2, students agreed that science is a human enterprise; scientists were raised/educated/trained under a certain culture, so their ways of research and the selection of research topics were strongly affected by the culture and demands of society. In CUL3, the majority of the students agreed that the testing processes are cross-cultural and sometimes cross-historical. The historical anecdote between Galilee Galileo and the Roman Catholic Church was a prevailing example of the conflict between science and religion. The Church asked Copernicus to correct the Julian Calendar (a society demands scientists to work a certain area), later, the Church tried to suppress Copernicus’s heliocentric view of the solar system, and to silence Galileo when he championed it (the societal religious restrictions on science). In
the students’ minds, science finally won out because the Church doctrine of the time did not have strong evidence (science is about objective, absolute truth).

Science should be respected as independent knowledge. This knowledge should not be suppressed by political, cultural power. A historical example is that finally the Catholic Church admitted, Galileo’s heliocentric views of the Earth. Thus, empirical evidence is most important in science and without any strong evidence scientific knowledge cannot be changed.

Due to Catholic dogma and other cultural background, Geocentrism instead of Heliocentrism was widely supported. Religious belief made people Geocentrism divine truth to people. However, finally, such prevalent religious and cultural ideas were not objective and truth was revealed.

6.2.4.3 Comparison between Canadian and Korean Students

There was no fundamental difference between Canadian and Korean students answering the close-ended questions. One difference was in examples that students provided. For instance, the students from School 3 discussed stem cell research in terms of the ethical issues and the benefits for medical treatments. School 3 must be seen as an exceptional case because coincidentally the students had had a discussion on stem cell research before the survey was administered. Canadian students mentioned ethical issues of science, and scientists’ responsibilities in society (e.g., dealing with pollution, and destruction of eco-systems). Korean students were oriented toward the political misuse of science and the abuse of scientists by those in power. This arises from North Korea’s recent scientific efforts to develop
nuclear weapons. This divergence may reflect differences both in the two countries’ science curricula and in their national concerns.

Aikenhead (2008) claims that the Science, Technology, Society and Environment programme (STSE) provides education on the nature of science in Canada. The Ontario Science Curriculum (2007) stipulates STSE as one of the three goals of its science program: “The STSE expectations provide the context for developing the related skills and conceptual knowledge necessary for making connections between scientific, technological, social, and environmental issues. The STSE expectations often focus on aspects of environmental education.” (p.16). Meanwhile, as is well known, tension between North and South Korea has continued unabated since 1945. Further, Korea’s National Science Curriculum equates scientific development with national prosperity (Korean Science Education, 2007).

To summarize: students voiced consistent views on socio-cultural embedded scientific knowledge in both open- and close-ended responses. Most students agreed that social, cultural and political values have influenced science, particularly in selecting what to study and how to study. In discussing the assessment of scientific knowledge, students tended toward universal views. No significant differences on this theme were found between Canadian and Korean students, except in the focus of their examples. Canadians stressed ethical issues, while Koreans emphasized political issues.

6.2.5 Scientific Methods

Many philosophers of science have attempted to establish a theory of knowledge development, and the number of the theories is as many as the number of philosophers. Not only epistemological anarchists, but scientists generally agree that there are no single, logical
step-by-step scientific methods used by all scientists. Since Francis Bacon’s experiment-intensive inductive approach to the knowledge of nature and Rene Descartes’ mind-centred, mathematics-intensive deductive approach, most scientists as well as philosophers of science have thought that the methods for studying nature should be different from other academic areas (Chalmers, 1999; Goldman, 2004; Goodfry-Smith, 2002; Reichenbach, 1951; Trefil & Harzen, 2010). The subjectivity of scientific observation (Hanson, 1958), Kuhn’s view of the normal science activities within a paradigm (1962, 1970) and methodological anarchism (Feyerabend, 1975), spurred the replacement of entirely rational and systematic methodological perspectives in science with a wide variety of scientific investigations.

However, Trefil and Harzen (2010) who are scientists argue that science has a unique method in its major procedures rather than a messy variety of approaches. They diagramed this method as follows:

*Figure 6-1 A Model of Scientific Method*

(Trefil & Harzen, 2010, p.10)
**Question:** With examples explain whether scientist follow a single, universal scientific method Or use different methods? For instance, the universal method means such steps as define the problem ➔ gather information ➔ form a hypothesis ➔ make relevant observations ➔ test the hypothesis ➔ form conclusions ➔ report results (SUSSI item 6)

**Intention:** This question was adopted to assess respondents’ views of diverse scientific methods. Naïve views are strict on objective inductive or deductive methods, precise experiments or objective observation whereas informed views are “Scientific knowledge is constructed and developed in a variety of ways including observation, analysis, speculation, library investigation and experimentation” (Liang et al., 2009, p.16).

---

**6.2.5.1 Common Themes**

Almost all of the responses to the open-ended questions stated that scientists should use diverse methods because 1) science covers broad areas from the human body to the universe (the solar system and galaxies), 2) no method has been known to discover unknown things; many students mentioned many discoveries come by serendipity, 3) individual scientists’ expertises are different, so scientists should use the methods which they feel comfortable and familiar with, and 4) the “best” method does not guarantee a successful result. A second group of responses embraced both universal and creative methods. For unsolved problems or things unknown, scientists do not know which methods they can use; thus, in a discovery context, scientists use creative methods that they prefer or they think appropriate, and in a confirmation context, scientists should use the universal method. A third group of responses suggested that scientists should exclusively use an established and proven method because 1) it can reduce trial-and-errors, and 2) it can be easily replicable and accepted by other scientists.
Table 6-7

**Summary of Common Themes on Diverse Methods of Scientific Research**

\[ N_{Ca}=114, N_{Ko}=158 \]

<table>
<thead>
<tr>
<th>Sub code (Description of Responses)</th>
<th>Canadian No. of response</th>
<th>Canadian %</th>
<th>Korean No. of response</th>
<th>Korean %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established scientific method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well established (proven methods)</td>
<td>12</td>
<td>11</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>To be accepted by other scientists</td>
<td>6</td>
<td>5</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>To reduce trial-and-errors</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Others (no reason)</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Both (diverse methods + scientific methods)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, at the discovery level, serendipities, limitation of existing methods, different areas and individual scientists’ preference</td>
<td>8</td>
<td>7</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>No, at the confirmation level (Confirming a theory)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diverse methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different disciplines</td>
<td>15</td>
<td>13</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Individual scientists’ preference</td>
<td>21</td>
<td>18</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Limitations of existing methods/discovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (open-mind, problem-solving, no reason)</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>No response</td>
<td>17</td>
<td>15</td>
<td>26</td>
<td>17</td>
</tr>
</tbody>
</table>

The following excerpts represent students who support diverse methods for broad areas of science, and who emphasize serendipities in the discovery contexts.
Science talks about everything that has happened in our world and what is to cane. It talks about animals, life, space, and electricity. I don’t think a universal method, whatever it is, can cover all these different things. (1-28)

Scientific discoveries start from chance [serendipity]. Scientists encounter curiosities and start research to know what they mean. The universal method is what they learned from their education and scientists were forced to follow and memorize the method. If they had not learned the method, they would use diverse and creative ways. The universal method does not guarantee they will discover things either. 과학의 발견은 우연에서 시작된다. 과학자들은 이 우연에 흥미를 느끼고 알아 내는 것이다. 그들이 사용하는 보편적인 유일한 방법의 과정은 살아오면 수업을 받으며 검증에 대해 배우고 주위에 의해서 이 방법이 기억되었기에 할 뿐이지 만약, 이방법을 배우지 않았다면 자신의 개성적인 방법으로 이 문제에 대해 알아 낼 것이다. 그리고 그 방법이 꼭 새로운 사실을 발견하라는 법은 없다. (5,115)

Few students pointed out that in a discovery context, scientists should use diverse and creative methods while scientists use systematic and logical methods in a judgment context. In the following excerpts, students explained why scientists use both methods.

Depending on the situations scientists can use universal or new method to solve problems. When investigating or experimenting to find new and unknown facts, with universal methods, I predict that it would be hard to discover anything. However, if scientists were experimenting or researching to confirm or solve [a] well-known problem, the universal method would be the fastest and easiest way. For example, in science lab activities, we will be given labs with universal instructions for us to observe how the actual experiment will be taken. However, for new discoveries following the universal method will only give the universal results instead of new results. (2-47)

To verify a theory or suggested hypothesis the systematic method would be useful because it is precise and correct. However, solving new problems or discovering something new needs diverse and personal methods. For things unknown, we do not know which methods are good or not. Therefore to make discoveries or tackle problems unsolved until now, different methods will be good. 이론이나 제안된 가설을 검증하기 위해서는 체계적인 과학적인 방법을 써야한다. 왜냐하면, 그 방법이 정확하고 맞기때문이다. 하지만 새로운 문제나 발견을 하려 할때는 다양하고 각 과학자에게 맞는 방법을 써야한다. 알려지지 않은 것에 대해서는 어떤 방법이 좋은지 나쁜지를 알수 없기 때문이다. 그러므로 새로운 것을 발견하거나 지금까지 몰리지 않는 문제에 대해서는 다양하고 창의적인 방법을 써야 줄다고 생각한다. (6-1)
About 24% of Canadian students and 35% of Korean students agreed to the idea that scientists should exclusively use universal methods in their study; if a scientist follows the universal step-by-step scientific method, he/she can reach a good theory, if not, he could get a less developed theory. Their reasons were 1) that great scientists have proven the universal method, so they can easily develop new knowledge; 2) that if scientists use the methods, they can reduce failure from trial-and-errors; therefore, they can save time and materials that they used during research; and 3) using the step-by-step scientific methods, other scientists can confirm theories easily. The following excerpts outline the students’ ideas:

Scientists use proven methods. The great scientists like Galileo, Darwin, Aristotle, Newton etc. established a very large base of science and we need to do keep accumulate knowledge like a pyramid. When we reach the top of the pyramid, all problems that we face currently will be solved. That would be the golden age of the history of humans. (4-43)

If a scientist uses his own ways to discover a theory, and he thinks the theory is true, he would reorganize his work using the universal methods in order to publish it an academic journal so that other scientists could easily gain access to the theory. That is, the scientist will adjust his own methods to the universal methods, thus, whatever the methods a scientist uses at the beginning, the final methods would be universal. 만약 한 과학자가 여러 다양한 방법을 통해 진실된 과학적 사실을 얻었다면 그 사실은 문제에서 언급한 방법으로 다시 정리가 될 것이기 때문이다. 과학자들이 자신의 발견을 간행물에 실을 때 논리적 순서로 기재하는 이유는 좀 더 그 실험을 간단화시키고 좀 더 보편적인 방법으로 증명함을 통해서 다른 사람들이 쉽게 이해할 수 있도록 하기 위해서라고 생각한다. 처음에 어떤 방법을 써던 간에 다시 정리할 때는 과학적인 방법으로 정리한다. (6-48)

In my opinion, I do not think scientists need diverse methods because scientists should focus mainly on facts. If the scientists come up with some creative idea, and try this and that, they may end up putting what is not true and their own opinion into their data. An example would be if the scientists who came up with the periodic table of elements had made up “educated guesses” and “logical thinking” if not, then the elements that we study now can be wrong. (Quotations original) (3-32)
Scientists should research using the universal method. Many scientists identify problems and formulate hypotheses, and then they try to verify their hypotheses. During these processes, they can find something wrong, and then they can go back to their data gathering steps (observations). Repeating these processes enables scientists to get some crucial results. Darwin’s natural selection theory was accepted over Lamarck’s theory of use and disuse, because Darwin’s data gathering method was more advanced and thorough than Lamarck’s. Due to a lack of scientific method, Lamarck’s theory came to a wrong conclusion.

6.2.5.2 Comparison between Close-and Open-ended Answers

Two different types of data on scientific methods were approximately consistent. The two close-ended questions (METH1 and METH2) addressed scientific methods and the third (METH3) dealt with the outcome of scientific methods. The ratios of agreement to the first two questions were METH1 vs. METH2=17%: 20%. With the third question, 32% thought that if scientists used correct methods, they could have correct results, which showed a significant increase of support for “correct” scientific methods. Figure 5.13 (e) presents the distributions of these three close-ended questions, and shows that most students did not agree with the idea of logical and systematic step-by-step methods. Relatively, a little higher proportion of open-ended responses, (about 24% of Canadian and 35% of Korean students) supported systematic and step-by-step scientific methods. Meanwhile, the responses that supported diverse ways of scientific research were 60% in close-ended questions. Some 63% of Canadian and 44% of Korean students agreed with this in the open-ended questions.

As shown in Table 6.7, some students (7% of Canadian and 13% of Korean) supported both step-by-step and diverse methods. For various reasons, such as scientists’ preferences, the
limitations of existing methods, and the broad ranges of scientific research, scientists use diverse methods. However, when they attempt to publish, or verify a study, they ought to remove their trial-and-errors, and make the study acceptable by using systematic methods.

6.2.5.3 Comparisons between Canadian and Korean Students

MANOVA results showed there was a country effect on students’ understanding of scientific methods in METH2 while no such effect was discerned in METH1 and METH3. The open-ended responses, especially those dealing with universal methods of scientific research showed that more Korean students (36%) supported universal methods than Canadian students did (20%).

6.2.6. Section Summary

Open-ended questions probed students’ concepts of NOS, clarifying and providing reasons for their close-ended answers. Common themes were extracted, and a few representative excerpts were cited. Across the five concepts of NOS, the open-ended answers were consistent with close-ended answers in the four concepts while they were only approximately consistent for the concept of Subjectivity. One possible reason for this degree of the discrepancy might be the difference of focus between observation in close-ended questions and interpretation in the open-ended question. Another marked difference emerged between students from a Christian school and those from public schools over the concept of Earth history. No students from the public schools made reference to the Bible on this topic while students from the Christian school did. In accounting for changes in scientific knowledge, a good number of students thought that scientific theories included incorrect information and errors. The students also had progressive views on theory changes although the majority did not agree to the notion that such progress approaches absolute truth. In the close-ended
answers, the students agreed that scientific knowledge needed evidence. Both types of answers consistently maintained that scientific knowledge should have support from experimental results. In examining mutual influences between culture, society, religion and politics and science, Galileo’s historical anecdote was a dominant example of how society and religion affected science in a negative way. Chapter 7 further discusses students’ views of historical anecdotes from a current point of view. Some of the students argued that ideally science should be independent of such values, but the reality is that society or politicians abuse science.

6.3 Semi-Structured Interview Analyses

6.3.1 Overview

After the quantitative data analyses, 4 Canadians (1 boy and 3 girls) and 5 Koreans (4 boys and 1 girl) participated in semi-structured interviews. Interviewees’ open-ended answers on NOS concepts had already been clearly stated, so most of the interview questions focused on their learning activities and tasks, and assessment. Canadian interviewees were from School 1 (1 boy and 1 girl) and School 4 (2 girls). As mentioned, School 2 participated in this research on condition that the students would not participate in the interviews. Many students from School 3, signed their names to the survey; however, when contacted through their science teachers, they did not want to participate in the interviews. This also happened in School 7. It was also found difficult to get teachers to participate actively in the research. For these reasons, the source material from interviews was rather limited.

Chapter 4 has already outlined the procedures and questions used in the interviews. All students were asked the same questions (see Appendix B); then, individual questions were posed based on their survey answers. To avoid undue burdens in writing, students’ perceptions of
learning activities and tasks (PLAT), and of assessment formats (PAFORMAT) and content (PACONT) were not included in the survey’s open-ended questions. Instead, semi-structured interview questions addressed these domains and examined how student perceptions on these themes related to their understanding of NOS. Since the interviewees signed their names on the survey, the close-ended answers and open-ended answers of the NOS concepts could be identified and reviewed before conducting their interviews.

Interview results are here reported in the following order: 1) common themes, 2) inter-related themes, 3) emerging themes, 4) comparisons among interviews and 5) comparisons with close-ended and open-ended responses. Pseudonyms are used.

### 6.3.2 Learning Activities and Tasks

Regardless of country or school, lecture-typed learning activities were dominant while presentations and class discussions were not frequent learning activities. Memorization of scientific knowledge and teacher-demonstrated and guided lab experiments were likely to occur more often in Korean classrooms while solving questions by students were often activities in Canadian classrooms. The PLS analysis showed students’ perceptions of their learning activities and tasks (PLAT) were good predictors for all 5 concepts of NOS in the combined sample, and for 4 NOS concepts in the separate Canadian and Korean samples (refer to Table 5.22, 5.25, and 5.26, the coefficients of the structural models, t-values and their significances).

The interviewees were asked what kinds of classroom activities they thought most effective to learn science; what their actual class activities were, and if their class activities affected their views on science. The actual classroom practices mentioned were consistent with
the quantitative results: lecture-types were most frequent while interviewees’ preferred
learning activities were 1) hands-on lab activities, 2) teachers’ explanations, and 3) enough
time to think and understand what the actual meaning of the knowledge and how to apply their
new knowledge.

The interviewees liked lab activities because, they thought, the activities were fun, and
made it easy to understand and memorize scientific knowledge (seeing is believing). A few
thought that hands-on lab activities enabled them to cultivate logical thinking while they were
planning and expecting what results would be.

Teresa: I love lab classes. When you do hands on activities, it makes learning fun yet
educating. In an ordinary class where you only listen to the teacher and write down
notes, that is effective to a point, but when I experience it with my five senses it goes
into my brain a lot more quickly and efficiently.

Bill: Experiments are the most effective way to learn science. There's nothing like
seeing a law or theory works before your eyes, and I think a lot of kids learn better
physically. The experiment becomes a personal memory when something happens in
front of my eyes. .. By doing the lab researched, it will make sure the student will learn
and understand why it happened and how it happened.

Kyu: I think doing lab activities are good. If I do experiments, I can understand easily
and I can remember the content longer. In addition, experiments enable me to think
logically and systematically while I am planning and expecting the results. (제
생각으로는 더 실험을 많이 했으면 좋겠습니. 실험을 하면, 이해도 쉽고, 또 오래 기억할 수
있으니까요. 그리고 실험을 어떻게 알 것인가 계획을 짜를 때 체계적으로 생각하도록 하며,
결과를 논리적으로 짐작할 수 있게하기 때문에 실험을 많이 하는 것이 좋다고 생각합니다.)

Christina: Since science is mostly about memorizing the materials, doing lab tests will
get the students to do hands on activities, which could help them memorize the
contents.

Noh: For me, the methods employing both studying and involving hands-on activities
are effective. First I learn theories and laws by teacher’s explanation. Later I do some
relevant experiments. This learning method can lead me to apply and expand my
As the above excerpts show, the interviewees saw science lab activities as a learning tool to gain scientific knowledge rather than as a learning goal of how to do scientific experiments: planning, measuring, observing, recording, graphing, finding patterns and interpreting. Hodson (2005) classed this view as cognitive argument (p.168) on lab activities.

One Korean interviewee in School 5 raised critical concerns about “cookbook-type” experiments. Hwan said that lab experiments, which involved just following the guidelines from teachers or textbook, were not very helpful for learning science. According to him, science experiments should be directed and performed by students themselves. They should stimulate and satisfy students’ curiosity and enrich problem-solving abilities. But “recipe-style” experiments could not enhance such abilities. A sample lab report was attached (Appendix F). The title, aims of an experiment, apparatus and procedures were given by their teachers, and students were expected to follow the process. What they had to do was to measure and record the data and calculations based on what they learned in class.

Hwan: I think experiments are a good method to learn science since our hands involve learning. However, without a curiosity, “Why?”, a problem solving of “How?” or prediction, “What?” an experiment just following what the textbook provides cannot be helpful to develop a high level of thinking abilities. 직접 체험하기 때문에 실험이 과학 학습에 좋은 방법이라고 생각한다. 그렇지만, “왜?”라는 호기심과 “어떻게?”라는 문제 해결력이나 “무엇?”이라 결과를 예상하는 실험이 아니라 단순히 교과서 대로 따라하는 실험이 크게 생각의 능력을 키우지 않는다고 생각한다."
Another Korean interviewee, Sang Hyun, pointed out that their current learning was too much concerned with memorizing factual knowledge rather than understanding the knowledge or applying what they learned during classes into diverse everyday life situations. He observed that usually students did not have enough time to think how a principle of science can be connected with other principles or other scientific theories.

SangHyun: I think current science education is more on memorization of scientific knowledge rather than focusing on understanding it. Too much knowledge within a short time. I would rather have enough time to understand and some opportunities to discuss the knowledge by myself. I think students need time to learn science by themselves. And I like classes that do not require copying or memorizing but focusing on understanding. I hope teachers ask questions for students to think and then to answer. Mostly teachers ask questions and expected students to answer immediately.

Two students out of nine preferred teacher-directed lecture-typed learning. Nancy said that lecture-typed learning activities were straightforward and Hwan believed that teachers’ explanations of complicated concepts were helpful to understand.

Nancy: I like the science lectures because you learned so many fascinating things about everything in the world, also, it is the most straightforward way for teachers to teach you the facts and theories.

Hwan: The teacher’s explanation is very helpful to learn and understand scientific knowledge. After I listen to my teacher’s explanation, when I review the content by myself, the understanding comes much easier than without any explanation. Full understanding enables me to apply the knowledge into diverse cases and to network with other knowledge. The networked knowledge helps me when I learn new knowledge. However, only listening to teachers’ explanation, copying what the teacher wrote on the board and memorizing scientific knowledge will make the class boring.
In considering discussions and presentations as class activities, Noh said that these methods could be too burdensome for students to prepare. SangHyun thought that ideally these sort of instructional methods could broaden students’ thinking and give them a chance to know their peer students’ thoughts and questions; but he was skeptical about the practicality.

Noh: My science class does many experiments, so I have to submit lab reports, and I occasionally have to do a few projects such as making science magazines or science fairs. Rather than how many times we do such activities but the activities took time to complete and they are a real burden. (프로젝트라기 보다는 실험을 많이 하니깐 실험 보고서 제출하는 것도 자주 있고, 또 간혹 과학 심문 만들기, 과학관찰 경진대회 같은 것이 있어서 많다고 생각합니다. 몇수로 짤자면 몇번 안되는 것 같지만, 그것을 완성하기 위해서는 많은 시간이 걸리고, 또 생각도 많이 해야되고 부담스럽기도 합니다.)

SangHyun: Group projects or class discussions on scientific issues during class enable me to learn from other students; how they think and what their opinions are. However, presentations and class free discussions do not work well, I think, because most students are not willing to present or share their ideas in class. They probably won’t listen to the presenter either. So presentations, discussions cannot be practical in my science classes. 그룹 프로젝트나 과학에 관련된 사회적 이슈 같은 것을 토론하면, 배움의 범위를 넓힐 수 있고 또 다른 친구들이 어떻게 생각하는지 친구들의 견해는 무엇인지 알 수 있어서 좋을 것 같은데 친구들은 발표를 거의 하지 않을 것이고 또 발표한다고 할 지라도 잘 듣지 않을 것이다. 그래서 발표 토론 수업은 현실성이 없다고 본다.

Teresa credited science fairs as a good learning experience whereas she did not positively evaluate class presentations and discussions. According to her, students could learn from peer students’ presentation but they were not professional presenters, so they quite often tended to read a computer screen or texts. She doubted whether these presentations were reliable. Therefore she preferred teachers’ instructions to peer presentations or discussions. If students should present something, the content should not be very important.
Teresa: I think teachers should teach the basic knowledge of the subject, and it is a good experience for students to present specific subtopics. For example, ecology, our teachers explain us how ecosystem works, how producers gain energy, and how consumers spend energy and how each niche contributes to nature. We learned all about limiting factors, tolerance ranges, carrying capacity, etc. … Students should teach subtopics because we as the audience don’t know if the research is reliable data. …Learning depends on presenter if the presenter communicates with audience or eye contact… I can learn a lot, but mostly students are not professional presenters, so I cannot learn as much as I learn something from my teacher.

Science fair was decent. I learned the topics [the effects of Helium gas on voice] that I did, but not other things. It was really a good experience. …You never know until you face it. Someone can train and train to be a fighter, but you don’t know how good you are until you actually fight against a person. ... So that metaphor ties in with the experience of the science fair because I can learn in the classroom and think I understand everything, but until I experiment I won’t fully understand.

As with the quantitative results, all interviewees said that their learning activities leaned heavily toward lecture-typed classes. Five out of nine interviewees thought the best ways to learn science were such hands-on activities as lab experiments since direct experience made knowledge easily understandable and memorable. Two interviewees preferred lecture-type activities but wished for more diverse learning content. Korean students doubted the practicality of student presentation and discussion because of their normally passive role in learning.

### 6.3.3 Interviews on Test Formats and Test Content

*(one correct answer vs. alternative formats & knowledge focus vs. understanding focus)*

In response to quantitative questions on test formats, most students claimed that traditional paper-and-pencil formats most significantly affect their science test scores.
Interviewees’ responses to this same question could be compared to quantitative results. All interviewees thought diverse formats were good in terms of fairness, for students have different abilities, strengths and weaknesses. However, students preferred short or alternative formats to measure their knowledge correctly rather than multiple choice or True or False types.

Bill and Teresa were from the same school, and their test formats consisted of mixed types: True/False, multiple-choice and short answers. Both said that about 50% of the scores were from close-ended formats and the rest were from short answer formats. A science test example given by their science teacher consisted of 24 questions: seven were True/False, ten involved matching items, one was multiple choice and six involved short answers. All close-ended questions were assigned 1 point, while questions with short answers were from 1 to 3 points. This sample test was consistent with Bill’ and Teresa’s responses.

**Bill:** Science tests are a mixed type. They have true or false, multiple choice and short answer questions. I think…for my test it’s [close-ended format] 50% because I always see the short parts in the front for like a page or a page and a half, and then it’s a long question, but usually the long questions are shorter than the short parts…… This could be in the form of short answer or formula questions. Experiments could also be part of a science test, based on how well the student is able to conduct the experiment and analyze the results.

**Teresa:** It [the number of close-ended questions] is about 50% maybe 60%. I haven’t counted the actual numbers. Almost all tests start with T/F questions, and then multiple choice or fill the blank questions are followed. Then, one or two long questions are followed.

When asked which test formats were more important to them, both chose questions with short answers. In the survey Bill indicated that short and close-ended questions were important while questions with alternative answers were not. In Teresa’s survey responses, short-answer

---

11 Due to the copyright, I cannot attach the test examples.
formats came first, while close-ended and alternative formats tied at second. Both thought that True/False and multiple-choice questions tested basic, memorized knowledge, while alternative and short answer formats tested for comprehension and deeper knowledge. Teresa seemed to emphasize memorization as much as comprehension.

**Bill:** I think it’s the short/long answered ones, because it tests your knowledge more and because it’s describes what you are thinking about science. Short answering types, I think, students just memorize the materials, but it doesn’t mean that they would learn it. Multiple choice questions.. only understand basic knowledge of whatever.

**R:** Multiple choices are about basic knowledge or they just without understanding of the context students just memorize the knowledge. Is that what you mean?

**Bill:** Or they could just guess it.

R: Do you mean true or false or multiple-choice questions are basic? If you could answer long sentence questions or open-ended questions, it means you can answer true or false too, right?

**Bill:** Pretty much, because to know how to answer long answer questions it means you have to know the basics. Yeah, I mean it’s like knowing algebra before you learn how to do adding and subtracting or multiplying and dividing, right? You just can’t do it. …. yeah, because if you fully understood what the question was asking, and you know how to answer it then the easy questions will be really easy for you.

**Teresa:** … most of the kids, how they do the test is memorization, because if you can memorize a lot, you can have high marks. I guess it depends on the student. Learn it and if you understand it, it’s easy to memorize. Or if you memorize it and then you start to think about what you memorized it’s easier to learn it.

Nancy and Christina also sat school tests with mixed formats. Nancy personally preferred short or alternative formats to multiple-choice or True/False questions. She thought tests should measure how much a student knows about the subject, not how “lucky” he/she may be at guessing. Some students without any real knowledge can score quite well on multiple choice
or True/False questions; she thought this was not fair. Like Nancy, Christina believed that alternative formats measure her abilities in science more accurately since they reflected her ideas rather than a score of chance choices.

**Nancy:** I think our science formats are very good because our teacher change the phrases in our notes to a completely different sentence and/or scenario. You would only get full marks if you studied and understood instead of memorizing the facts and notes. The tests often ask you to explain how something happens in your experiments. The most proper way to assess the test is to have different ways to evaluate our knowledge.

**Christina:** I think essay tests formats properly measure my ability and knowledge of science because it will actually show how much I’ve learned. I can explain my answers on how and why I have got it instead of guessing on a true or false or multiple-choice question. It would be harder for me to guess the right answers on an essay question.

Christina experienced two different styles of teaching in Grades 7 and 8. She said that her tests in Grade 7 were harder than in Grade 8 because they largely consisted of short-answer or explanatory formats. In Grade 8, her tests were largely in True/False or multiple-choice formats, and she changed her studying style to fit this, even though she felt short-answer or explanatory formats assessed her learning properly. She also said that she changed her study strategies depending on the test formats.

**Christina:** I find it always depends on teachers. As I had two different types of science teachers I had past two years. My grade 7 teacher Ms. A, she was always hard, her tests were always really hard. Most of her questions were short answers and long answers, and she asked us to explain your knowledge, which you actually understood. But this year we have Ms. B and her test is really easy. Most of her test questions are fill in the blanks and multiple choices. Because she gives us the test, she is easier to mark. So recently, I am seeing more on the blanks and multiple choices and short answers. But it honestly depends on the teachers. Every teacher has their [his/her] own style of teaching.

**R:** Which types of test questions make you more study?
Christina: I would definitely short and long answer questions. Usually they are most important on tests. Because teachers weigh more on those long and short answer questions. Short and long answering questions actually show students’ knowledge and their understanding. The answers actually show students’ understanding. Rather than true/false or multiple choices question, mostly I study more for the short and long answering questions.

When asked to discuss varied kinds of test content – laws, theories, factual knowledge, or application – interviewees provided quite fragmented information. They actually could not distinguish clearly between laws, theories, and applications. For example, Bill said scientific terms, facts and a few examples of theory (e.g. particle theory) were important in his science tests. However, his response to the close-ended question on this matter was ambiguous; there, none of these items were important. His open-ended answer on the definition of science resembled his ideas about science tests. He answered, “Science creates new vocabulary (e.g., unipolar neuron, trachea, parasympathetic, etc.).” He thought science tests were mostly about science terms and facts, but the survey actually had no items on vocabulary. Perhaps having no other choice, he answered the close-ended questions ambiguously.

   R: Which kinds of contents are most frequently appeared in you science test?
   Bill: Vocabulary, and… the other theory we learned in grade seven and continuing to grade 8 was the particle theory. So we do learn some theories here and there but mostly vocabulary, not so much of examples.

Teresa also emphasized that memorization of the facts. For her the application of the knowledge or knowing how people use the knowledge in daily life was an extra and it could not be the major test content.

   Teresa: the facts, not so much of the law, because we haven’t gone into many laws yet. So far the only law we learned in grade 8 was Pascal’s law. We’re not in high school, and we’re not in that advanced scientific knowledge yet, so, so far it’s vocabulary and facts.
R: So you mentioned Pascal’s Law. Did your test ask the meaning of Pascal’s law or the use of it in our every day life, or what kind of material uses Pascal’s law. What kinds of questions appear in your science test?

Teresa: For the water unit, you have a cup, it bends on the walls of the surface, that was one. …. But those parts of the test is [are] not the main part. If you don’t know the application the laws to our everyday life, or examples, it’s not the biggest concern. But, if you know and example, then it’s better to know it.

Korean science teachers confirmed students’ responses about their test formats. Their tests do not include True/False formats, because a student has a 50% of answering correctly without any real knowledge. Their school board guidelines recommend teachers to devise questions with five choices (or a 20% probability of a correct answer). Sample tests are found in Appendix F. Like the Canadians, Korean students prefer diverse test formats. They feel multiple-choice questions can measure wide ranges of knowledge within a limited time, while alternative test formats can measure how much a student understands and can apply the knowledge acquired.

Two Korean interviewees pointed out the systemic weakness in alternative formats. According to Hwan and Kyu, these formats cannot differentiate between those who know and do experiments and those who do not, although the formats can quite clearly detect students’ abilities and attitudes to learning. Their teachers acknowledged that their critique was correct; the quality of lab activity reports was important but not as crucial as expected. Lab activities

12 According to a science teacher in School 5, a school board-wide guideline mandates that a maximum 70% of science scores come from paper-and-pencil tests and a minimum 30% come from alternative formats (e.g. lab reports, projects or presentations). For that 30%, she set these standards: 1) attendance at lab activities; 2) submission of a lab report, and 3) the quality of the work submitted. So most students who submitted lab reports or other types of alternative assessments would score at least 20% of the total 30% allotted. The difference between the highest and lowest scores in this category was usually about 12. Of course, a student would hardly have any marks without submitting something.
were performed in groups, so data would be shared with those who actually had not done the work. Appendix F presents sample lab report.

Hwan: Personally, I like current test formats, which include multiple choice [one or two correct answer among five choices] and short answer types. The alternative methods are also good, but they lack the differentiation among students. All test types have their own strength and weakness. I think, a test should be fair for all students and evaluate students’ learning appropriately by adopting diverse methods. The multiple choice format can test if a student knows whether scientific facts or experimental results and can cover wide and variety contents within a short time while process and open-ended questions can test students' understanding of science. Project or lab experiment reports, and presentations have benefits to evaluate a student's abilities in solving problems, or applying scientific knowledge in different situations. (저는 오지선다형과 주관식을 섞어 놓은 지금의 시험 방법도 괜찮다고 생각합니다. 단지, 보고서 평가나 다른 것 평가들이 변별력이 떨어지는 것이 단점이지만, 그리고 모든 타입의 문제들은 각기 장점이 있고 단점이 있다고 생각합니다. 시험일반 다양한 방법으로 정확하게 학생들의 능력을 평가해야하기 때문에 여러가지 방법으로 해야 공평하다고 생각합니다. 오지선다형의 문제는 학생들의 과학적 사실이나 실험 결과를 광범위하게 평가할 수 있지만, 과정 평가나 주관식 문제들은 학생들의 이해 정도를 평가할 수 있어서 좋다고 생각합니다.)

R: You said that the alternative methods couldn’t differentiate among students. What do you mean by that? (실험보고서나 다른 형태의 평가 방식이 학생들의 능력차이를 제대로 평가하지 않는다고 말했는데, 왜 그렇게 생각하는지?)

Hwan: For instance, the lab reports mainly score group work not individuals work. A few students do not participate in the experiment, but they can have good marks without any effort except copying the work of their group members. And, there aren’t very low scores in lab reports, I think, the base line is about 70%. (예를 들자면, 실험은 그룹으로 하는데, 실험이 전혀 하지 않고 다른 학생들것 배껴서 제출해서 좋은 점수를 받고, 또 점수도 대체로 70점이상을 기준으로 주기 때문에 변별력이 떨어 진다는 것입니다.)

Kyu: Science tests should put more weight on alternative methods than on paper-and-pencil tests. Alternative tests should help learn science and not just meet the requirements of the guideline. Actually, test scores should include diverse methods. Evaluating lab reports should also include how actively involved a student was. The report should show the problem-solving procedures used, rather than merely the results. Science should be a subject in which students develop their ability to think deeply and solve problems, not to memorize facts and laws. (과학시험은 필기시험에만 의존하지 않고, 다양한 방법으로 치는 것이 학생들이 얼마나 아는가를 제대로 평가할 수 있다고 볼니다. 단지 형식적으로 그 바탕을 맞추기 위해서 실험보고서라든가 문제 풀이과정을 넣기 보다는 실질적인 반영이 중요합니다. 그리고 실기 평가의 경우 결과보다는 문제풀이과정, 얼마나 실험에
Hyun answered in the survey that factual knowledge, theories and laws and experiment were important in his science tests. He said that what would be in the test completely depended on teachers’ decisions, but definitely the questions should be what they learned. Other Korean interviewees had similar opinions. Sanghyun said that science tests should measure not only students’ learning but also their attitudes and enthusiasm for the subject; he, then, preferred to alternative formats over multiple-choice tests.

**Sanghyun:** You can get scores in most multiple choice or short answer questions by chance without knowledge and full understanding. Of course memorizing scientific laws and theories is important, but more important thing is passion and curiosity that makes you delve into what nature is. Science lab reports or projects are better methods to contain such elements. 내 생각에는 과학 실험 보고서, 과학 신문만들기 같은 것이 제대로 된 평가방법이라고 생각한다. 단답형은 막 찍어도 맞을 수 있고 또 단답형은 개념을 외우고만 있어도 (이해하지 않고) 맞출수 있는 것이 대부분이다. 개념과 법칙들을 외우는 것도 물론 중요하지만, 과학에서 중요한 것은 그에 대해 파고드는 열정이나 호기심이 있어야한다. 실험보고서나 프로젝트 같은 것이 그런 것을 잘 나타낼 수 있을 것 같다.

### 6.3.4 Strategies to get high scores in science tests

Previous literature (Biggs, 1987, 2003; Crooks, 1998; Scouller 1998) has suggested that student strategies to prepare for science tests could reflect the ways their learning were tested. Interviewees were asked how they studied to get good marks in science; their preparation would vary according to the science test format. All students answered that because traditional paper-and-pencil type tests were almost always in the same formats, they do not take test format into account in their study style. For Christina, as her Grade 8 science teacher used more of close endend questions, she adjusted her study strategy. They simply study the textbook or class handouts, or notes copied during class. However, when faced with
alternative formats like science fairs or group projects, students definitely changed their 
approach to study. They would search the Internet, or go to a library to find relevant sources. 
For Korean students, lab reports were the major alternative format. These interviewees tried to 
use a reference book if allowed to\textsuperscript{13}. Therefore, being well used to paper-and-pencil test 
formats, the interviewees felt no need to adjust their strategies to prepare for science tests.

Generally students prioritized understanding the content, and then applying the 
knowledge to different situations over merely memorizing what they learned. All agreed that 
memorization without understanding was not very helpful to get high marks. These interview 
results contradicted their responses to close-ended questions. Items 2 and 4 asked for student 
perceptions about solving questions and memorizing laws, theories and formulae, as learning 
activities. About 21\% of Canadian and 39\% of Korean students responded that memorization 
activities occurred in almost every class. On the other hand, only 9\% of Canadian and 5\% of 
Korean students mentioned student solving questions activities.

On this point, Teresa and Bill, though from the same school, held diametrically 
opposite perceptions of their science learning and NOS understanding. In her responses, 
Teresa said that open-ended test formats were most important; memorization was a frequent 
activity in class; and solving questions was less common. She held to realistic and rationalistic 
views, except on the construct \textit{Diverse Scientific Research Method}. She said understanding 
and application of what she learned were her best way to get high marks. But if she did not 
have time, she relied on simple rote memory of her science knowledge. She also said that her 
teacher prepared worksheets practicing solving questions to be completed as homework. In 

\textsuperscript{13} Most Korean science textbooks do not provide answers to the results of an experimental 
activity, or exercise questions. However, publication companies publish reference books 
including detailed background content and answer keys to all questions.
contrast, Bill asserted in his responses that his science class had more solving questions than memorizing activities. Here are his comments, when asked how he prepared for tests, and how he dealt with material he could not understand:

**Teresa:** The best way to my idea is first to understand, and then apply your knowledge to other cases. But there are times when you were lazy and stupid so you didn’t ask a teacher what you didn’t understand, in those cases it’s all up to memory. …My teacher gave a lot of work sheets, and mostly worksheets were for homework.

**Bill:** I first try to learn it, but it turns out I can’t learn it and I don’t fully understand it I memorize it then try learn it, and if I still can’t memorize it I ask someone in my class to explain it in a non-scientific way. Sometimes how the textbook explains it is a bit too confusing, so I ask a classmate or my teacher who understands it to explain it in a more of a casual form. In order to understand it better, then I read the textbook all over again to see if I actually understand what it’s talking about.

In the survey Noh remarked that lectures, experiments and memorization occurred very often. Unlike other students, he thought presentations and discussions occurred quite often while problem solving did not. For him, close-ended, multiple choice and short answer test formats affected his science test scores, while science lab reports or projects were neutral factors. He seemed to take a relativist and naturalist stance toward the tentative nature of science and the diversity of research methods. On the other hand, he took more of a Universalist outlook on the culture-embedded nature of science. He took the unique view that solving questions was not common, while test content most often dealt with the application of scientific knowledge:

**Noh:** Mostly teachers cover a good amount of content within 45 minutes, so they cannot give students enough time to think or to solve questions. The teacher asked questions and required students to answer the questions immediately.
R: You said that your science classes do not have enough time to solve questions whereas you are often involved in experiments and listening to your teacher’s explanations. Meanwhile you think that your science test content is about application of scientific knowledge, Noh, do you think your science tests do not reflect your science class learning? (노는 문제 풀이 시간이 적다고 대답한 반면, 프로젝트 하는 것은 많고 시험에서 응용 문제가 많이 출제된다고 대답했는데, 수업 활동과 시험문제의 방향이 일치하지 않는다고 생각하는지?)

Noh: My science class does many experiments, so I have to submit lab reports, and I occasionally have to do a few projects such as science magazines or observation or science fair. Rather than how many times we do such activities does not matter, but the activities take time to complete and they are a real burden. (프로젝트라기 보다는 실험을 많이 했기 때문에 보고서 제출하는 것도 자주 있고, 또 간혹 과학 신문 만들기, 과학관람 경진대회 같은 것이 있어서 많다고 생각합니다. 횡수로 떨지자면 몇번 안되는 것 같지만, 그것을 완성하기 위해서는 많은 시간이 걸리고, 또 생각도 많이 해야되고 부담스럽기도 합니다.)

Individual students perceived their learning activates differently. Considering Noh’s answer, students’ view of these activities was determined not by their frequency but by the workload they entailed. Whatever their view of learning activities and of test formats, almost all the interviewees agreed that understanding content was the most important factor to get high scores in their science tests.

6.3.5 Knowledge of Science and Knowledge about Science

One of the major questions of this research is to clarify the relationships between students’ perceptions of their learning activities and tasks and their understanding of NOS. To probe this issue, interviewees were asked which, they thought, was more important: knowledge of science or knowledge about science. All the interviewees thought that learning knowledge of science was more important than learning knowledge about science. They offered such reasons as that the knowledge of science is the basis to understand nature.
Nancy’s response well reflects how she viewed science and knowledge about science. Teresa’s excerpt shows that knowledge of science rather than knowledge about science is the premium for a successful life. In her reply, “an intelligent and meaningful conversation with someone” can be understood as scientific literacy. This literacy comes from knowledge, so, knowledge of science was more important for her.

**Nancy:** Learning scientific knowledge helps other to explain why the world works the way it does. Learning knowledge about science is based on the experiences of others. I think learning science is more important than learning about science because knowledge, theories, laws and facts are more accurate than opinions and experiences. It can be passed on as a fact for many more generations.

**Teresa:** For scientific knowledge, it’s important, because then you know how the world works, so many questions are answered through the knowledge, theories, laws and facts. It’s knowledge that is needed to be known to have a basic understanding of anything, you need this knowledge to have an intelligent and meaningful conversation with someone or to have a good and successful life. I think that this is more important than learning about what science is, that should come second, because if you don’t know what the science is then how will you know what effects our society and our environment.

Christina, also emphasized knowledge of science over knowledge about science in her science class, and insisted that her views on science had been widened through her knowledge of science. Christina said the content she wanted to learn was knowledge of science covering such broad ranges as biology and chemistry. She did not consider knowledge about science at all.

**Christina:** I think the kind of contents that should be taught in science class is biology, how earth is formed and chemistry. I think learning biology can teach us about the living things on earth, which will help us understand more about the creatures around us. Chemistry can help us understand more about the different substances, which would become useful when mixing different chemicals. The contents that should be on my science test are biology, and lab researches. Lab researches can be about many contents in science like testing
out chemical reactions or pulleys. Biology can teach us many things that will expand our knowledge towards life forms of different sorts.

I didn’t think my science class had taught knowledge about science. I have always thought science was about chemicals and lab tests. But, when I started to learn this subject, I found out that there is much more to science. I learned, science can be about pulleys, chain reactions, animal cells and much more.

The Korean interviewee, Sanghyun, mentioned a potential positive aspect of knowledge about science. Knowledge about science could be an effective tool for learning the knowledge of science. As with the benefits of doing experiments in science class, Sanghyun’s idea of knowledge about science was for the effective knowledge of science. His point concurs with Hodson (1998) and McComas and Olsen (1998) who argue science curricula should include NOS content for this very reason:

Sanghyun: While I listen to teachers’ explanation, copy what the teacher wrote on the board and memorize scientific knowledge during class, I feel science is full of knowledge. If science class includes the history of science, science would be more interesting and science class would be more fun; I mean, who discovered a theory, how the theory was discovered, how the initial theory has been changed, how we use the theory in everyday life, etc. Such science classes can widen my understanding of science.

Kyu acknowledged the importance of knowledge about science as well as knowledge of science. In discussing the effect of his learning activates and assessments on his views of science, he stated that knowledge about science is a way to enhance critical thinking. He insisted that the lack of critical thinking in education is ruinous. For this he cited the historical
example of Aristotle’s theory of the four elements. This untested idea had controlled people’s thought for two thousand years from ancient times to the Middle Ages.

**Kyu:** Science teachers should teach students tentative nature of scientific knowledge. Aristotle’s wrong idea of 4 basic elements had governed people’s understanding of material components for 2000 years. The reason why people had believed his idea as truth for such long time was, I think, due to science education. Science educators had taught students to believe scientific theories as truth instead of teaching them critical thinking. Galileo said, ‘We need to reconfirm a scientific theory which has been proved.’ If teachers teach students to believe a scientific theory as a truth, which could be revealed as a wrong theory in the future, the education would be an obstacle for scientific development. So science teachers should teach students the fact that current scientific knowledge that we are learning now could be wrong, and teach to enlarge our critical thinking.

6.3.6 PLAT, PAFORMAT and PACONT and NOS

The quantitative analysis results (Section 5.6) showed that PLAT, PAFORMAT and PACONT were effective exogenous variables to predict students’ understanding of NOS. The $R^2$'s ranged from 0.19 (Sociocultural Embeddedness) to 0.63 (Diverse Scientific Research Methods) for the Canadian sample, and from 0.28 (Sociocultural Embeddedness) to 0.47 (Empirical Evidence Based Scientific Knowledge) for the Korean sample. All $Q^2$'s of the significant independent latent variables, Stone-Geisser’s indexes, were positive so that the model could well reconstruct the observed variables of the NOS concepts. If the interview data supports these quantitative results, the model postulated in this research would have solid grounds.
Interview questions were formulated to test whether how students learned and how they were assessed affected their views of science. Two of the interviewees said that their views on science and scientific knowledge might not change because of their learning, the test formats, or the content of science. These two students had very firm opinions about assessments. One of these was the Korean student Noh. His answer carried two implications: 1) for Noh, NOS concepts raised the question whether or not scientific knowledge is about truth; and 2) the tests conducted in school do not measure how much truth a student knows but how much the student has learned in class.

Noh: I don’t think they affect my views on science and scientific knowledge. I don’t think a test is about truths. Tests are for how much I know what I learned during classes. They don’t test about truth. So I don’t think what are in tests affects my views on science. It cannot be an overstatement that nothing that I learned in school is truth. (내 생각에 과학 시험은 과학을 보는 관점을 변화시킨다고 보지 않는다. 시험은 사실이나 진실에 관한 것이 아니라 내가 배운것을 얼마나 알고 있느냐를 테스트하는 것이다. 그러므로, 과학 시험이 과학을 보는 관점을 변화시킨다고 생각하지 않는다. 사실 학교에서 배우는 것 중에서 아무것도 진실이 아니라고 해도 과언은 아니다.)

The other was the Canadian student, Teresa. She thought that diverse test formats were a way for teachers to “go easy” on students who do not study hard, so they can get some scores by chance in True/False and multiple-choice questions.

Teresa: No, not really, I think that the teachers put it there so that a student can get easy marks, so that they won’t fail at least. “Teachers don’t fail you, you fail yourself” pg 43 of The Scripture Passage. That is my religion textbook, and it talks about blame. The teachers don’t purposely fail you, they are paid to teach you and set you up to get ready for life, but there are some students that still don’t study, and I think that true or false, and multiple choice questions are used for those students that can at least get some marks to that they hopefully won’t fail. So no, as long as you have an understanding of the subject, it doesn’t affect the ways I think of science.
R: Do you think science is difficult, so science teachers give those students easy questions to answer?

Teresa: No. The test formats are not subject matter. Whatever the subjects are, some students don’t study at all. For most of the kids if they don’t know it’s an educated or just random guess. Especially for True and False, it’s a 50-50 chance just guess……And multiple choice and true or false questions are benefits both students and teachers. Students can have marks without knowing, and teachers mark them with easy.

The rest of the interviewees thought that how they learn science could affect their views of science. Kyu, acknowledging the influence of how one learns science on how one views science, insisted that teachers should teach students diverse facets of science although there may be only one “correct” answer in their test questions. His opinions were compatible with claims in comprehensive literature that explicit instructional approaches to NOS are more effective than implicit approaches (Khishfe & Abd-El-Kblick, 2002). Kyu instanced the two thousand years in which Aristotle’s theory of the four elements was accepted as absolute truth. According to him, educational institutions failed to teach students to think critically (as cited above).

Nancy expressed the view that students were strongly affected by their teachers because the teachers determined how and how much they learned.

Nancy: I think science classes affect me in a great way on my view in science because I learn science the way it is taught to me. I would not know whether or not the teacher is teaching the correct information, and us students are all learning it the teacher’s way and his/her explanation. Last year I thought science was boring, but this year I like science, because she teaches very well.

Hwan actually showed great enthusiasm for reading diverse books, but he admitted that most of his science learning came from school, so school science was his most important learning resource. At the end of the following excerpt, Hwan gives his view of science indirectly. To
him, science is a way to improve one’s social status, to enhance one’s logical thinking, and to understand nature without superstition. This connection between learning science and improving social status could be on the same continuum with the position of Korea’s National Science Curriculum, that science aids Korea’s technological advancement as well as the well-being of individuals and Korean citizens as a whole.

**Hwan:** I think they do. Most of my science learning was through school science classes. Middle school science is not very profound this basic knowledge does not bring direct relations with the high technology society. But I am sure that it will be the foundation for further learning.

I know whatever the test formats are, and whatever the tests are about, the answers to science tests are not absolute truths, but I also know they are the best answers we can have so far. Science tests are a way of learning science, and without tests learning will less occur. Although my teacher teaches certain knowledge of science, if the content will not be in the test, almost all students will not study it, including myself. So, tests are important in learning science. If we do not learn science, and public education does not provide science, science will belong only to a few people of high social status. Then, the gaps between low and high SES will be widened. Another reason why science education is important is, that we have to consider, although science and scientific knowledge are not true, it is worthwhile for people to learn. It can raise logical thinking ability, and problem solving abilities, and our understanding of nature but not in a superstitious way.
A major issue in this research was whether student learning activities and assessments could be predictive of their understanding of science. The PLS results of the quantitative data and the interviewees’ responses revealed that PLAT, PAFORMAT and PACONT were indeed predictive for student understanding of NOS. Two of the interviewees said how their learning was assessed could not affect their views on science since they regarded the tests as measures of learning but not of truth. Another reason offered was that teachers provided close-ended questions as a source of easy marks for poor students. The remaining seven students agreed that their learning activities and tests did indeed affect their views of NOS since school remains their major source of scientific knowledge.

6.3.7 NOS

Teresa was from School1. Her responses to the question of theory-laden observation did not meet the original intention of the question. She rejected both of the theories argued by scientists to explain how dinosaurs became extinct (the collisions of meteorites or the explosion of volcanoes). She declared dinosaurs became extinct because of Noah’s flood. Here is her response to the survey:

Teresa: I do not believe that either of those things happened. I believe in creation and that God created the universe and science too. I believe the dinosaurs were extinct because of the flood of the world (Noah’s flood). There was most likely a change of pressure in the air at the time, so they could not survive in that climate.

The interviews did not intend to study or challenge students’ religious beliefs at all, but to confirm students’ perceptions of NOS.

R: If there is a contradiction in explaining something like the extinction of dinosaurs between science and your religious beliefs, which explanation do you accept?

Teresa: I do believe the Bible. It is always correct but science is changing all the time.
**R:** Do you think scientific knowledge approaches to an absolute truth?

**Teresa:** I am not sure whether it will ultimately reach the knowledge of Bible, but every new theory is better than old theories.

For Teresa, the knowledge of the Bible was the summit of knowledge; science is a human way to study nature as God’s creation. Her answers to the open-ended questions show her ideas quite well:

**Teresa:** If a scientist develops a theory, it doesn’t always have to be true. It is something that is most very likely possible. For example, the particle theory isn’t exactly true. We cannot see particles even with a microscope. The theory picture makes sense. …It [Science] shows how wonderfully our world was made, and has a lot to do with nature.

Bill’s response to the survey regarding the extinction of dinosaur was blank. His close-ended question showed that he knew various scientific terms; he enumerated what science does in our world and he admired science.

**Bill:** Science is different from other subjects because science is a combination of all and creates new vocabulary. And we learn about our environment in science by interacting with other organisms, studying cells, learning about particles, the building blocks of our world with science we can discover natural disasters before they take place. We use science to discover history like discovering mummies in Egypt and locating preserved artifacts. Science is even in physical education. With science we can discover the trajectory, velocity, and kinetic energy of a kicked soccer ball. Science is amazing.

According to his teacher, all students at Bill’s Christian school were regarded as Christians. His responses to the two other open-ended questions revealed that he seemed to have a keen interest in science and had learned a lot. Consequently, it was intriguing to explore his opinions about the extinction of the dinosaur and on the differences among scientists on the
topic. Like most of his classmates, Bill believed that Noah’s flood led to the extinction of dinosaurs.

**Bill:** More Christian scientists should work to prove that Bible is correct. Science is not the enemy of Bible, science is all about discoveries and continuous developing.

These two interviewees’ views on science and scientific study resemble the outlook of Isaac Newton and of the 18th century Deists. Newton, as a natural philosopher\(^{14}\), wanted to understand what nature was, and how it operated, as God’s creation. He would search the Bible for scientific data and evidence. Further these two students’ responses were very consistent with those of other students of this Christian school. About 30% of these students said that the Noah’s flood in the Bible was the real reason for the dinosaurs’ extinction, rather than the interpretations of two groups of scientists. Also in response to the open-ended question what science is, students from the Christian school said that science reveals God’s handiwork in creation. As Student 1-45 stated, “It [Science] teaches me that I have such an intricate God who knew exactly what He was doing when He created this earth.”

Noh had a strong relativist perspective on theory-laden observations, so further questions prompted him to clarify his views. Her explanation on why different groups of scientists reached different conclusions on the same event was very persuasive. The authors of the question, Abd-El-Khalick et al. (1998), had predicted that to say different conclusions arise from lack of evidence or direct observation could be a clue that the person had “a naïve realistic view.” On the other hand, they forecast that to give such reasons as the limitation of

---

\(^{14}\) At Newton’s era, the term “scientist”, had not been coined. Whewell (1840, cited in Gregory, 2006) proposed the word, “scientist” in 1840 in his book *The Philosophy of the Inductive Sciences.*
human beings or scientists’ background knowledge could be clues to an “informed view.”

Noh’s example stressed scientists’ prior knowledge using a linear function, interestingly his response was the same with what a scientist answered in Wong and Hodson’s study of scientists (2009). The following response was consistent with his close-ended and open-ended responses to Subjectivity.

**Noh:** I think individual scientists’ background knowledge is most important. If we find an answer to this equation, $1+3=\,?$, the answer is clearly 4, but if you find answers to this equation, $x+y=4$, there are many answers depending on conditions. If a person knows only natural numbers, the answer will be within these three cases, $(1,3), (3,1)$ or $(2,2)$, if a person knows integers, the answer will be completely different from the person who knows only natural numbers. Likewise if a person knows real numbers, the answer will be different from those who have limited knowledge of numbers. Like the answers of this equation, when the conditions or the data are not clear, scientists’ background knowledge, and their degree of knowledge are a crucial element in observations and interpretation of data.

Hwan explained how a theory could be accepted by societal pressures as well as by empirical evidence using an example of a fake skull of “Piltdown Man.” His acknowledgement of the mutual interactions between science and culture, society, and politics in both close- and open-ended questions went well beyond the guidelines provided by the authors of the VONS.

**R:** I was really impressed by your open-ended responses. For example, the story of “Piltdown Man” was a unique answer that I found among all other students’ answers. Do you have any interests in anthropology or the history of science? Does your reading affect your views on science? (환의 대답이 굉장히 인상적이었어. Piltdown Man 같은 예는 전체 학생들 중에서 유일한 대답이었거든. 혹시 인류학이나 과학사같은 것에 흥미가 많은지? 그리고 너가 읽은 것들이 과학을 보는 관점에 영향을 미치는지?)
Hwan: I like reading on scientists’ biographies. Science textbooks mostly provide only laws or theories, but do not give any stories about the scientists, their lives, and personalities, how they could learn the theories or laws. But their biographies include plenty of interesting episodes and how they learned them. When I learn some theories, and I know the scientists I am motivated to learn about the theories. (과학자 전기 같은 것을 읽는 것을 좋아 합니다. 과학교과서에는 어떤 과학자가 무슨 법칙이나 이론을 알아냈다는 것만 나오지 그 과학자가 어떤 사람이었고, 어떻게 그 법칙을 알아내었는지 하는 것은 없지만, 전기를 읽으면, 과학자들의 성격 같은 것도 알 수 있고, 또 어떻게 하다가 법칙을 알아냈는지, 과정도 알 수 있어서 과학 시간에 그 과학자와 관련된 것을 배우면 더 재미 있습니다.)

R: You said that societal demands and people’s development of cognition can affect whether or not a theory was accepted in spite of the lack of evidence. Is my understanding right? (사회적인 요구와 사람들의 인식정도의 발달에 따라 어떤 증거가 부족한 과학 이론이 받아지고 거부된다는 말지만? 내가 잘 이해했나?)

Hwan: I don’t think societal demands or people’s degree of understanding are a decisive element for accepting a theory, but it absolutely affects the acceptance. For instance, Galileo was accused and was subjected to the Inquisition because he supported Heliocentrism while Newton supported Copernicus, but he was not accused. If Darwin had published his theory in Galileo’s era, he would have had the Inquisition. Of course, many laughed at him and still some extremists refuse to accept the theory, but most Christians accept the evolution theory nowadays. 완전히 사회적 요구가 과학 이론을 받아들이는 것을 결정하지는 않지만, 많은 영향을 미친다고 생각합니다. 예를 들자면, 지동설을 지지하다가 갈릴레이는 종교 채권을 받았지만, 뉴턴에 와서는 지동설을 지지한다고 해서 채권을 받지는 않았습니다. 만일, 다윈이 갈릴레이나 그 이전 시대에 진화론을 주장했다면 다윈도 종교 채권을 받았을 수도 있지만, 그렇지 않았습니다. 물론, 많은 사람들이 조롱하였고 지금도 진화론을 전혀 받아 들이지 않은 기독교인들도 있지만, 대부분 지금은 받아 들이고 있습니다. 이것은 사람들이 인식이 진화론을 받아 들일 수 있도록 발전했다고 볼니다.

R: You think that people’s understanding of a theory is an important factor to accept a scientific theory. Then, what do you think the role of empirical evidence? You know that Darwin’s evolution theory has been modified and corrected a lot since he suggested it. Do you think that regardless of its faults, because people were willing to accept the theory, it was accepted to scientists’ community? (너는 사람들이 어떻게 과학 이론을 이해하는가가 그 이론을 받아들이고 거부하는데 중요하다고 생각하는데, 그렇다면, 실험적인 증거는 어떻게 생각하니? 너가 말하시피, 다윈의 진화론은 발표 되고 난 후 지속적으로 수정 보완되었잖아요. 오류에도 불구하고 사람들이 그 이론은 받아들이고 싶어해서 과학자들이 받아 들였다고 생각해?)

Hwan: I think subjective ideas and theories without any empirical evidence basis cannot be accepted as scientific knowledge regardless of societal demands or people’s understandings across the era. I provided the example of “Piltdown Man” because at that time people did not know the evidence was not true. People believed that the skull
was real. Evidence and societal demands were well matched. If a theory were far advanced or its time, so that people could not accept the theory, but it had strong evidence, it would be accepted later, like the Copernicus’s theory. I am sure that science is a way to approach to truth, and through changes and the endeavor to change by science itself, science can approach truth. 주관적이고 증거가 바탕이 되지 않는 이론은 어느 시대를 막론하고 과학적인 이론으로써 받아들여지지 않는다고 생각한다. 아주 획기적인 이론이라서 사람들이 잘 받아들이기 힘들다 할 지라도 증거가 확실하다면 나중에는 과학적 이론으로 받아들여진다고 생각한다. …(지동설)…. 한가지 확실한 것은 과학이 진리에 한층 더 가까워지는 길이며 과학 그 스스로도 변화를 통해 진리에 가까워 지고 있다는 것이다.

6.4 Section summary

In follow-up to the survey, four Canadian students participated in semi-structured interviews. Five Korean students joined the interview process through e-mails and Internet chats. The analysis of quantitative data and of individual interviewees’ responses to the survey generated a series of probing questions. There were no open-ended questions on PLAT, PAFORMAT and PACONT in the survey, so the interview questions focused were on these constructs.

Five interviewees chose hands-on activities as their most effective ways of learning science. They had the perception that “seeing is believing”. Traditional paper-and-pencil test formats were decisive for the interviewees’ science test scores whereas diverse tests formats were supported in terms of fairness for different students. In discussing the distinction between knowledge of science and knowledge about science, the interviewees regarded knowledge of science as the foundation of science education. They thought of scientific theories and laws as “hard science” while they felt that knowledge about science was a mere matter of people’s opinions.

Probing questions for the concepts of NOS were based on interviewees’ responses to the survey. The interviewees from School 1 supported a Biblical interpretation of the changes
of living things while other interviewees made no appeal to the Bible to support their ideas. Although the interviewees put greater stock in knowledge of science over knowledge about science in their science learning, they agreed that the school science curriculum should include content about how science has been developed, what effects science has on society, and how science can enhance critical thinking.
7. DISCUSSION AND CONCLUSIONS

This chapter includes a summary of the study and discusses the major findings from the analyses of quantitative and qualitative data made in Chapter 5 and Chapter 6 with respect to the research questions and related literature where possible. The later parts include recommendations for future studies and limitations of the study. The content of the sections are as follows:

- Section 7.1 overviews the research
- Section 7.2 summarizes the answers to the research questions
- Section 7.3 discusses the major findings
- Section 7.4 draws conclusions from this study
- Section 7.5 recommends some possibilities for future research, and finally,
- Section 7.6 takes account for a few limitations of this study.

7.1 Overview of This Study

The Nature Of Science (NOS) has been recognized as a way of enhancing scientific literacy; however, understanding this field is not easily achievable. Reasons for this may lie in the complex and broad concepts of NOS as well as in discrepancies in the classroom practice of science teaching, in such areas as learning and teaching activities and student learning assessments. The aim of this research is to identify students’ perceptions of their learning activities and tasks (PLAT) and of their perceptions of assessment formats (PAFORMAT) and content (PACONT) on their understanding of NOS. It also intends to examine the relationships among these perceptions and their understanding.
Previous research into students’ learning indicates that students’ perceptions of their learning environment and assessment are good indicators of their attitudes toward and achievement in a subject. Such previous research results facilitated formulating the theoretical framework for this research; that is, students’ understanding of NOS is the outcome of science education, just like their general understanding of science knowledge. Therefore, their understanding should be explained and predicted by their learning. This research hypothesized that students’ views of science have been formed by how they have learned science (learning activities and tasks) as well as how and what learning has been assessed (assessment).

Adopting the natural settings of Canadian and Korean middle school classrooms, data were collected using a survey designed with 5 point Likert scaled items and open-ended questions. The Likert scales were assigned to these close-ended items: 1) 5 items examined \textit{PLAT}, 2) 3 items examined \textit{PAFORMAT}, 3) 4 items looked at \textit{PACONT} and 4) 15 items checked for their understanding of the five NOS concepts (tentative nature of science, subjectivity/ theory laden scientific observations, empirical evidence based science, sociocultural embedded scientific research and diverse/ pluralistic methods of scientific research). Differences and similarities between the two countries and among the seven participating schools were identified using MANOVAs. Two factors were country and school (nested in country). Partial Least Squares (PLS) examined whether these factors could predict the observed variables of NOS. Semi-structured interviews were conducted with 9 volunteer interviewees. The conceptual framework and the results of quantitative data analysis guided the interview questions and the qualitative data analyses.
7.2 Answers to the Three Research Questions

7.2.1 Identification of Students’ PLAT, PAFORMAT, PACONT and Understanding of NOS

Research Question 1: What are the junior middle school students’ perceptions on their learning activities and tasks, achievement assessment and the nature of science?

7.2.1.1 Perceptions of Learning activities and Tasks

The quantitative results of the PLAT and 9 interviewees’ ideas on their science learning were consistent. The quantitative close-ended items covered relatively more participants whereas the small sample of interviewees enriched the contexts and uncovered idiosyncrasies from each classroom (Denzin & Lincoln, 1994). Regardless of country or school, the most frequent science class learning activities were lecture type. Above 60% of students thought memorizing scientific laws and theories were frequent activities while solving questions, student presentations, class discussion and project-based learning were rare. Lab experiments occurred frequently in Korean schools. The lab activities were pre-defined, and students had to follow these procedures, so that little room was left for students’ autonomous learning (refer to an example lab report Appendix, F). As Bell and Lederman (2003) and Clough (2006) argue, such cookbook- style experiments gave students the impression that science experiments were a process to confirm what they learned during classes or a tool to make learning science easy.

As a general trend, science class learning activities and tasks were teacher-directed rather than student-directed. The students’ perceptions of their learning activities implied that they accept knowledge passively. They learned scientific theories, laws and facts without ruminating on them or applying them to different situations, and without experiencing trial and
error in discovering them. More problematic was the finding that students did not value sharing their ideas with their peers. Interviewees regarded class discussions and presentations, though infrequent, as burdensome work, and the opinions of fellow students as unreliable. This might be because so little class time was assigned for these activities.

7.2.1.2 PAFORMAT & PACONT

Participants thought the main method of assessing science learning remains traditional paper-and-pencil tests. Such formats have forced them to choose one correct answer; no alternatives were allowed. To keep assessments fair, interviewees supported diverse ways of evaluation, but some criticized their current science tests. The interviewees believed multiple-choice questions could not evaluate their real abilities because this format emphasizes and measures how much content they remember. This view was similar to what Sambell et al. (1997) reported. In their study, some students thought a test that measured only their memorization was irrelevant. Considering the influence of tests on students’ learning, students have been trained to have the idea that only one answer is correct through 12 years of the formal science education, whether or not teachers intended it.

The Canadian schools and the Korean schools barely included any content of knowledge about science in their science tests, while scientific laws and theories appeared most frequently. Both countries’ science curricula explicitly stipulate the goal of science education is that every student gain a certain level of scientific literacy rather than recruit all students to be scientists. Although there is no complete agreement on what science and scientific knowledge is, scholars of science education, philosophers of science and scientists
agree students should obtain knowledge about science, but in real practice such content was less likely to be taught and assessed.

With students trained for 12 years to choose only one “correct” answer, and with teachers providing no assessment for their knowledge about science, it is ironic to expect such science education to provide any level of NOS understanding.

7.2.1.3 Understanding of NOS

Generally students held relativistic views on science and scientific knowledge. They agreed science is continuously changing due to errors in scientific theories, and due to discoveries using advanced technologies. Those students who accepted the tentative nature of science had a strong view that scientific knowledge has progressed; the newer is better. Yet, despite these progressive views, they did not think scientific knowledge approaches absolute truth. Some other students held scientific knowledge does not change. For them, empirically supported major theories such as Newton’s gravitation law do not change but are refined to be more accurate.

On Theory-laden observation/Subjectivity, the students in this research admitted that the prior knowledge of individual scientists affected their observation, analysis and interpretation of data. In particular, when data were not clear, students felt scientists filled in gaps from their own assumptions and imagination. For this they would draw upon their prior knowledge. Previous research shows that this idea was widely found among groups of professional scientists (Schwartz, 2004; Wong & Hodson, 2009). Not many students in the open-ended question reasoned that the lack of direct observation led to different interpretations. Instead, the students tried to determine why scientists reached different interpretations. They
concluded that either ambiguity in the data or similarities between two natural causes led scientists to different interpretations.

Among the five concepts of NOS, the concept of empirical evidence based scientific knowledge had lower means; 63% of students agreed that scientific knowledge needed to have experimental tests and should be replicable by other scientists. They thought the efforts to find new evidence and the process of trial-and-error inquiry lead to continuous changes in science. The following excerpt well represents student ideas on the role of evidence in changing scientific theory:

Science is an unstable subject to compare with other subjects. Scientific theories, regardless of biology or astronomy, can always be changed by adding discoveries or by removing existing theories due to new experimental results or discoveries. …..So, science can change with high probability by new results of scientific experiment. (6-45)

The majority of the participants agreed there was mutual interaction between science and society, but they thought science has contributed to the development of a society, while politics and religion have abused scientists and science. Students in both countries cited the same historical anecdotes: the conflict between Galileo and Catholic Church, and Darwin’s evolution theory and religious views. In these historic episodes, the final winners were science and scientists. These examples may give students the impression that scientific theories are innately superior to philosophical or religious views. Student-cited examples are discussed further in section 7.3.11.

Participants agreed that scientists should use diverse methods because science covers a broad range of levels from the subatomic to the universal, and there is no known method in new areas. Also, individual scientists’ preferences are different. However, a considerable
number of students stated that scientists should follow universally established methods to render their theories acceptable to other scientists, because universal methods are best.

Most students took their position from either the discovery context or the judgment context and described their opinions on part rather than the whole process. Very few mentioned both. Those students who focused on the discovery context (social demands, motivation and scientists background knowledge), tended to hold more or less relativistic and Multiculturalist views, while students who highlighted the judgment context (the acceptance and eligibility of a theory) leaned toward realist and Universalist perspectives. To these students scientists start with messy methods and finish with a single neat scientific method.

7.2.2 Differences and Similarities

Research Question 2: To what extent are the students' perceptions similar or different in the two contexts?

7.2.2.1 PLAT

Lecture types of learning activities were most frequent in both contexts. The item lacked discerning ability, so it was excluded in the inferential analyses. The country effects on the learning environment were statistically significant in this study, which was consistent with other international studies on learning environments (Aldridge & Fraser, 2000; Dorman et al., 2002; Fraser & Lee, 2009). Both country and the nested factor of the school could explain the differences in the variances of PLAT from 12% and 1%, respectively.

The Canadian students were likely to have more opportunities of solving questions themselves than were the Korean students. Korean students were likely to have more teacher-directed activities or textbook-guided science lab experiments and memorizing scientific knowledge than were Canadian students. Class presentation and discussion activities were the
least frequent activities in both countries but Canadian schools were likely to have more than were Korean schools.

**7.2.2.2 PAFORMAT & PACONT**

Both Canadian and Korean students thought that the traditional paper-and-pencil tests were decisive in their science scores. More than 50% of students from the five schools (Schools 1, 2, 5, 6, and 7) answered that close-ended questions were most significant followed next by short-answer questions. Only the students from School 4 ranked alternative formats as “very significant”.

Canadian students were relatively more likely to consider their science test formats flexible, and close-ended questions less important than were Korean students. For instance, Canadian interviewees mentioned that the test formats differed teacher by teacher. Close-ended questions were easy for teachers to mark, or teachers generously gave some marks by chance so that the marginal students did not fail the course. On the other hand, the Korean interviewees thought that tests should assess students’ learning rather than their luck; therefore, no True or False questions should be in tests. The test formats across the three Korean schools were the similar. Students’ emphasis on paper-and-test implies that most science tests elicit stereotyped, “correct” answers rather than diverse possible answers seen flexibly as “acceptable” or “open.” Section 7.3.4 discusses the unique, neutral responses that about 40% of the Korean students gave to questions on alternative test formats. In addition, a further discussion on Korean schools’ the same test formats was made in Section 7.3.8.

Both of the Canadian and Korean students responded that the content of knowledge about science was rarely assessed in their science tests. Even if such content were assessed, its proportion in tests was negligible. In contrast, knowledge of science such as scientific laws,
facts and theories was assessed frequently. Consistent with Korean students’ perceptions of their learning activities, test content on lab related questions were likely to be assessed frequently.

7.2.2.3 Understanding of NOS

The country effect on the variance differences of understanding of NOS was found across four concepts except for Sociocultural Embeddedness. As mentioned in 7.2.1.3, both Canadian and Korean students took a relativistic stance; however, generally, Canadian students were more likely to hold relativistic positions than were Korean students.

The major difference in Sociocultural Embeddedness was the examples students used to explain their views. Korean students used political examples, and they identified scientific development and national development while Canadian students used stem cell research and a few mentioned global warming.

In the open-ended question on scientific methods, a considerable proportion of Korean students argued that theories be established by universal step-by-step methods in order to win acceptance from other scientists, since universally preferred methods were well established. Only where no method was known in discovering something new, would scientists give up stereotyped approaches. Students probably did not know Kuhn’s (1962) the crisis of normal science, but their explanation about new methods was similar to Kuhn’ view; discovering something new (anomaly) does not fit any of existing methods (paradigm-defined methods), so scientists try new, unknown methods.
7.2.3 Associations

**Research Question 3:** What are the relationships, if any, among students’ perceptions of their learning activities and tasks, their perceptions of the formats and content of their assessments, and their understanding of NOS?

The results of PLS showed $PLAT$, $PAFORMAT$ and $PACONT$ were effective predictors for the variances of students’ understanding of NOS. Taking consideration of the structural models of the theoretical model, almost all the significant inner paths were positive (42 out of 45 paths) except the path from $PACONT$ to *Empirical Evidence Based Science* in the Canadian sample. The positive inner paths implied that the students who perceived their learning activities and tasks as more teacher-directed, emphasizing factual knowledge, were likely to have universalist and realist views of science. As well, students who had the perception that science tests required correct, inflexible answers were likely to understand science is a body of accumulated, universally accepted knowledge. On the other hand, students who perceived their learning activities were more student-directed, and experienced flexible assessment formats emphasizing the application of knowledge to diverse situations, were likely to see scientific knowledge as multicultural, and not absolute truth.

The results of semi-structured interviews supported the predictability of $PLAT$, $PAFORMAT$ and $PACONT$ to the observed variables of the concepts of NOS. Almost all interviewees agreed that how they learn science and how/what the tests assess affect their views of science. They reasoned that school science education is the most important standard to assess knowledge from other resources, and that they gain most of their science knowledge from school education. Students said that they had a variety of sources for science learning, among them television programs, books, the Internet, science journals and magazines.
However none of these sources can compete with the amount of learning, time and effort given to science classes. According to TIMSS Report 2007, Grade 8 students in Ontario have 96 hours of science classes in a year and Grade 8 students in Korea have 104 hours; these are comprised of formal education exclusive of time for homework and self-study. Simply, because of the proportion of time involved, and considering the importance of the content, interviewees felt their formal schooling in science most largely affected their ideas about science.

Two interviewees, who claimed that their views of science were not affected by assessment formats, nevertheless had particularly strong opinions about assessment. One student held the idea that all science knowledge taught in school was not objective truth but pragmatic information that is useful for everyday life. According to her, teachers and educational leaders forced students to learn science because the content could be useful for people’s lives, not because it was true. Tests measured how much a student had learned during classes. As with other subjects, science was not special to her. Another student expressed the view that teachers formulated close-end answers (such as True/False questions or multiple choice questions) out of concern to keep marginalized students from failure. Since there was a 50/50 chance of getting right answers to such questions without actually knowing anything about the subject, these formats helped surface learners pass the course and made marking easier for teachers. To her, assessments filtered and judged students’ abilities in handling the subject matter.

Lederman (2007) says, the study of knowledge about science is “an “input-output” model with little known about the in-depth mechanism that contributes to change in teachers’ and students’ views” (p.860, quotation original). This association model identified the
relationships between students’ learning science, assessment and understanding of NOS could partially clarify that “input-output” model, i.e., how they learn knowledge of science explains how they view knowledge about science.

7.3 Discussion of Major Findings of This Study

7.3.1. Valid Instrument Development

One of the intended contributions of this research to NOS studies was to develop a valid instrument to gauge students’ perceptions of their learning activities and tasks, of their assessment formats and content and of their understanding of NOS. To date no such instrument was available. The field tests of Canadian, Korean and combined samples demonstrated that this instrument reliably predicted the observed variables of the NOS concepts.

The development of the instrument faced several challenges. To examine one domain fully, sufficient numbers of items were required. However, this research covers three domains, and could not obtain sufficient numbers for each domain. Too many items could induce test fatigue in the participants, and result in outcomes of poor quality. The solution was to make two types of survey, dividing the open-ended NOS questions into two sets of three, designing the latent independent variables (the exogenous variables: PLAT, PAFORMAT and PACONT) in formative models, and designing the latent dependent variables (the endogenous variables: the concepts of NOS) in reflective models.

7.3.1.1. Validity and reliability of the instrument

For the reflective measurement models, composite reliability was used. The minimum value was .77 for the Diverse Research Methods of the combined sample, and the maximum values were about .80 across the Canadian, Korean and the combined samples. The internal
consistency of reliability was thus established for the reflective measurement model.

The observed variables for the five NOS concepts reviewed in Chapter 2 were developed to include the meaning of the concepts. Average Variance Extracted (AVE) was used for convergent validity; the Fronell-Lacker criterion and cross-loadings were used for discriminant validity. Across the three samples, all AVE values were greater than .5 in the reflective models. That is, 50% of the variance of an observed variable could be explained by the construct. Applying the Fronell-Lacker criterion, all the AVE values of each latent variable were greater than the squared correlations with all other latent variables; therefore, each concept of NOS shares more variance with its own block of observed variables than with the other latent variables representing different blocks of indicators. In addition to Fronell-Lacker’s criterion, the cross loadings were well established (refer to the tables, 5.5, 5.6 and 5.7).

The formatively designed measurement models were examined for multicollinearity instead of internal consistency reliability since they did not require high correlations among the observed variables (Bagozzi, 1994; Bollen & Lennox, 1991; Chin, 1998; Diamantopoulos, 2006). The Variance Inflation Factor (VIF) values for PLAT, PAFORMAT and PACONT were much less than 10 (the cutoff value) in the three samples. Two outer loadings out of 30 cases were not significant: PLATpresentation (p=.14) and PACONTknowledge (p=.35) for the Canadian and the Korean samples, respectively. Otherwise, all outer weights were significant. The formative measurement models were therefore relatively well established.

The structural models were assessed using $R^2$ of the NOS concepts, the effect size of $f^2$, and the predictive relevance, Stone-Geisser index (predictive relevance) ($Q^2$) and its effect size ($q^2$). The values of $R^2$ ranged from .19 (weak) (Empirical Evidence Based Scientific
Knowledge) to .63 (substantial) (Diverse Scientific Research Methods) for the Canadian sample. The corresponding effect sizes, $f^2$, were from .01 to .65. All values of Stone Gessior Criterion were positive. So the model’s capacity to predict the observed variables of the NOS concepts was established for all three samples.

A few outer loadings were not significant, but when the sample size was large, the coefficients of the loadings became significant. As structure equation modeling methods generally require the number of participants to be 10 times the number of the observed variables, the instrument would be more stable if the sample size proved large enough.

7.3.1.2. Students’ responding pattern (attitudes)

In assessing instrument validity and reliability, the attitudes of students in responding to the questions should be considered. Liang et al. (2009) advise researchers to be cautious with a heavily open-ended response type of questionnaire. Too often, answers left blank, or containing a few words without clear reasoning can undermine the quality of the questionnaire. The instrument used in this research tried to avoid respondents’ test fatigue and to enhance the quality of responses, so the open-ended questions were 3 of each kind, not as many as in previous instruments (VNOS A, B, C and D) (Abd-El-Khalic & Lederman, 1998; Lederman et al., 2002), VOSTS (Aikenhead et al, 1987), SUSSI (Liang et al., 2008).

In the quantitative standardized questions, the missing rates for Canadian and Korean were 8.8% (21 cases out of 238) and 18.4% (72 cases out of 391). For the open-ended questions, the average rates of non-response were 10.6% for Canadian and 17.5% for Korean data. Here, non-response cases did not include responses lacking clear reasons.

Missing cases with a considerable number of blank responses would lower the study’s quality extremely. On comparison, the response rates between close-ended and open-ended
questions were not significantly different. The missing portion of the Korean sample was high. The Korean science teachers accounted for this from circumstances when the survey was administered. The teachers wanted to administer the survey after final exams to minimize the interruption of regular class activities; School 5 and 6 administered the survey in December before the winter vacation and School 7 in late June. Typically, students’ enthusiasm in class drops after final exams. The same happened in School 3, which had the highest missing cases in the Canadian sample. The school data were gathered in June. Even students who wrote down their names to participate in the interview declined to join the interviews. The high rate of missing responses then could have arisen from the time of the survey in the school year.

7.3.2 Gaps between real and preferred learning activities

Gaps were found between what students received and what they preferred in their science classes. In the quantitative data analysis results, the majority of students were involved far less in learning activities where lecture-type learning and memorizing theories and laws prevailed. On the other hand, the interviewees preferred active student involvement to passive learning in science classes. Others (Fraser, 1998; Fraser & Griffiths, 1992; Kim & Kim, 1995) have found similar results. Particularly, when Kim and Kim (1992) examined middle and high school students’ preferred and actual laboratory/classroom environments, students clearly desired a much more open, flexible and hands-on approach to laboratory lessons to what they actually experienced.

Most interviewees assumed that such activities would be helpful to their science learning. And indeed hands-on learning has been recommended as an instructional tool to enhance student understanding of NOS (AAAS, 1993). However, empirical studies (Clough, 2006; Moss, 2001; Ryder & Leach, 1999) did not support the intuitive assumption that doing
science could automatically facilitate understanding science. As Clough (2006) has pointed out, “cookbook”-style lab experiments can neither lead students automatically to authentic learning nor give them an easy grasp of NOS. A Korean interviewee criticized his cookbook-style lab activities that could cultivate neither creativity nor learning but that forced students to imitate teacher or textbook.

The interviewees’ preference for hands-on activities might not guarantee a certain level of understanding. Therefore, these activities need to be accompanied by well-defined connections between curriculum content and hands-on activities. As mentioned in Chapter 2, researchers recommend explicit and reflective approaches to NOS (Abd-El-Khalick & Lederman, 2000; Akerson et al., 2000; Bell et al., 2003; Khishfe & Abd-El-Khalick, 2002).

The following subsection shows the same continuum of perception regarding science lab activities.

### 7.3.3. Cognitive aspects of students’ perceptions on lab experiments

The interviewees of this study preferred hands-on activities for their cognitive aspects. Hodson (2005) defines the cognitive argument for hands-on learning: “practical work assists and promotes conceptual understanding, and is sometimes expressed in terms of ‘what you see and do for yourself, you understand’ or ‘practical work provides concrete reinforcement of abstract ideas’” (p. 168, quotation original). According to the interviewees, when they directly watched and touched something concrete, they could more easily understand the abstract meaning of a scientific theory and they could remember it longer. The direct hands-on activities that students cited as examples were predominantly science lab experiments, science fairs, and class presentations. Unfortunately, these activities were not frequently practiced. Although about 60% Korean students reported that science lab activities were frequent, these
were recipe-style activities, performed or guided by the teacher with minimal student involvement (A sample lab report (Appendix F) predefined the goal, apparatus, procedures and students’ activities).

Reflecting NRC (2000), emphasis has been placed on teaching science as inquiry. This includes teaching skills of scientific inquiry as well as teaching science content through inquiry. And science lab activities are regarded as a way of learning science through inquiry because the activities are assumed to be interesting, to enhance curiosity, and to develop creative thinking and problem-solving ability (NRC, 2000; Shulman & Tamir, 1973). As mentioned above, the interviewees in this research regarded the lab experiments as a tool of learning, not a goal to achieve. In other words, students conceived their lab activities as a process confirming what they already learned by lecture or textbook, enabling them to memorize the content longer and understand the content better than without demonstration or experimentation. The students did not grasp the importance of gaining the skills necessary to pursuing real experiments, which involve a complex process of planning, predicting, observing, recording and interpreting. Most of the students interviewed understood only the cognitive aspects of their experiments.

### 7.3.4 Korean students’ perceptions on lab activity assessment

Taking into consideration students’ perceptions of lab activities as test content, it was found Korean students perceived science lab activities as much more important than did the Canadian students. School 5 and 7 teachers reported that their total science scores had to include at least 30% from alternative formats (e.g., science lab reports); as well, their schools have laid special emphasis on lab activities. This recommendation applied across the school
board could account for significant difference between these Korean schools and Canadian schools.

Uniquely, neutral responses were dominant in relating lab activities to test content. These equivocal answers could be explained by a lack of differentiation in students’ outcomes. With a minimum effort students could get above 70% in lab reports, and there was no functioning differentiation between good and poor works. Therefore, students were likely to think the content of their lab activities in the alternative formats was not quite as important as the proportion of scoring assigned suggests.

7.3.5 Lack of time to ruminate

The interviewees expressed a desire for sufficient time to ruminate about what they had just learned and to integrate their new knowledge with their existing knowledge so as to deepen their understanding and apply it to diverse situations. Quantitative data showed that students did not have enough time to solve problems and questions by themselves although this was improved from 2007 when TIMSS were administered (Martin et al., 2008).

One interviewee pointed out that teachers rushed to cover the curriculum content within a limited time frame, so they asked students questions expecting immediate answers. Acknowledging their intention to transmit as much knowledge as they could within class, teachers then remarked that they could give some homework to students to help them consolidate and understand fully their newly acquired knowledge.

7.3.6 Homework related to the lack of time to ruminate

Following up on the relation between giving students homework and giving them time for their own knowledge consolidation, TIMSS reports (Martin et al., 2008) and the homework
assignment policies of the school boards were consulted. Toronto District School Board Policy P.036 (2008) advises that the amount of homework should be one hour’s worth for Grade 8 students. However, no information about homework guidelines or policy was found in the Busan Metropolitan City Office of Education. Strangely, according to TIMSS reports, Korean teacher survey results report that Korean teachers had a homework policy while Ontarian teachers did not. Korean teachers from School 5 and 7 said that they had guidelines at the school level. The 2007 TIMSS reports showed about 76% of Korean students had homework less than twice a week compared to 63% of Canadian students. In terms of the achievement levels in TIMSS, no difference was found in the Korean data between its high science homework group (552) and its low science homework group (554) while in the Canadian data, the high homework groups achieved higher scores (544) than the low science homework groups (527). According to the TIMSS results, science homework was not effective for Korean students’ learning. Canadian students had more class time to practice what they learned and had frequent homework. Also an interviewee, Teresa, said her homework was useful to prepare science tests.

The push to cover curriculum content within a constrained time frame could provide one reason for the lack of time to consolidate and integrate newly acquired knowledge. Homework did not seem to prove effective for the Korean students. While seeking to save class time, teachers need to pay more attention to help students’ actual learning through effective homework assignments.
7.3.7 PACONTphs in science tests

According to Lederman’s broad review of NOS studies (2007), student understanding of NOS has not reached a desirable level despite worldwide efforts in science education reform. Here a discussion of test content may partially explain this.

The item on the assessment content of social, cultural and historical aspects of science, PACONTphs, was excluded in the inferential analysis due to its lack of discerning abilities; almost all students, particularly almost all Korean students (95%), answered that it was “not significant.” This result implies that most schools did not assess students’ knowledge about science at all. Yet as section 4.4 (research context) notes, the Ontario Science Curriculum (2007) and the Seventh Science Curriculum Reform of Korea both explicitly state what the nature of science is, and why the understanding of nature of science is important. Because such content has not been assessed, students may have gained the impression that knowledge about science is not important; consequently, they probably do not pay it much attention.

The result of the item on the test content of historical, social and cultural meaning of science was consistent with the interviewees’ opinions on knowledge about science. All the interviewees mentioned that learning knowledge of science was more important than learning knowledge about science. The interviewees viewed knowledge of science as directly related to understanding nature, while seeing knowledge about science as a matter of people’s experiences and opinions, which were inaccurate and very tentative. Only two interviewees thought that learning knowledge about science could be helpful to understand scientific knowledge and cultivate critical thinking.

Not only a lack of assessment, but improper assessment could be problematic. Previous studies (Biggs, 2003; Ross & Siegenthaler, 2006) report that a proper assessment enhances
effective learning even though the learning may not be voluntary. Actually assessments are one of the most important stimuli for students to sit at their desks. In this research, however, both the lack of assessment and improper assessment were observed. Although the science curriculum guidelines of Ontario, Canada and Korea stipulate that understanding NOS is as important as understanding scientific concepts, such content was not assessed in science tests. It is, therefore, quite understandable that all interviewees held the knowledge of science to be much more important than knowledge about science.

It was also consistent with views skeptical of teaching knowledge about science (Matthews, 1994; Miller & Osborne, 1998; Turner & Sullenger, 1999). Some educators contend that education is very expensive for individuals as well as for the nation, and class time is of utmost importance. Given the time limits of science class, students should learn the best that current science has achieved. According to their argument, there is no time to waste on anything of secondary or lesser important. For such educators, the prevalent outlook of curriculum designers and teachers rightly values knowledge of science far above knowledge about science.

Hodson (1998) and McComas (2004) have pointed out the significant benefits of NOS in science education. NOS demystifies and emphasizes science as humane work, encouraging more students to study science and to pursue science-related professions. Crucially, it helps instill scientific literacy in the coming generations. However, the responses of the interviewees did not value learning the concepts of NOS. Ironically that is why science education should introduce the concepts of NOS. Of course, without learning knowledge of science, students cannot achieve scientific literacy. Hodson (2001) says, good science education is “a well designed and well constructed curriculum delivered in an appropriate and effective way” (p.3).
He adds that in good science education, gaining knowledge of science, mastering knowledge about science and practicing actual science are harmonized. For the expense incurred, science education should be good.

7.3.8 Same test formats among different schools vs. different formats within the same school

The means of PAFORMAT varied wide ranges among Canadian schools, but the means were similar within the Korean schools. Christina (School 4 interviewee) said that her Grade 7 science teacher used more open-ended and long answering formats while her Grade 8 teacher employed close-ended more. Nancy, the interviewee from the same school with Christina, has different test formats. School 1 science test formats were different from that of School 4. School 1 used more of True/False questions and School 4 used filling the blanks. Depending on the units the proportions would vary. However, the Korean students had exactly the same tests within a school although their teachers were different. Furthermore across the schools, the test formats were the same.

What makes different Korean science teachers employ the same test format? What is the test model? As reviewed in Chapter 3, Korea is a high stake test driven educational environment (Kang, 2005; Kwak, 1998), therefore, the university of netrane examination was referred to. The test format was exactly the same; multiple-choice questions have five choices and there are no True and Force questions. This is the reason that the means across Korean schools were similar. Although there is no rule that all teachers should follow the test format, this result shows an example how a high stake test controls the practice of teaching and learning.
7.3.9 Informed or naïve views on Tentativeness

The students agreed with the close-ended statements that scientific knowledge is subject to change. It was not easy to determine whether they had informed or naïve views based on these results alone. Open-ended responses were therefore reviewed to clarify the vagueness of close-ended responses. According to Lederman et al. (2002), an informed view of tentativeness admits that scientific knowledge is not an absolute explanation of nature. It could be subject to change owing to new interpretations of existing phenomena or new data even though in the past it had the support of evidence and had explained and predicted well.

Responses to the open-ended questions suggested that the main cause for change in science was the imperfection of scientific knowledge for such reasons as inaccurate measuring instruments or the lack of related knowledge. Because very few students explained what errors in science were in, it was important to learn what they meant through the examples they provided (e.g. the atomic theory). In their view, at the initial stage of a theory, scientists do not know much about a certain phenomenon; they make a few hypotheses, and when a particular hypothesis is supported empirically, it becomes an accepted theory. However, such an initial theory may include errors. As science progresses and accumulates further evidence, scientists correct errors in the initial theory, make it more accurate, so that the theory reaches the final stage, as a scientific law.

If one examined whether students agree to the idea of changes in scientific theory, then, students’ views would be thought to be “informed”. But, here, if one examined their reasoning about changes in scientific theory, then a gap emerges between the students and the scholars. Lederman et al (2002) and Liang et al. (2008) see such change in terms of a paradigm shift.
rather than the incremental correction of error or discovery of new data. Students see it otherwise. That is, students could not see how scientists could change their explanations and understanding of nature without the impetus of error correction or further discovery. For instance, some open-ended responses held that a law of science backed by accumulated evidence does not change.

Of course, discoveries and observable evidence are important in science; however, discoveries are interpreted by scientists and integrated into their prior scientific knowledge. Students seemed to miss the element of interpretation in discovery, so that they likely saw empirical evidence as absolute. Previous researchers (Duschl, 1990; Lederman, 2007; Roth & Roychoudhury, 1994) have voiced concern about students’ views of scientific evidence as absolute facts rather than theory-driven results. Some students mentioned that scientific knowledge deals with an objective, physical world that is uniform and stable, unlike humanity’s subjective, mental or psychological worlds; therefore, scientific knowledge once proved as correct does not change and has the power of prediction. This idea that the phenomena of the material world are uniform is akin to the view of inductivists (Hickey, 2005; Skym, 2000); that the future will be the same as the past. This view is more or less close to logical empiricism: a scientific theory supported by consistently observed evidence represents a true description of the relevant natural phenomena.

While close-ended responses reflected the students’ relativistic views of science, a close examination of their open-ended responses revealed gaps between scholarly and student thought, which confirm the concerns raised in previous research.
7.3.10 The unique response from the religion based school about Subjectivity

This question invited students to comment on what scientific observation meant and why their conclusions delivered from the same data were different. The open-ended question used an example of the current discussions on the extinction of dinosaurs. This uncovered a difference between students in public and religious schools. When scientific explanations conflicted with their religious tenets, students from the religion-based school were likely to prefer the teachings of their faith to the ideas presented in science class. On the other hand, no students from public schools mentioned religious explanations.

While the question involved did not address the theory of evolution directly, it was in some ways related to issues between evolution and creationism or intelligent design. This long-running controversy was widely cited in open-ended responses discussing the mutual interactions of science with society, religion and politics. Studies conducted worldwide on the conflict between theories of evolution and creation (Curry, 2009; Deniz et al., 2008) have concluded that there exists in both teachers and students ambivalence, or antipathy toward the theories of evolution or creation. Although no Korean students in this study offered any explanations for the extinction of dinosaurs related to the Bible, other studies from Korea (Ha, Lee, & Cha, 2006; Lee & Lee, 2006) report that a considerable segment of the student population does not agree with the theory of evolution. Other differences between religious and public schools relevant to this study are taken up in Section 7.3.12.

7.3.11 Examples of historical anecdotes

Science, as a human enterprise, can affect and be affected by its societal, cultural, religious and political milieu. The students explicitly accepted this multicultural interaction of scientific knowledge, and used historical anecdotes to support their ideas. Students’ views
largely depend on contemporary perspectives favourable to science. They described conflicts
in the history of science as the dramatic, heroic victories of scientist and science over the
impediments of society, culture or religion. The Roman Catholic Church’s suppression of
Galileo, Isaac Newton’s triumph over the received dogmas of Aristotle, Darwin’s challenge to
European religion and society loomed large in their minds.

Many studies have discussed the advantages and disadvantages of learning the history
of science together with learning science itself (Abd-El-Khalick & Lederman, 2000; Gooday et
al, 2008; Matthews, 2000; McComas et al., 1998). Particularly, Gooday et al. argue that
teaching the history of science gives students the opportunity to read and interpret the primary
sources of scientific knowledge and so develop their abilities for critical thinking. One
interviewee raised this very point, in discussing why Aristotle’s theory of the four elements
had controlled people’s view of nature for nearly two thousand years. He pointed out that all
that time students were presented the theory as an established fact without understanding how
it had been developed. In these circumstances students would simply accept a received body of
knowledge and would have difficulty challenging its authority or thinking critically about it.
This wrong approach to science education ultimately entrenched an incorrect and untenable
theory.

Historians of science warn that scientific anecdotes should be understood in their larger
context, and not treated as isolated and heroic events (Allchin, 2002; Kuhn, 1962, 1970;
scientific knowledge along a continuum of historical events can be beneficial in cultivating
students’ critical thinking, in bridging the two cultures, and in increasing student interest in the
subject. However, the examples the students provided show that they still need to be guided
toward a balanced understanding of the history of science in context. Otherwise, as Allchin (2002) warns, students could get a dichotomous, “black and white” view of science as correct or incorrect.

7.3.12 Discussions about School differences based on MANOVA results

The school effect nested with the country effect was significant for variances of PLAT, PAFORMAT, PACONT and across four concepts of NOS except for Sociocultural Embeddedness. These results raised two questions: “What are the reasons for the school difference within the same country?” and “Can the cause be systemic or teacher-based? Two obvious differences among schools within a country existed: Religious based versus public schools in Canada and coeducational schools versus boys-only schools in Korea. Using LMATRIX a contrast was conducted to test whether these differences were systemic.

7.3.12.1 Comparisons between Religious and Public Schools in Canada

School 1, which participated in this research, was a Christian school. It provided a unique response to an open-ended question in this research. Similarly, other previous research has reported how different kinds of schools contribute to differences in student understanding of NOS. Ethnic-centered schools, academic or vocational schools (Tuncer et al., 2005) and literature focused vs. science-focused schools (Dogan, 2011). Haidar and Balfakih (1999) also showed that religion and culture affected students’ views on NOS. Therefore, this research tried to examine whether there existed differences or similarities between a religion-based school and the public schools across the constructs.

Before making these comparisons, two cautions need to be given. The Christian school in the study was just one among thousands of schools of all kinds, so it would distort reality to over-generalize the study’s findings beyond its research context. Furthermore, as with other
results of these analyses, no value judgments were applied to the results since learning and teaching are complex human endeavors. Each learning environment has its own idiosyncratic features. Consequently it is wrong to make rash judgments that either public or religious schools are superior one to the other.

A contrast (LMATRIX) on PLAT showed that School 1 was significantly different from the other three Canadian schools. Comparisons among the three public schools were not significantly different. Learning in School 1 was more teacher-directed than student-directed. Schools 1 and 2 were very similar in their students’ perceptions of test formats and content. Schools 3 and 4 differed from each other for PAFORMAT, whereas all schools resembled each other for PACONT. Schools 1 and 2 showed no significant difference of means in understanding of NOS. In fact, across all five constructs, these two schools did not show any significant differences. In contrast, while Schools 3 and 4 were much alike they differed with each other in the Empirical Evidence Based Scientific Knowledge and Diverse Scientific Research Methods.

These results indicate that while differences existed among schools, these differences could not be traced to the kind of school but to factors unique to individual schools. The results of Schools 1 and 2 were very similar to each other, and School 2 was statistically significantly different from the other two public schools. Additionally, School 3 and 4 were similar in PACONT and Tentativeness, Subjectivity and Sociocultural Embeddedness while they were different in other construct. Unlike other studies (Dorgan, 2011; Haidar & Balfakih, 1999; Kilic et al., 2005), this research found that any differences between a religion-based school and the public schools in students’ perceptions of their learning science, assessment and the concepts of NOS were not significant.
7.3.12.2 Comparisons between Boys-Only and Co-educational Schools in Korea

All Canadian schools and two Korean schools were for both boys and girls, but School 5 was for boys only. Generally School 5 had lower means than School 7. The data analyses identified a gender effect. Of course, the comparison among Schools 5, 6 and 7 cannot strictly speaking reveal any gender difference since Schools 6 and 7 were for both genders and not for girls only. Because country was a significant factor, this comparison was kept within the Korean schools to avoid a compounding country effect.

As with the comparisons between the public and religion-based schools, the univariate LMATRIX was used to contrast School 5 with School 6 and 7, and School 6 with School 7 across all the constructs. School 5 was different from School 6 in PAFORMAT and PACONT out of the total 8 constructs. School 5 also differed from School 7 in Tentativeness, EMP and PACONT. Meanwhile, there also were differences between the both gender Schools 6 and 7, in PAFORMAT and Empirical Evidence Based Scientific Knowledge.

Comprehensive research has reported gender differences in learning style preferences (Baron-Cohen, 2003; Bridgeman & Lewise, 1994; Glazer, 2005; Sanchez & Wiley, 2010); in particular, boys excel over girls in taking multiple-choice format tests (Bridgeman & Lewis, 1994). Some studies (Baron-Cohen, 2003; Browne, 2002) have even argued that children are predisposed to these disparities from birth. Korean studies on student perceptions of learning environment and of interpersonal interaction with teachers have uncovered gender differences across all examined constructs (e.g. Kim, Fisher & Fraser, 2000). NOS studies have revealed similar differences between boys and girls (Kilic et al., 2005).

Differences found in this study could derive from causes other than gender, since no notable difference of construct emerged between the “boys only” and co-educational schools,
and yet other differences were discovered among co-educational schools. Therefore gender-based differences could not account for variance differences found on $PLAT$, $PAFORMAT$, $PACONT$ and understanding of NOS.

### 7.3.13 Teacher Effects on Students’ Understanding of NOS

The MANOVA results found that school differences significantly affected students’ understanding of NOS and $PLAT$, $PAFORMAT$ and $PACONT$. The contrast analyses tested two potential causes: public verses religious, and co-educational versus gender-specific. But these two facets did not strongly support school differences. Of course, there could be many reasons why one school differs from another: the school could be public or religious; the school could be co-educational or gender-specific; the school could have an advantageous location, etc. Teacher differences can be a further cause of difference. Unfortunately, teacher data was too limited to examine whether differences of teachers determined school differences. The present research cannot discuss the relationships between teacher differences and students’ understanding of NOS.

### 7.3.14 Partly Mediated Effects of Assessment on Learning Activities

Based on the previous research results (Dorman, 2001, 2003; Dorman et al., 2002; Dorman Fisher, & Waldrim, 2006; Fraser, 1999), the theoretical model for this study was postulated upon the assumption that students’ perceptions of assessment formats and content affected their perceptions of learning activities and tasks ($PLAT$). Although the inner coefficients were significant, the meaning of the paths were not discussed in Chapter 5 since the effects of test content on the learning activities mediated by the test formats were not central concerns for this research. In this section of the discussion, the significant paths among $PLAT$, $PAFORMAT$ and $PACONT$ were examined and how $PAFORMAT$ mediated to predict
PLAT by PACONT. The full discussion of the effect of PACONT and PAFORMAT on PLAT was on the combined sample not for the Canadian and Korean samples. Because the path from PACONT to PLAT was not significant for Canadian sample and the path from PACONT to PAFORMAT was not significant for the Korean sample. Figure 7.2 represents this model. PAFORMAT is mediated between PACONT and PLAT.

Figure 7-1. Partly Mediated Causation Effect Between PACONT & PLAT
(PAFORMAT is a mediator) Model

The initial examination of PLAT indicated that there were positive correlations; (r=.41: PLAT and PAFORMAT, r=.24: PLAT and PACONT). The indicator variables, PAFORMAT and PACONT, r=.205 (p<.01) did not indicate high multicollinearity (VIF=1.04). In the multiple regression to identify the effects of PAFORMAT and PACONT on PLAT, R=.44 and $R^2=.19$, which meant that about 19% of the variance in PLAT could be predicted by PAFORMAT and PACONT. The adjusted $R^2$ was .19. The overall regression was statistically significant, $F(2, 533)=62.31$, p<.001. Complete results for the multiple regression are presented in Table 7.1.
Table 7-1

Results of the Standard Multiple Regression to Predict PLAT (Y) from PAFORMAT (X₁) & PACONT (X₂)

<table>
<thead>
<tr>
<th>Variable</th>
<th>PLAT</th>
<th>PAFORMAT</th>
<th>PACONT</th>
<th>b</th>
<th>β</th>
<th>sr² unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAFORMAT</td>
<td>.41**</td>
<td></td>
<td></td>
<td>.31**</td>
<td>.38</td>
<td>.13</td>
</tr>
<tr>
<td>PACONT</td>
<td>.24**</td>
<td>.21**</td>
<td></td>
<td>.12**</td>
<td>.16</td>
<td>.02</td>
</tr>
<tr>
<td>Means</td>
<td>2.33</td>
<td>2.25</td>
<td>2.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.74</td>
<td>0.88</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N=536

$R^2=.189 \quad R_{adj}^2=.186 \quad R^2=.435 \quad ** p<.001$

PAFORMAT was significantly predictive of PLAT when the variable PACONT statistically controlled: $t(533)=9.40, p<.001$. The positive slope for PAFORMAT as a predictor of PLAT indicated that about .38 points in PLAT increases one point of PAFORMAT, controlling for PACONT. As PAFORMAT and PLAT were arranged as low scores meant more inclination of one correct answer required formats and teacher-centred learning activities while high scores meant more flexible test formats and student-centred learning activities. The squared semi-partial correlation that estimated how much variance in PLAT was uniquely predictable from PAFORMAT was $sr^2=.13$. About 13% of the variance in PLAT was uniquely predictable from PAFORMAT when PACONT was statistically controlled.

PACONT was also significantly predictive of PLAT when PAFORMAT was statistically controlled: $t(533)=3.96, p<.001$. The slope to predict PLAT from PACONT was approximately $\beta=.16$; in other words, there was 0.16 point increase in the standardized PLAT scores when PACONT increased 1 standardized point. The $sr^2$ for PACONT (controlling for PAFORMAT) was .02. Thus, PACONT uniquely predicted about 2% of the variance in PLAT when PAFORMAT was statistically controlled.
The conclusion from this analysis was that the original zero-order correlation between \textit{PAFORMAT} and \textit{PLAT} ($r = .41$) was partly (but not entirely) accounted for by \textit{PACONT}. When \textit{PACONT} was statistically controlled, \textit{PAFORMAT} uniquely predicted 13\% of the variance in \textit{PLAT}. And when \textit{PAFORMAT} was statistically controlled, \textit{PACONT} uniquely predicted 2\% of the variance in \textit{PLAT}. This result could be understood as \textit{PAFORMAT} and \textit{PACONT} might be partly redundant as predictors of \textit{PLAT}. However, each predictor was significantly associated with \textit{PLAT} (refer to Table 7.1) even when the other predictor variable was significantly controlled; both \textit{PAFORMAT} and \textit{PACONT} contributed uniquely useful predictive information about \textit{PLAT} in the combined sample.

The predictive equation for the combined sample was as follow:

\[
\hat{\text{PLAT}} = 1.33 + 0.31 \times \text{PAFORMAT} + 0.12 \times \text{PACONT}
\]

In addition to the \textit{PAFORMAT} mediation effect on \textit{PLAT}, the two test constructs’ mediation effects on the concepts of NOS were also significant. For instance, when the effects of the exogenous variables on \textit{Diverse Scientific Research Methods}, the coefficients of \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT} were .37, .31 and .15, respectively. SmartPLS provided the total effects of the constructs, and the values were .37, .44 and .31. For \textit{PLAT} the coefficient and its effect were the same but for \textit{PAFORMAT} and \textit{PACONT} those values were increased. The results meant that the mediation effects of these two constructs indirectly affect students’ understanding of NOS. Another example is the case for \textit{Subjectivity}. Three inner coefficients for \textit{PLAT}, \textit{PAFORMAT} and \textit{PACONT} are .23, .31 and .37, respectively. The total effects for the constructs were .23, .43 and .49, respectively. As these examples showed, two constructs of test directly and indirectly affected students’ understanding of NOS.
7.4 Conclusions

This study examined students’ perceptions of their learning activities and tasks (PLAT) and of assessment formats (PAFORMAT) and content (PACONT) and their understanding of the Nature Of Science (NOS). Attention was paid to the relationships among these constructs to identify factors influencing students’ understanding of NOS in formal science education. A survey was administered to 217 Canadian and 319 Korean grade 8 students. In most constructs, the country and school effects accounted for the variance differences significantly. Also, the perceptions could effectively predict students’ understanding of NOS. Different data resources such as the quantitative and qualitative data from the survey and semi-structured interviews and other auxiliary data resources were congruent.

One of the significant contributions of this research to NOS studies can be developing a reliable and validated instrument. The measurement and structural models were consolidated when the sample size was over 500. When sample sizes were small, one outer loading from each country did not significantly contribute to the targeting construct. Otherwise, convergent and discriminant validity were both satisfied. All indexes such as AVE, Fronell-Lacker criterion, VIF, and the predictive relevance met the requirements. The composite reliabilities for the reflective constructs across the three samples were greater than .70. Therefore, the instrument can be used to identify the relationships among constructs, PLAT, PAFORMAT and PACONT and the NOS concepts, for middle school students.

The quantitative and qualitative results indicated that there existed differences between Canadian and Korean students in PLAT, PAFORMAT and PACONT and understanding of NOS. There were school effects on the differences of the examined constructs, but the country effect was greater than the school effect. Along with other cross-cultural international studies,
this study provides further empirical evidence that students from different contexts perceive science learning, assessment and NOS concepts differently.

The third significance of this research is whether PLAT, PAFORMAT and PACONT can be relevant to predict the variables of NOS. Across three different samples (Canadian, Korean and the combined samples), these constructs can predict the variances of the observed variables of NOS. Using all the three constructs, the variances were predicted from 19% to 63% for the Canadian sample, from 28% to 47% for the Korean sample, and from 27% to 46% for the combined sample. This result implies not only how to teach but also how and what to assess should be considered to enhance students’ understanding of NOS. The signs of the weights are consistent with the argument from the previous studies (Duschl, 1990; Hodson, 1998; Kuhn, 1970; McComas, 2004; Roth & Roychoudhury, 1994). Learning activities that demand students to memorize factual knowledge and scientific theories and laws, and assessments that force them to choose only one correct answers, are highly likely to give the impression that science is a body of knowledge so proven by empirical evidence that they are approximate truth. On the other hand, when students are exposed to the activities of trial-and-errors, exploration of new methods, discussion about scientific controversial issues and the assessments allow flexible and pluralistic answers, they are likely to think that science is a continuous process and effort to understand nature.

7.5 Recommendations for Future Research

Schwab (1962) pointed out that scientific knowledge does not consist of “facts”, but rather of interpreted facts. Through the process of interpreting data, individual scientists’ background knowledge and cultural norms became involved. Without understanding the values and assumptions inherent in such interpretation, students can have the image of science
as a collection of isolated facts (Schwab, 1962; Hodson, 2008; Lederman, 2007; Lederman et al., 2002; McComas & Olson, 1998). Current science education efforts encourage the teaching of NOS concepts within various disciplines of science in order to develop a disciplinary-based knowledge of subject matter. However, Lederman (2007) contends that learning how conceptual principles were involved in the interpreted facts of scientific knowledge is very broad and vague.

Regardless of the difficulty in clarifying the model, this research recommends the following for science education (1~3) and for future studies (4~6):

1) Class activities and tasks should encourage student involvement;

2) Science lab activities should allow students to explore new methods and challenge the existing ideas rather than cookbook style activities

3) All science assessments should include the content of knowledge about science;

4) A collaborative study is needed to assess knowledge about science to provide teachers with usable content and rubrics;

5) A sizable concept of NOS is needed for teachers to apply their teaching and learning activities; and,

6) More studies need to generalize the relationships between assessment formats and the content and understanding of NOS.

The results of this research showed that the inner coefficient of \( PLAT \) directed to the concepts of NOS were high. This meant \( PLAT \) was a good predictor to the variance of the concepts. Thus, learning activities should encourage student involvement rather than remain largely teacher directed. This recommendation is consistent with what Roth and Roychoudhury (1994), Duschl (1990) and Lederman (2007) argue. Teachers’ content
knowledge on the concepts of NOS must be enlarged and teachers’ pedagogical content
knowledge and willingness to deliver the knowledge to his or her class must be encouraged.
NRC (1996) also recommends the shift of teachers’ roles in classrooms from final authority to
guide and facilitator. Through engaging in content-embedded inquiry activities, students are
to develop conceptual understanding of subject matter, scientific inquiry and NOS.

Reynolds, Doran, Allers, and Agruso (1995) argued that for an effective learning to
occur, congruence must exist between instruction, assessment and outcomes. The results of
this study demonstrate that teaching the concepts and doing science are not enough. There
should be a proper assessment of what has been taught. With a flexible format, content needs
to be assessed. If knowledge about science is assessed as well as knowledge of science, not
only teachers but also students would pay more attention to learning that knowledge, though
the learning may not be willingly. Here, how to assess is an issue. Korean students’ assessment
of alternative formats was a notable example how students’ abilities should not be assessed.
The portion of lab reports was 30% of their test scores, but the format did not have discerning
abilities of how a student performed and reported the activities. In such cases, the assessment
cannot play a proper role that enhances learning. Therefore, rubrics for how to assess and what
to assess on knowledge about science need to be developed.

This raises yet another problem: “Who should develop these rubrics?” Research shows
that science teachers’ understanding of NOS has not reached a desirable level (Lederman,
2007). Plus, it would involve too great a workload to expect science teachers to develop
guidelines. As Pedretti (2003) study showed, there were gaps between what a science teacher
knew and what he/she practiced in the classroom when the science teacher should cover a
heavy amount of content within a given time. A collaborative study on NOS and on
assessment needs must provide a practical guideline for the assessment of knowledge about science, which can be used in science class with little or no time adjustment required.

The next issue at variance is: “How, and to what extent, can NOS concepts be adopted in science education?” Without a clear definition and scope for NOS, developing standards and assessing content would be jumbled. The literature related to NOS has revealed that too broad, philosophical definitions did not help teachers apply NOS concepts in practice. Therefore, a manageable guideline, tightly related to the science curriculum, should be developed. If guidelines are not properly developed, misconceptions will be reinforced and knowledge of NOS concepts will not be improved.

Furthermore, novel relationships between students’ perceptions on their assessment formats and content and their understanding of NOS were reported in this study. More evidence is required to generalize these relationships beyond the contexts of this research. A single study cannot provide solid evidence for a causal relationship. So more studies on the relationship will accumulate further evidence. This study has established that the assessment of NOS concept can enhance student understanding of NOS; specifically, diverse test formats which move away from the traditional selection only one “correct” answer can enlarge students’ views of science. But to prove a causal relationship, more research is required.

To forward the development of scientific literacy thorough enhanced understanding of NOS, reform advocates recommend scientific inquiry experiences as a context for learning. Philosophy of science also advocates contextualized scientific knowledge. Thus, subjects must be inquiry-embedded (AAAS, 1990, 1993; NRC, 1996; Schwartz, 2004). The argument for the use of inquiry asserts that engagement in active scientific inquiry along with actual scientists,
or in a manner similar to actual scientists, will help learners develop an understanding of the scientific process, and in turn an understanding of NOS.

The intuitive assumption that if students were involved authentic and doing science they would naturally get the concepts of the nature of science does not in fact have strong support from empirical evidence. Moss (2001), Ryeder and Leach (1999) and Kishfe and Abd-El-Khalick (2002) argue that both an authentic science learning environment and an explicit instruction and reflection opportunity are effective approaches to enhance students’ understanding of NOS. For instance, regarding instructional methods, Lederman (2007) summarizes “Conceptions of NOS are best learning through explicit, reflective instruction as opposed to implicitly through experiences with simply “doing” science (p. 860, quotation original). As with these studies, this research showed authentic experiences in science-participation in scientific research experience (Kishfe & Abd-El-Khalick, 2002), and an inquiry-focused science curriculum (Meitry, 1992) were not very effective in shaping students’ or teachers’ understanding of NOS. Therefore, science educators need to pay more attention to designing experiments and making students understand that empirical evidence has been initiated with a theory, and depending on the theory the evidence can be interpreted differently.

7.6 Limitations of the Study

This study has several limitations, which inhibit any external generalization. First, the samples were not randomly selected. They were recruited from convenient sampling procedures, so they are not representative of the whole population of Canadian and Korean middle school students. Both Canadian and Korean samples were from urban cities (Toronto is the largest city in Canada, and Busan is the second largest city in Korea). Furthermore, the interviewees were volunteers from the survey participants, and not randomly selected.
Particularly, the Korean interviewees were outstanding students in science. Two were volunteers, and three more were teacher-recommended. Four of these five aimed to enter science-focused high schools, which means they were within the upper 1% of students in science and other subjects. Due to this lack of representative sampling, caution must be exercised in generalizing the research results and findings beyond the samples.

In addition, this research was conducted in natural settings, not with an experimental design. The effective factors identified in this research were good predictors for understanding of NOS but the relationships between the factors and the understanding were not causal. Warner (2008) advises researchers of some conditions when she/he argues a causal inference between two variables: 1) the variables should be systematically associated, 2) the cause variable should precede the effect variable (chronological order of events), 3) there must not be any other variable confounded with the treatment variable, and 4) there must be a reasonable theory to predict or explain a cause-and-effect relationship (p. 17). In the present study, variables may be confounded through its non-random sampling procedures. Student perceptions could be affected not only by school education but also by influences outside school, such as media, access to knowledgeable adults, books and the Internet. Thus, the factors considered in this study are good predictors for students’ understanding of NOS but they may not be the causal factors.

Third, the limited number of items in the measuring instrument could not cover the complexity of the learning and teaching enterprise. It could catch in rough outline the features of class activities and assessments, but it could not sift through their subtleties, nor through the idiosyncrasies of individual schools. To catch a detailed description of a classroom, more items and interviewees are needed. As aforementioned in the methodology chapter, this
research covered several different domains (students’ perceptions of learning activities and
tasks, assessment formats, assessment content and the concepts of NOS) and each domain
required a sufficient number of items. Too many items could provoke test fatigue, blank
answer sheets, or insincere responses. In addition to the small number of items in each domain,
more interviewees are required from each school. The nine interviewees involved were from
only four schools; no interviewee came forward from the other three schools.

Finally, the interpretation of the open-ended responses and semi-structured interviews
was inevitably influenced by the researcher’s views. All interpretation is indeed theory-laden,
and this research could not be an exception. The author is confessedly the outcome of science
education in the post-Sputnik era. Korea’s educational system has been strongly affected by
US system and curriculum, which took heavily empirical positivistic perspectives of science.
Regardless of personal efforts to adopt a relatively neutral stance, the researcher’s own
knowledge and experience in science education has inevitably contributed to frame the study
findings. These probably favor a Universalist approach.
References


http://www.ngspscience.com/profdev/Monographs/SCL22-0449A_SCI_AM_Bell_lores.pdf


Chin, W. W., Marcolin, B. L., & Newsted, P. R. (1996). *A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and voice mail emotion/adoption study*. Paper presented at the 17th International Conference on Information Systems, Cleveland, OH.


Cobern, W.W., & Loving, C.C. (2001). In defense of realism: It really is commonsense. *International history, philosophy and science teaching group meeting, USA, 162*.


Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. Journal of Marketing Research, 18(3), 328–388.


Reshaping Undergraduate Science and Engineering Education: Tools for Better Learning, p. 53-64.


Appendix A Survey

Students’ perceptions of the learning activities and tasks, and of assessment formats and content in science education, and their understanding of the nature of science.

Name: ________________________________________ (if you want to participate in this research as an interviewee, please write down your name; if not, your name is not required)

Are you a boy ( ) or girl ( )

Date: ________________________________________

Which two subjects do you like the most? 1) ____________ 2) ____________

Instructions

I am interested in your experience in science class, with science tests, and what you think about science. Remember, there are no “right” or “wrong” answers to these questions, there is only your opinion. Thank you.

A. What is Your Experience in Your Science Class?

<table>
<thead>
<tr>
<th>Your experience of science classroom activities: How often do you do these things in your science class?</th>
<th>almost always</th>
<th>often</th>
<th>sometimes</th>
<th>seldom</th>
<th>hardly ever</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I listen to the teacher explaining the lesson or copy out what she/he writes on the blackboard.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I solve questions using formulas or scientific laws by myself with the teacher’s help.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I do experiments following teachers’ guidance and my</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
B. What is Your Experience with Science Tests?

<table>
<thead>
<tr>
<th>Which types of questions do you think most significantly affect your science test scores?</th>
<th>1 most important</th>
<th>2 important</th>
<th>3 moderate</th>
<th>4 little important</th>
<th>5 least important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>True or false questions and multiple choice questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Write answers by your own (short answers: a few words or one or two sentences)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Presentation, discussion or an essay about scientific issues or projects.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please list how often you get the following kinds of questions in your science tests

<table>
<thead>
<tr>
<th>Please list how often you get the following kinds of questions in your science tests</th>
<th>1 almost always</th>
<th>2 often</th>
<th>3 sometimes</th>
<th>4 seldom</th>
<th>5 hardly ever</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
III. What Do You Think About Science?

For each statement in this section check off whether you

- STRONGLY AGREE (SA)
- AGREE (A), OR
- DON’T HAVE AN OPINION (N),
- DISAGREE (D),
- STRONGLY DISAGREE (SD),

Tentative Nature of Scientific Theories

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A new interpretation of data can change our present scientific knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. We accept an idea as scientific knowledge only if it does not have any error.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. As scientific knowledge develops, it will approach to absolute truths.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
After scientists develop a scientific theory (for example, cell theory), does the theory ever change? If you believe that “yes, they do” then explain why we bother to learn scientific theories. If you believe that “no, they don’t” … Explain your answer by using examples (VNOS- B, item 1)

Science textbooks often represent the atom at a central nucleus composed of protons (positively charged particles) and neutrons (neutral articles) with elections (negatively charged particles) orbiting that nucleus. How certain are scientists about the structure of the atom? What specific evidence, or types of evidence, do you think scientist used to determine what an atom looks like?

Observing Things in Science

<table>
<thead>
<tr>
<th>4</th>
<th>Scientists’ observations of the same event may be different because the scientists’ prior knowledge may affect their observations.</th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Scientists’ observations of the same event will always be the same regardless the related theories change because observations are facts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Scientists’ observations of the same event will the same because scientists are objective.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is believed that about 65 million years ago the dinosaurs became extinct. Two different theories explain this; both enjoy wide support. The first, formulated by one group of scientists, suggests that a huge meteorite hit the earth 65 million years ago and led to a series of events that caused the extinction. The second hypothesis, formulated by another group of scientists, suggested massive and violent volcanic eruptions were responsible for the extinction. How are these different conclusions possible if scientists in both groups have access to and use the same information to derive their conclusions?

Empirical Evidence Based Scientific Knowledge
Among different scientific theories, the acceptance of scientific theory entirely depends on experimental evidence.

An experimental test is not mandatory for scientific knowledge.

Scientists invent scientific knowledge, so it does not need observable evidence.

What makes science (e.g., physics, biology, chemistry) different from other subjects that require investigation and experimentation? (VNOS-C item 1)

Science is not influenced by cultural, societal values because science is independent from those values.

Scientific knowledge is universal because it applies everywhere.

The values and expectations of a culture determine the research scientists’ conduct and the results they accept.

Do you think social and cultural values decide what scientists should work on and what kinds of knowledge they accept as scientific knowledge?
### SCIENTIFIC METHODS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Scientists use different types of methods to conduct their research.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Scientists follow the same step-by-step scientific method.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>When scientists use the scientific method correctly, their results are true and accurate.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**O3 type B)** With examples explain whether scientist follow a single, universal scientific method Or use different methods? For instance, the universal method means such steps as define the problem→ gather information→ form a hypothesis → make relevant observations → test the hypothesis → form conclusions → report results.
설문지

이 설문지는 과학 학습활동과 과제, 시험 방식과 내용, 그리고 과학본성에 대한 여러분의 생각을 묻는 것으로 3 부분으로 이루어져 있습니다. 질문에 따라서 동의 여부를 묻거나 얼마나 자주 발생하느냐를 묻고 있습니다.

이름: ___________________ (만일, 인터뷰에 참여할 의사가 있으면 이름을 적고, 그렇지 않다면 이름을 적지 않아도 됩니다.)

남(     ) 여(    )

날짜: 년 월 일

줄어 하는 두 과목은 무엇입니까? 1) _______ 2) _______.

질문의 내용에 대한 여러분의 생각, 의견이 무엇인가를 대답하는 것으로 올고 물린 답은 없습니다.

학습 활동과 과제에 대한 질문

<table>
<thead>
<tr>
<th>번호</th>
<th>과학 시간에 하는 학습활동: 아래 활동이 얼마나 자주 여러분의 과학 시간에 일어남니까?</th>
<th>항상</th>
<th>자주</th>
<th>때때로</th>
<th>가끔</th>
<th>전혀</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>나는 수업시간에 선생님의 설명을 들거나 철반에 적은 내용을 공책에 적는다.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>나는 과학 공식이나 법칙을 이용해서 내 스스로 또는 선생님의 도움을 받으면서 문제를 풀다.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>나는 선생님께서 지시한 대로 또는 시범을 보여 주시는 대로 실험을 한다.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>과학 선생님께서 수업시간에 여러분에게 무엇을 하라고 요구하십니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>선생님께서는 나에게 과학적 사실이나, 이론 법칙을 알려주라고 요구하신다.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>선생님께서는 나에게 과학에 관한 주제에 대해 발표하거나 또는 과학 프로젝트 (신문</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


과학 시험에 대한 질문

<table>
<thead>
<tr>
<th>어떤 시험문제 유형이 가장 과학 시험점수를 좌우 합니까?</th>
<th>아주 중요</th>
<th>중요</th>
<th>보통</th>
<th>조금</th>
<th>전혀</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 오지연형</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 용어를 적거나, 한 두 문장으로 대답할 수 있는 주관식 문제.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 발표, 토론, 서술형, 실험 수행, 실험보고서 등의 형태</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>어떤 내용의 시험문제가 가장 과학 시험에 자주 출제 됩니다?</th>
<th>항상</th>
<th>자주</th>
<th>때때로</th>
<th>간혹</th>
<th>전혀</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 과학적 사실, 이론, 법칙에 대한 문제</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 과학 시간에 배운 것을 응용하는 문제</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 실험에 관한 문제</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 과학이 어떻게 사회, 우리 생활에 영향을 미치는지? 또, 과학이 어떻게 발달되었는지 등에 대한 문제</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

과학의 본성에 대한 설문

아래 설명에 대해 여러분의 동의 정도를 묻습니다.

1. 적극적으로 동의함,
2. 동의함,
3. 보통,
4. 동의 하지 않음,
5. 전혀 동의 하지 않음.
과학적 이론의 가변성

| 주 1 가 유형 | 어떤 과학이론이 (예, 원자설) 제안, 발전된 후, 그 이론들은 변화 할 수 있다고 생각하는가? 만일 변화한다고 생각한다면, 왜 과학 선생님들이 변화하는 과학 이론들을 학생들에게 가르쳐야한다고 생각하는가? 이유를 예를 들어 설명하라. |
| 주 1 나 유 | 과학 교과서에 중간에는 원자핵이 양성자와 중성자가 있고 그 주위에 음전기가 떠고 있는 전자들이 있는 원자 구조에 대한 그림들이 종종 보인다. 과학자들은 어떻게 원자구조에 대해 확신을 하는가? 어떤 특정한 증거를 과학자들이 이용하여서 원자의 모습이 그와 같다고 생각하는가? |

과학적 관찰에 대해서

| 주 2 가유 | 과학자들에 의하면 6500 만년 전에 지구 상에 공룡들이 멸종을 하였다고 합니다. 두 가지 다른 가설들이 공룡 멸종을 설명하고 있고 두 가설 모두가 광범위한 지지를 받고 있습니다. 첫째 가설을 지지하는 과학자들은 6500 만년전에 엄청나게 큰 운석이 지구를 떨어져 일련의 지구 환경 변화를 초래하여 공룡이 멸종하였다고 합니다. 두 번째 가설은 거대하고 활발한 화산 폭발이 공룡 멸종을 초래하였다고 본다. 두 가설을 주장하는 과학자들은 같은 자료를 바탕으로 하였는데, 어떻게 다른 결론들이 가능했는지 볼까? |

| 4 | 똑같은 자연현상을 과학자들 사이에서 다르게 관찰될 수 있는 이유는 과학자들이 가지고 있는 기존 지식이 관찰에 영향을 미치기 때문이다. | 1 2 3 4 5 |
| 5 | 관련된 이론이 변하더라도 똑같은 자연현상을 관찰한 것은 항상 같다.왜냐하면, 관찰은 사실이기 때문이다. | 1 2 3 4 5 |
| 6 | 똑같은 자연현상을 관찰한 것은 항상 같다.왜냐하면, 과학자들은 아주 객관적이기 때문이다. | 1 2 3 4 5 |
증거에 바탕을 둔 과학적 지식

<table>
<thead>
<tr>
<th>7</th>
<th>여러가지 과학 이론들 중에서 인정되는 과학이론은 실험적 증거가 뒷받침된 것들이다.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>과학적 지식은 실험을 필수 요건이 아니다.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>과학자들이 과학 지식을 만들어 낸 것이기 때문에, 과학지식은 관찰 가능한 증거가 필요하지 않다.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>주3</td>
<td>과학 (물리, 생물, 화학 등)과 다른 학문의 다른 점은 무엇입니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>가유형</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

과학과 사회, 문화와의 관련성

| 10 | 과학은 문화적, 사회적인 가치나 기대에 영향을 받지 않는다. 왜냐하면, 과학은 이런 것들에 의사의 독립적인 것이기 때문이다. | 1 | 2 | 3 | 4 | 5 |
| 11 | 과학적인 지식은 보편적인 것이다. 왜냐하면, 모든 곳에 독립적 적응되기 때문이다. | 1 | 2 | 3 | 4 | 5 |
| 12 | 문화적 가치나 사회적 기대가 과학자들이 무엇을 연구하고 그 결과를 받아들일지 결정한다. | 1 | 2 | 3 | 4 | 5 |
| 주2 | 어떤 학자들 중에는 과학이 사회 문화 정치적 가치가 과학자들이 무엇을 연구하고 받아들일지 결정한다고 생각하십니까? | |
| 나유형 | 여러분의 생각은 어떤지, 예와 함께 자신의 생각을 응호할 수 있는 설명을 적으시오. | |

과학적인 방법

| 13 | 과학자들은 서로 다른 방법을 이용하여 연구를 한다. | 1 | 2 | 3 | 4 | 5 |
| 14 | 과학자들은 같은 방법인 일련의 조사 연구를 실행한다. | 1 | 2 | 3 | 4 | 5 |
| 15 | 과학자들이 올바른 방법을 이용하여 연구한다면, 그 결과는 정확하고 | 1 | 2 | 3 | 4 | 5 |
| 주 3 (나유형) | 과학자들이 보편적인 유일한 과학적인 방법 (예를 들자면, 1. 문제 설정, 2. 관련 정보 수집, 3. 가설 설정, 4. 유효한 관찰, 5. 가설 검증, 6. 결론 7. 결과 보고)을 따라야 하는지 아니면 서로 다른 방법으로 조사 연구를 하여야 하는지, 여러분의 생각은 어떤지, 예와 함께 자신의 생각을 응호할 수 있는 설명을 적으시오 |
Appendix B: Interview Protocol for Students

(Potential Questions for Semi-structured Interview)

Name: _______________________________
Date: _______________________________

Interviewer:

There are no “right” or “wrong” answers to the following questions. I am only interested in your opinions and the reasons for your opinions on the Nature of Science, on your learning, and on your achievement assessments in science class. Most of the interview questions explore your responses to the survey in greater detail; a few deal with more general ideas.

[Please note: The particular questions used in an interview will depend on the type of survey questions, which the student answered. If the student completed the Type A Survey, the interviewer’s probing questions will focus on the Type B open-ended questions. Also, the individual student’s answers will influence what kinds of questions are asked. The following outlines the main contents of the interview.]

Science Classes:

1. What kinds of learning activities are most frequent?
2. Among your science classes, how many lab activities do you have?
3. What is the most effective way to learn science?
4. Which do you think is more important learning of scientific theories/laws/factual knowledge or learning about how science has been developed, what the effects of science on our society or how society or culture has affected science?

Achievement test

1. What types of questions in your science test? (True/False, matching, multiple choices or short answering questions, essay types or science project)
2. Which kinds of tests, do you think, measure most fairly and correctly your knowledge in science? (Writing an essay / Multiple-choice/ oral presentation/Group or individual project)
3. What are your strategies to get high scores in your science tests? Do you change your studying habits depending on your test methods?
4. Do you think how you learned and your tests of science affect your ideas (opinions) on science?

**Nature of science**

Unlike the questions on learning activities and tasks and the assessment, the interview questions on NOS were based on interviewees’ responses to the survey questions.

1. The probing questions for the interviewees from School 1 were on the comparisons between their religious views and science learning.

2. A Korean interviewee had relativistic views on NOS; thus, the question was the standards to accept a result or assertions of scientific research as valid ones.
Appendix C: A Letter to Principals

Date: dd/mm/yyyy

Dear Sir/Madam:

I am writing to introduce myself to you, inform you of a study I am seeking to conduct in science education, and invite the participation of your school students in it.

My name is HyeRan Park, a doctoral student at the Ontario Institute for Studies in Education at the University of Toronto. My study is exploring key factors influencing students’ understanding of the Nature of Science. I wish to ask your permission to invite Grade 8 students to participate voluntarily in this study.

The following is a short synopsis of the research.

**Title of Study:** Understanding of the Nature of Science: A Comparative Study of Canadian and Korean Students

**Purpose:** The research seeks to understand students ‘concepts of the nature of science, and how their perceptions on science achievement test and the learning science experiences relate to the concepts of nature of science.

**Procedures:** Participants will be asked to complete a survey questionnaire about students’ opinions on the Nature of Science, their perceptions of science achievement and learning science. To complete the questionnaire students will need about 40 minutes. The survey will be taken when students’ teacher thinks it proper to administer.

**Potential risks:** This study involves no risk, no detriment to classroom work, and no adverse effect on achievement scores. Students will not be identifiable, and confidentiality will be maintained.

**Potential benefits:** The content of nature of science is closely related to the concepts of scientific literacy: “The knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity (The Ontario Science and Technology Curriculum, 2007, p.163). Participants in this study should benefit from reflecting on their learning activities during science classes. By answering the questions, they will be prompted to think overtly about the characteristics of science and scientific knowledge and their strategies of studying science. This may motivate them to take science more seriously as a possible life pursuit. On request, they may learn the results of the survey.

**Payment for participation:** A specially prepared “Certificate” will be given to each participating students as ‘junior researchers in science’ signed by Dr. Woodruff and HyeRan
Park. Recognition of the teacher and school’s participation will receive a gift: for the teacher, a personal gift and for the school, a student microscope.

**Confidentiality:** All data generated during this study will remain confidential. Neither the names of school, teachers, nor students will be used in the doctoral thesis. Only Professor Earl Woodruff and I will have access to the primary data. All data will be digitally encrypted, and will be destroyed after the study is concluded.

**Participation and withdrawal:** Participating in this research is entirely voluntary. Participants can withdraw at any time without consequences of any kind.

Included is a copy of my approval letter (approval number 25723) from the University of Toronto, and the Toronto District School Board. If you have any questions, please feel free to contact me by phone (647-835-8519) or by email (hyeran.park@utoronto.ca). If you have any concerns regarding the study and questions about your rights as a research participant, please contact the Office of the Research Ethics from the University of Toronto, at 12 Queen’s Park Cres. West, McMurrich Building, 2nd Floor, Toronto, ON M5S 1S8, phone 416-946-3273, fax 416-946-5763 or email: ethicsreview@utoronto.ca

I do look forward to hearing from you soon. Thank you for your kind consideration of this letter.

Sincerely

HyeRan Park

---

**Contact Information**

<table>
<thead>
<tr>
<th>Principal Investigator:</th>
<th>Supervisor:</th>
</tr>
</thead>
<tbody>
<tr>
<td>HyeRan Park</td>
<td>Dr. Earl Woodruff, Associate Professor</td>
</tr>
</tbody>
</table>

Department of Curriculum Teaching and Learning Ontario Institute for Studies in Education/UT,

252 Bloor Str. W., Toronto, ON M5S 1V6

Room 9-176

Tel: 647-835-8519

E-mail: hyeran.park@utoronto.ca

Department of Human Development & Applied Psychology Ontario Institute for Studies in Education/UT,

252 Bloor Str. W., Toronto, ON M5S 1V6

Room 9-156

Tel: 416-926-1068

E-mail: earl.woodruff@utoronto.ca
Consent Form

I understand the procedures and conditions to participate in this research and agree to my school students’ participation.

________________________  ____________________________  ________________
Name of Principal          Signature of Principal           Date
Appendix D: A Letter to Science Teachers

Dear Sir/Madam,

I am contacting you asking your voluntary participation in a study in science education. Please let me introduce myself, and my research.

My name is HyeRan Park and I am a PhD student in the Curriculum Teaching and Learning department at the Ontario Institute of Studies in Education, University of Toronto, working under the supervision of Professor Earl Woodruff. I come to this study after 17 years of hands-on science teaching in middle schools connected to the Busan District School Board in South Korea. The study forms the basis of my doctoral dissertation on “Understanding of the Nature of Science: A Comparative Study of Canadian and Korean Students”. I invite your students to take part in this study. I believe the experience will enrich their understanding of how science works, and help improve their scientific literacy.

From a reading of the Ontario Secondary School curriculum, I have noticed how the curricula encourage enhanced understanding of the concepts of the Nature of Science (NOS) in elementary and secondary school science classes. I am interested in factors which can affect students’ understanding of NOS such as their perceptions of the learning science and of achievement assessment. Comprehensive studies on NOS have been conducted in such areas as the effects of teachers, science textbooks, and instructional approaches. Not many have made students’ perceptions of achievement assessment, learning experiences of science and the understanding of NOS. Therefore, to gain insights into these areas of inquiry; the study will collect data from students using a survey questionnaire established in each domain.

For this reason, I am inviting your students to participate in the study. The questionnaire should take approximately 40 minutes. The questionnaire consists of Likert scales and three open-ended questions about the concepts of NOS, their perceptions of achievement test. The survey has no “correct or wrong answers”. In order to reduce students’ workload, open-ended questions are divided into two different sets (the multiple choice questions are the same). Each student completes either Type A or Type B.

All data generated during this study will remain confidential. Neither the names of schools, teachers nor students will be used in the doctoral thesis. Only Professor Earl Woodruff and I will have access to the primary data. All data will be digitally encrypted, and will be destroyed five years after the study is concluded. Your students are free to refuse to answer or skip at any question, if it seems uncomfortable for them. Also, students are free to withdraw at any time if they want. Please be assured that your students are under no obligation to participate in this study.

The findings of this study will not only benefit you directly, but by participating in this study, you will contribute to the growth of new knowledge factors affecting student understanding of the Nature of Science. Understanding the Nature of Science is a field of growing importance to advance public scientific literacy not only in Ontario but also worldwide. At your request, you may receive a copy of the summary of findings from the study and a digital version of the entire thesis once complete.
Payment for participation: A specially prepared “Certificate” will be given to each participating student as ‘junior researchers in science’ signed by Dr. Woodruff and HyeRan Park. Recognition of the teacher efforts will receive a gift, and for the school, a student microscope will be given.

Please feel free to contact me by phone at 647-835-8519 or by email (hyeran.park@utoronto.ca) with any questions you may have about this study. If you have any concerns regarding the study and questions about your rights as a research participant, please contact the Office of the Research Ethics from the University of Toronto, at 12 Queen’s Park Cres. West, McMurrich Building, 3rd Floor, Toronto, ON M5S 1S8, phone 416-946-3273, fax 416-946-5763 or email: ethicsreview@utoronto.ca

Informed consent is required for your participation. In the spaces provided below, please indicate your willingness to participate by placing your signature.

Thank you for your kind consideration. I look forward to working with you.

Sincerely,

Hyeran Park

CONTACT INFORMATION

Principal Investigator
HyeRan Park (PhD student)
Department of Curriculum Teaching and Learning
Ontario Institute for Studies in Education
University of Toronto.
252 Bloor Str. W. Toronto, ON M5S 1V6
Room 9-176 (Tel: 647-835-8519)
Email: hyeran.park@utoronto.ca

Supervisor
Dr. Earl Woodruff (Associate Professor)
Department of Human Development & Applied Psychology
Ontario Institute for Studies in Education
University of Toronto.
252 Bloor Str. W. Toronto, ON M5S 1V6
Room 9-156 (Tel: 416-926-1068)
Email: earl.woodruff@utoronto.ca
Consent Form

I understand the procedures and conditions of my student participations described above and agree to participate in this study

_________________  ______________________  ____________
Name of Participant  Signature of Participant  Date
Appendix E: A Letter to Parents and Students

Dear Sir/ Madam:

As science and technology advance and influence all aspects of our society, it is vital that our children not only gain current scientific knowledge, but also come to understand the nature of science as a form of human learning, with all its human history and human limitations. I am asking your permission to invite your child to participate in a study in science education that should enhance his/her understanding of these issues, and improve classroom practice in this field.

To introduce myself: My name is HyeRan Park, a doctoral student at the Ontario Institute for Studies in Education at the University of Toronto. My concern for this topic comes from 17 years of hands-on middle school science teaching in South Korea. My dissertation study is exploring key factors influencing students’ understanding of the Nature of Science. With the permission of your child’s school principal and science teacher, I ask your permission for your child to share in this study along with 200 Grade 8 Science students from the Toronto District School Board.

The brief summary that follows outlines what we plan on doing, and how:

**Title of Study:** Understanding of the Nature of Science: A Comparative Study of Canadian and Korean Students.

**Purpose:** The research seeks to understand students’ concepts of the Nature of science, and how their perceptions of achievement assessment and learning science experiences relate to the concepts.

**Procedures:** Participants will be asked to complete a survey questionnaire about students’ opinions on the Nature of Science, and the perceptions of science achievement tests and learning science. To complete the questionnaire students will need about 40 minutes. The survey will be taken when their teacher thinks it proper to administer.

**Potential risks:** This study involves no risk, no detriment to classroom work, and no effect on achievement scores. Students will not be identifiable, and confidentiality will be maintained.

**Potential benefits:** Participants in this study should benefit from reflecting on their learning activities during science classes. By answering the questions, they will be prompted to think overtly about the characteristics of science and scientific knowledge. This may motivate them to take science more seriously as a possible life pursuit. On request, they may learn the results of the survey.
Confidentiality: All information obtained from this study will be secured and entirely confidential. Participants taking the survey are not to place their names on the questionnaire.

Participation and withdrawal: Participating in this research is entirely voluntary. Participants can withdraw at any time without any consequence.

If you would like to receive a copy of the completed study, or if you have any questions, please contact HyeRan Park at 647-835-8519 or by email at hyeran.park@utoronto.ca. If you have any concerns regarding the study and questions about your rights as a research participant, please contact the Office of the Research Ethics from the University of Toronto, at 12 Queen’s Park Cres. West, McMurrich Building, 3nd Floor, Toronto, ON M5S 1S8, phone 416-946-3273, fax 416-946-5763 or email: ethicsreview@utoronto.ca

Thank you for your support in the conduct of this study.

Sincerely,

Hyeran Park
Parent Consent Form

Your son or daughter’s participation in this study is entirely voluntary. You may refuse to have your child participate at all. You may also withdraw your child from the study at any time without any consequence.

Your signature below indicates that you have received a copy of this consent form for your own records, and that you agree to allow your child to participate.

I consent/do not consent (circle one) to my child’s participation,

(name:_________________), in this study.

__________________________  _____________  __________________________
Signature                      date                          Please print your name here

Student Assent

Your participation in this study is entirely voluntary and you may refuse to participate, or withdraw from the study at any time without any consequence to your class standing.

Your signature below indicates that you have received a copy of this assent form for your own records, and that you are willing to participate in this part of the study.

__________________________  _____________  __________________________
Signature                      date                          Please print your name here
Appendix F: Lab Report

여러 가지 금속의 비열 측정하기

<table>
<thead>
<tr>
<th>실험목표</th>
<th>금속의 비열을 측정하는 방법을 이해하고 여러 가지 금속의 비열을 측정할 수 있다.</th>
</tr>
</thead>
<tbody>
<tr>
<td>실험내용</td>
<td>• 관련단원: 8학년 열 에너지&lt;br&gt;• 관련 기본 개념: 비열, 열용량, 열량&lt;br&gt;• 용구요소: 실험&lt;br&gt;• 소요시간: 45분</td>
</tr>
<tr>
<td>저비물</td>
<td>알루미늄(또는 냉동이 있는 스티로폼 컵), 온도계 2개, 비커, 비열 측정용 금속(0.1 0.2kg 정도), 알코올램프, 성발이, 석면색그물, 메스실린더, 온도시계, 유리막대</td>
</tr>
</tbody>
</table>

활동과정

1. 금속 시료의 질량을 온도시계들로 측정한다.
2. 열량계에 놓은 물의 질량을 저울 또는 메스실린더로 측정한다.
3. 물을 열량계에 부은 후 온도를 측정한다. 이 때 물의 양은 금속 시료가 충분히 장갑 정도의 양이어야 한다.
4. 비커에 물을 2/3 정도 놓고 실로 묶은 금속 시료를 유리막대에 매달아 비커 속에 넣는다.

5. 비커의 물이 끓을 때까지 가열한 후 비커의 물의 온도를 측정한다.
6. 금속 시료를 비커에서 깨내어 즉시 열량계에 넣고 온도계를 보면서 채는 막대로 물을 천천히 툂저어 물이 최고 온도에 도달하여 열연령 상태가 되었을 때의 물의 온도를 읽고 기록한다.

7. 열량 보존의 법칙을 이용하여 금속의 열용량을 계산한다. 

\[ \Delta t = \frac{Q}{c \cdot m} \]

\[ \Delta t = \text{관물의 열량} \times \text{온도 변화} \]

\[ \text{관물의 열량} = \text{금속의 열용량} \times \text{온도 변화} \]

8. 금속의 열용량은 금속의 질량으로 나누어 금속의 비열을 구한다.

\[ \text{비열} = \frac{\text{열량}}{\text{질량} \times \text{온도 변화}} \]

9. 실험결과를 다음과의 표에 정리한다.

<table>
<thead>
<tr>
<th>시료</th>
<th>열량계 속의 물(또는액체)의 질량</th>
<th>시료의 질량</th>
<th>열량계 속의 물(액체)의 처음온도</th>
<th>물의 물속의 시료의 온도</th>
<th>열량계의 상태의 온도</th>
<th>열량계</th>
<th>비열</th>
</tr>
</thead>
<tbody>
<tr>
<td>금속1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>금속2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>금속3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

유의사항

- 본 실험에서는 정확한 온도의 측정이 매우 중요하다. 온도계는 올바르게 보정된 것을 사용한다.
- 열량계 내의 온도계가 금속 시료에 닿지 않도록 주의한다.
- 물이 끓을 때의 온도를 100도로 하지 말고 온도계로 측정한 값을 사용한다.
- 금속 시료를 열량계로 이동시킬 때는 신속하게 하여 외부로 유출되는 열을 최소화한다.
- 알코올램프에 물을 적재로 이동하기 꼭, 화살에 조심하며, 물을 끓 때에는 알코올램프의 두경을 잘 덮는다.

결과 및 토의

- 열량계의 물의 온도와 질량을 측정하는 이유는?
- 비커와 열량계에 물을 끓는 이유는?
- 열량계의 온도가 미리 이상 변하지 않는 이유는?
- 금속 시료의 열용량과 비열은 얼마인가?
- 실험으로 구한 값과 실제 비열의 값을 비교할 때 차이가 있었다면, 그 차이가 생긴 이유는 무엇일까?