UNIVERSITY OF GOTHENBURG

FACULTY OF EDUCATION

DEPARTMENT OF EDUCATION AND SPECIAL EDUCATION

A CROSS-COUNTRY STUDY ON THE EFFECTIVENESS OF INQUIRY-BASED AND TRADITIONAL DIDACTIC APPROACHES

Kadir Isik

Thesis:	30 credits
Program/course:	L2EUR (IMER) PDA184
Level:	Second cycle
Term/Year:	Spring 202O
Supervisor:	Stefan Johansson
Examiner:	
Report nr:	

Abstract

Master Thesis:	30 Credits
Program:	L2EUR (IMER) PDA 184
Level:	Second cycle
Term/Year:	Spring 202O
Supervisor:	Stefan Johansson
Examiner:	
Report nr:	
Keywords:	Inquiry-based approach, traditional didactic approach, science instruction

Aim : In past decades, teacher practices in science teaching have changed from perceived traditional ways of teaching to more inquiry-based approaches. The driving force behind this change is the assumption of inquiry-based approach being more effective in terms of student science achievement than the traditional didactic approach. This study aims to examine the extent of these two approaches in a cross-country perspective. Moreover, it investigates the effectiveness of these two instructional approaches on student science achievement.

Theory : Cognitive Load Theory (CLT) proposed by Sweller suggests that learning happens best under conditions that are aligned with human cognitive architecture. According to CLT, instructional design principles must be based on our knowledge of the brain and memory. CLT was used to ground the assumption that is investigated in this thesis.

Method : Single level Structural Equation Modelling (SEM) modelling is used to identify the relationship between two latent constructs of instructional approaches and student science achievement while Socio- economic status (SES) and student confidence (CON) are used as statistical controls. This study used 8th grade dataset in Trends in International Mathematics and Science Study (TIMSS) 2015 by performing statistical analyses in Mplus version 8.2 software together with IBM SPSS Statistics 25.

Results : Findings across 12 countries indicate no clear evidence in favour of neither both instructional approaches, with the exception of the results from Italy in which the traditional didactic approach is found to be negatively influencing student science achievement, explaining 21% of the variance in achievement.

Acknowledgements

This thesis becomes a reality with the kind support of many individuals. I would like to extend my sincere thanks to all of them.

Foremost, I would like to express my very great appreciation to my supervisor Stefan Johansson for his valuable and constructive suggestions during the development of this thesis. His willingness to give his time so generously has been very much appreciated.

I also would like to thank all the lecturers who taught us in the IMER programme. My special thanks are extended to our coordinator Dr. Ernst Thoutenhoofd who has been very helpful, supportive, and friendly all along with the master programme.

As this thesis is the product of two years of studying, I feel I should mention my gratitude for my fellow classmates and friends. I believe they also contributed my learning journey through the discussions we carried out in the lectures, and their support outside of the classroom.

At last but not least, I wish to thank my parent and siblings for supporting me throughout my life and specifically during my studies.

Kadir ISIK

Amsterdam, Netherlands

May 2020

Table of Contents

LIST OF ABBREVIATIONS

EU	The European Union
NRC	National Research Council
TIMSS	Trends in International Mathematics and Science Study
IEA	International Association for the Evaluation of Educational Achievement
CLT	Cognitive Load Theory
INQ	Inquiry-based Approach
TRA	Traditional Didactic Approach
DI	Direct Instruction
SEM	Structural Equation Modeling
SES	Socio-economic Status
CON	Student Confidence on Science

1. Introduction

Today, the importance of science instruction is well established since it's vital role in terms of fulfilling the skill gaps in science, technology, engineering which are the dynamics of a growing economy. (Condon & Wichowsky, 2018) Therefore, good science instruction which helps students learning science adequately becomes an important matter for building tomorrow's competent workforce. By mid-20th century, good science instruction was associated with the term inquiry. (Anderson, 2002) So that, inquiry-based science instruction has been promoted across the world. For instance, Europe Union was funding several EU-projects focusing on inquiry-based instruction. (Rundgren, 2018) Moreover, American Association for the Advancement of Science (AAAS) and National Research Council (NRC) have developed guidelines for inquiry-based instruction that is ,as they highlighted, reflecting current scholarship on nature of science. (Abd-El-Khalick et al., 2004; Blanchard et al., 2010; Furtak, Seidel, Iverson, & Briggs, 2012) Furthermore, they also call attention to inquiry in science education and suggest that it supports students to acquire critical thinking. This call has led to, especially in many western countries, inquiry-based approaches to be more dominant throughout school systems and defining the curriculum standards.(Rowe, 2006) Hereby, teacher practices from perceived traditional ways of teaching give place to more inquiry-based approach. (Akkus, Gunel, & Hand, 2007)

A key reason for this shift away from traditional teaching practices to inquiry-based approach was the increase in the number of educational research that is critical towards traditional ways of teaching.(Heaysman & Tubin, 2019) These critics shaped around that they "are very formal focused on the memorizing of the facts without any deeper understanding of the processes in the nature."(Kubiatko, 2016, p. 4) and "possess endemically low levels of student engagement." (Scott, Smith, Chu, & Friesen, 2018, p. 37) In contrast, inquiry-based instruction is described as engaging students in the thinking process and scientific activities. Thus, it "includes students drawing upon their scientific knowledge to ask scientifically oriented questions, collect and analyze evidence from scientific investigations, develop explanations of scientific phenomena, and communicate those explanations with their teacher and peers." (NRC, 1996 as cited in Furtak et al., 2012, p. 301) Such characteristics of the inquiry-based approach are argued to be better aligned with how people learn. (Blanchard et al., 2010) Therefore, it is expected to help students to reach desired learning outcomes.

Even though the countries relied on the assumption of the inquiry-based approach being more effective in influencing student science achievement than traditional didactic approach(Gao, 2014), this assumption still needs persuasive confirmation. Because, the empirical support for this claim is weak. (Blanchard et al., 2010) There is a remarkable number of empirical and theoretical studies that stress the effectiveness of inquiry-based approaches in science teaching, and rather argues the efficiency of the traditional didactic approach.(Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004; Stockard, Wood, Coughlin, & Rasplica Khoury, 2018) The argument is mainly generated around the inefficiency of such minimal guided instructions and the necessity of guidance supporting the cognitive processing.

The inquiry-based approach, however, is the trend in educational circles, there is a need for more studies that examine the effectiveness of such instructional approaches in order to yield proper directions to science teaching. As well as, the number of studies that investigate this phenomenon in a cross-country perspective is limited, especially using the large-scale dataset assessments. Before drawing on a certain conclusion, the reliability of these approaches must be argued. In this context, depreciation of other teaching methods, such as the traditional didactic approach, should be avoided. This study aims to examine the extent of these two approaches in a cross-country perspective. Moreover, I will explore how these two instructional approaches relate to student science achievement.

2. Background

2.1. Traditional-Didactic Approach

By the 20th century, when education started to become a model, which is similar to today's conventional schooling, initially, the behaviourist approach was dominant in terms of teaching practices and classroom set-up. (Ertmer & Newby, 2013) Behaviourism arose in 1913 when John Watson wrote an article entitled 'Psychology as the behaviorist views it'. John Watson set out a number of underlying assumptions regarding methodology and behavioural analysis. B.F Skinner, one of the most outspoken behaviourism psychologists, adopted a learning model in which teachers are seen as the source of the knowledge in the class; and students act as the receivers of the knowledge that send by teachers. That traditional didactic approach was prominent model of learning for centuries. (B. Khalaf, 2018) The most distinctive criterion of traditional learning is that teachers talk more than students and the learning process is based on a whole class participation where no individual or group activities enforced. (Rashty 1999, as cited in B. Khalaf, 2018)

Moreover, it is rooted the direct instruction (DI) of Siegfried Engelmann (Bereiter & Engelmann as cited in Magliaro, Lockee, & Burton, 2005) Direct Instruction (DI) can be considered as advanced and revised instructional model of the behavioural theory. Even though it suggests specific guidelines that go beyond the behavioural theory such as aiming to get participation active by all students, in essence, it is a highly organized, teacher-directed approach in which skills are divided into small units, ordered sequentially, and taught explicitly. (M. Cohen, 2008)Herewith, it aims at avoiding the misconceptions that may occur during the learning process, and eventually allow for accelerated and more efficient learning. In the present study DI is also underpinned Traditional-didactic approach in which a body of knowledge transmitted from teachers to learners that are considered passive recipients of the knowledge and that leads to a teacher-centred classroom.(Kaymakamoglu, 2018) These instructional model also referred as transmissionist model, teacher-led learning, or direct instruction by researchers as stated in Klahr and Nigam (2004).

The traditional didactic approach in science education was criticised by not helping students to achieve a deeper understanding of knowledge, i.e. student' memorized knowledge rather than understand it. (Biggs, 1996) This has thought to be causing challenges and drawbacks in practical science education. Hereby, the instructional approaches, models which favours the student engagement gained popularity. Especially, since 70s there have been calls for reform of the old traditional methods of teacher-centred learning into practical methods that are more focussed on learners.

2.2. Inquiry-Based Approach

Inquiry based approach in science education can be tracked to 1950s, when Jean Piaget investigated the different ways in which children thought and processed information.(Kubiatko, 2016) Especially since the 1960s, the inquiry-based approach has become a popular subject through the emphasis of researchers e.g. Schwab (1962) in terms of the effectiveness of teaching approaches. Suchman (1966) describes inquiry as "a form of human behaviour in which person acts to increase the meaningfulness of his knowledge and experience." (p.178) Therefore, according to Suchman, learners' meaningful encounters with a concept or knowledge are more valuable than the teacher attempt to feed meanings to the children directly through verbal and other symbolic means. Since children are natural inquirers who have many questions and they seek to find explanations for these questions by interacting with their environments and others as well as using their prior knowledge actively, instead of providing ready-made answers, teachers should encourage students to seek answer themselves.

However, 'inquiry' was not a new conceptualization of the learning. Its roots go back to the famous works of Jean Jacques Rosseau, Emile. It also can be found in the influential writings of John Dewey (1910).(Krahenbuhl, 2016) Dewey (1938) has emphasized the importance of experience in learning. He has been critical to "static" teaching methods. Later, Papert's report (as cited in Heaysman and Tubin, 2019) shows Dewey's view towards the traditional didactic approach as it does not value interaction and discourses. Piaget (1973) and Vygotsky (1976) were also critical towards the traditional approach as being static in which students do not take an active role unless their teacher asks to do so. They suggest that learners' involvement in learning process is more meaningful in developing learner' skills, experience and knowledge.

In Dewey's proposed model student takes an active role while the teacher operates as a facilitator or a guide. In this model, students are encouraged to "..address the problems they want to know and apply it to the observable phenomena."(Barrow, 2006, p. 266) Constructivism that gained its popularity by 1970s and 1980s can be attributed to Dewey's model. According to constructivist theory "learning occurs best when it is self-constructed, initiated by students themselves in response to their interests with the teacher acting as a facilitator or guide." (Heal, Hanley, &Layer as cited in McMullen & Madelaine, 2014, p. 147) As noted, it can be seen that constructivist theory set out the roles of student and teachers similar to Dewey's model.

The theoretical foundations of inquiry-based approach are based on the constructivist learning theory. (B. Khalaf, 2018) This can be seen when looking at the characteristics for the central characteristics of constructivist learning explained by Brunning, Schraw, and Ronning (as cited in Krahenbuhl ,2016): (1)Learner constructs their own learning, (2)Social interaction plays a key role (3) Authentic learning tasks are crucial for meaningful learning, (4) Learning dependent on existing understanding.(p.98) NRC (2000) describes these core components of inquiry-based approach which is very similar to those characteristics of constructivism, as following (as cited in Bevins & Price, 2016, p. 18):

(1) Learners are engaged by scientifically oriented questions.
(2) Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions.
(3) Learners formulate explanations from evidence to address scientifically oriented questions.
(4) Learners evaluate their explanations in light of alternative explanations, particularly those reflecting scientific understanding.
(5) Learners communicate and justify their proposed explanations.

The inquiry-based and constructivist approaches seem to share many educational objectives. In this regard, the discussions of inquiry cannot be separated from the discussions of constructivist approach. (Abd-El-Khalick et al., 2004) Consequently, this study includes the literature that refers to the constructivist theory and underpins it as an inquiry-based approach since they are profoundly similar.

3. Theoretical Framework

3.1. Cognitive Load Theory

Learning theories are essential for effective teaching as they shed light on different aspects of the learning process. (Yilmaz, 2011) Nevertheless, according to Cognitive Load Theory, to what extent these theories can be effective depends on whether they attach importance to the characteristics of human cognition. Therefore, it is important to determine the conditions in which learning is maximized and effective. (Sweller, van Merrienboer, & Paas, 1998) "Cognitive load theory integrates the origins of human cognition in evolutionary theory with the structures and functions of human cognitive architecture to provide effective instructional design principles."(Sweller, 2008, p. 370) Moreover, Sweller (2008) claims that an efficient instruction must rely on the characteristics of human cognitive architecture, and he emphasizes the need to apply instructional design principles based on our knowledge of the brain and memory. Well-known cognitive structures such as working memory and long-term memory are interrelated because schemas held in long-term memory, acting as a "central executive", directly affect the manner in which information is synthesized in working memory. (Sweller et al., 1998) In the absence of schemas, instructional guidance must provide a substitute for learners to develop either their own schemas. In this sense, it challenges the constructivist perspective in which learners are supposed to discover or construct essential information for themselves. Furthermore, it supports the idea that Direct Instruction which explains the concepts and procedures that learner is required to learn should be provided and the learner should not be left to discover those procedures by themselves. (Kirschner et al., 2006) This perspective challenges Vygotsky's (1978) argument that the children learn at their ZPD (zone of proximal development) which is the distance between what learners already know and can do independently and what they can do with the help of a teacher or a peer.(Shabani, Khatib, & Ebadi, 2010) Contrarily, according to Sweller's theory leaners should be provided guidance in order to acquire knowledge and construct meaning.

Furthermore, Condon and Wichowsky (2018) noted that there is a broad consensus on inquiry-based science teaching that it provides a structure in which students guide themselves. In this context, teachers' role is to facilitate such construction of knowledge.(Rowe, 2006) However, the teacher's and student's roles in inquiry-science teaching are defined differently by researchers. Especially, the arguments differ around the level of guidance that will be given by the teacher and the level of student autonomy. Abrams, Southerland, and Evans (2007) defines the levels of inquiry-based instruction on the '*guidance given by teacher*' and '*open to student*' and introduces 4 different levels. Then, they discuss the appropriate amount of guidance in terms of the most efficient learning. These different interpretations cause some arguments on what inquiry means and constitutes. Accordingly, the implementation of inquiry-based teaching shows variation. Thus, researchers ambition to define inquiry science teaching has led to an extensive literature. (Anderson, 2002) This situation, as noted earlier, causes a lack of shared terminology and precise definitions of an inquiry-based approach. (Anderson, 2002; Blanchard et al., 2010)

Besides the level of guidance that should be given to learners is unclear, the inquiry-based approach is also struggling to answer what will learners do when they deal with a novel of information. In this scenario, unlike Direct Instruction, the inquiry-based approach is inefficient to provide information to learners, and thus help them to develop a conceptual schema to integrate the new information with their prior knowledge. (Darling-Hammond, Flook, Cook-Harvey, Barron, & Osher, 2020) According to CLT, when learners are left to explore a highly complex environment, they will end up with heavy working memory which is detrimental. (Kirschner et al., 2006) The *Worked Example* effect, present solutions to this heavy working memory problem and provides strong evidence for the superiority of directly guided instruction over minimal guidance. Using *worked examples*, which learners are shown step by step solutions, reduces the cognitive load through promoting sharing representations. (Valcke, 2002) It is also found to more efficient in terms of retaining integrated knowledge than constructivist approaches.(Vogel-Walcutt, Gebrim, Bowers, Carper, & Nicholson, 2011) These arguments founded around CLT generate a strong foundation against the efficiency of inquiry-based approach. Moreover, it puts forward the necessity of Direct Instruction, especially for novice learners.

4. Literature Review

4.1. Inquiry-Based Approach Found to Be More Effective

In his studies, (Colburn, 2000b) found many pieces of evidences support that inquiry-based instruction is superior to other instructional modes for student's achievement. Then he questioned if the inquiry is effective why its implications into practice cannot be seen. (Colburn, 2002a). Anderson (2002) claims that research about the effectiveness of the inquiry-based approach has matured. The focus of the research has changed from the effectiveness of the approach to the dynamics of such instruction and its implications. He argued whether the inquiry-based approach can be placed to teaching practice in schools on a widespread basis, besides the effectiveness of the approach. Then he further made suggestions regarding teacher's and student's roles for the inquiry-based approach.

On the other hand, the argument on the effectiveness of inquiry-based approaches still was ongoing due to such studies Klahr and Nigam (2004) that remarked the superiority of traditional direct instruction over discovery learning. As a follow-up study to Klahr and Nigam (2004), Dean Jr and Kuhn (2007) conducted a research on the same age group as Klahr and Nigam but for an extended time period. They compared three groups of 15 fourth grade students of diverse socioeconomic backgrounds on problems that required them to control variables to reach an effective solution related to forecasting an earthquake. One group engaged in only discovery learning. Another group received direct instructions on a concept before engaging in the same activity. A third group received only the direct instruction without any engagement or practice. Dean and Kuhn concluded that, in this longer-term framework, direct instruction is neither a necessary nor sufficient condition to acquire or to maintain the knowledge over time. This study unintentionally points to another argument concerning the efficiency of these instructional practices since the student on discovery learning group spent more time on tasks than the ones in the direct instruction only group.

The results from the study of Akkus et al. (2007) indicated another aspect of the argument on the effectiveness of inquiry-based approaches. In this study, they compared the effectiveness of the inquiry-based approach known as the Science Writing Heuristic approach as a treatment to traditional teaching practices on students' post-test scores in relation to students' achievement level and teacher's implementation of the approach. The considerable finding of the study was the quality of the implementation does have an impact on student performance. However, a more remarkable finding of the study was that low-achieving science students benefit most from the implementation of the SWH approach. The effect size difference between high achieving and low-achieving students in high traditional teaching was 1.23 standard deviation units, while for high SWH teaching the effect size difference was 0.13 standard deviation units. the mean score for the high-achieving students in either treatment condition did not vary—thus, either approach was equally valuable for high-achieving students because they were able to adapt. These results contradict a more recent study from Blanchard et al. (2010) with a sample of 1700 students of 12 middle school and 12 high school science teachers. Blanchard et al. (2010) compare the efficacy of Level 2, guided inquiry-based instruction to more traditional, verification laboratory instruction in supporting student performance on science learning. In their finding, they argued the quality of students' inquiry skills and their own prior knowledge are essential to conducting inquiry-based learning. The greater the skill level and the knowledge of students, the higher level of inquiry that can be reasonably employed. Additionally, they found evidence that lower socioeconomic status refers to lower achievement in both instructional methods. However, they claim that inquiry-based teaching methods are more effective on the achievement of students' in lower income schools over students from schools that more traditional instruction applied.

Furtak, et al. (2012) conducted a meta-analysis on studies published between 1996 and 2006, a decade during which inquiry was the prominent instructional approach in science education reform. Within their framework, 37 experimental and quasi experimental studies were coded. The findings of the study showed that inquiry-based teaching has an effect on student learning with the overall mean effect size .50. Besides, they found evidence that supports the superiority of teacher-led activities over student-led activities through a 0.4 higher effect size difference. These findings lead to further studies which question the impact of guidance in inquiry-based approaches e.g. Lazonder and Harmsen (2016). Lazonder and Harmsen conducted a meta-analysis where they synthesized 72 studies in order to compare the effectiveness of different levels of guidance for different age categories. The results showed that guidance has a significant positive effect on inquiry learning activities, performance success, and learning outcomes. These findings agree with plenty of studies that documented teacher-led conditions and guidance has a positive impact on learning outcomes despite they address the superiority of the different kind of instructions. (Alfieri et al., 2011; Hattie, 2009; Kirschner et al., 2006; Schroeder, Scott, Tolson, Huang, & Lee, 2007);

The increase in the number of studies that defend the positive effects of guidance on learning outcomes might cause the popularity of inquiry-based approaches to come to a standstill. However, in a recent study of Scott et al. (2018), such criticisms to inquiry-based approaches derived from that it is directed at discovery learning in theory and research in the field and the argument against curricular shifts towards inquiry reflect the limitations of discovery learning. Scott et al. (2018) draw attention to guided forms of inquiry, such as problem-based learning, and approaches to inquiry aligned with the authentic education movement. They noted that these are adopting approaches to inquiry that have demonstrated significant educational affordances as well as discovery learning. Furthermore, they claim that these frameworks do not oppose key elements of traditional forms of education, such as direct instruction. Finally, they emphasize the specific instructional supports needed for processes of inquiry to promote elements, such as critical thinking skills and flexible problem-solving abilities, necessary for success in a rapidly changing world.

Gao (2014) compared the effects of inquiry-based practices and traditional didactic practices on student achievement. 8th grade dataset from Singapore, Chinese Taipei, and the US from TIMSS (Trends in International Mathematics and Science Study) 2011 was selected in order to examine the research questions. This is one of the few studies drawing on the data from international assessment tests such as PISA, TIMSS on this field. He used a two-level hierarchical linear modeling (HLM) approach and simultaneous multiple regression and controlled Social Economic Status (SES), student self-confidence in learning science, and three affective teaching practices as these variables might have confounded the effects of teaching approaches on student science performance. The findings of the study revealed no robust association between teaching practices and student achievement. However, some negative observations for didactic practices were shown in different regions and in either low, medium, or high achieving students. On the other hand, none of these inquiry-based or traditional didactic science-teaching practices were found to be positive predictors of science performance in all three countries/regions except for the case of two inquiry-based teaching practice items that were positively related to Chinese Taipei students' achievements. In light of these findings, the positive effects of inquiry-based practices on student performance cannot be inferred.

4.2. Traditional Didactic Approach Found to Be More Effective

Klahr and Nigam (2004)compared the effectiveness of direct instruction and discovery learning with the sample consists of 112 third and fourth-grade students. This study is referred to as evidence for the superiority of traditional didactic approaches over inquiry-based approaches in terms of learning outcomes. They had two groups as a direct instruction group and discovery learning group. In the direct instruction group, students all phases of instruction controlled by the teacher, however, in the discovery learning group teacher's agency was absent. On the first day of study, students learned the control-of-variables strategy (CVS) which is a method for creating experiments in which a single contrast is made between experimental conditions. Then, one week later, they expected to assess posters through CVS. Klahr and Nigam found that number of students who mastered at CVS were higher for direct instruction then discovery learning, respectively 40(%77) and 12(%23). The same year, Mayer (2004) conducted a study to demonstrate sufficient evidence which will lead to questioning of discovery learning. He reviewed research on the discovery of problem-solving rules culminating in the 1960s, discovery of conservation strategies culminating in the 1970s, and the discovery of LOGO programming strategies culminating in the 1980s. He concluded that guided discovery was more effective than pure discovery for each case. He pointed out the importance of instructional guidance on learning as well as the need of including physiology to the argument for educational reform.

Similar to Mayer (2004), a more psychological study was conducted on the effectiveness of inquiry-based learning by Kirschner, Sweller, and Clark (2006). They challenged the inquiry-based instruction being less effective than guided instructional methods. They qualify inquiry, discovery learning, problem-based learning, and experiential learning instructions which have originated from the constructivist approach as minimally guided forms of instructions. They advocate that these approaches ignore the human cognitive architecture, and the evidence from empirical studies from the last decade which demonstrated the effectiveness of guidance in student learning. They highlight the importance of teacher guidance since there is a body of research supporting these approaches. More recent studies, even though some of them demonstrated that the inquiry-based approaches are more effective in terms of learning outcomes, found that teacher's guidance has a positive effect on student learning (Furtak et al., 2012; Lazonder & Harmsen, 2016).

A relatively recent study of Alfieri et al. (2011) established favourable results for direct instruction. They conducted 2 meta-analyses using a sample of 164 studies. In the first meta-analysis, they examined the effects of unguided discovery learning and explicit (direct) instruction. Within the second meta-analysis, they searched evidence for the effects of enhanced and/or guided discovery (M. Cohen, 2008) other types of instructions. 580 comparisons from the first meta-analysis demonstrated that explicit (direct) instruction has positive effects in terms of learning outcomes compared to unguided discovery learning. On the other hand, analyses of 360 comparisons from the second meta-analysis revealed that outcomes were favourable for enhanced discovery when compared with other forms of instruction. Alfieri et al. (2011) in their conclusion, propose a change in the focus of the argument from the limitations of discovery learning to the consequent empirical investigations which concern the implementation of what these studies suggest. Another but a quite recent meta-analysis with a larger sample from Stockard et al. (2018) presented results that support earlier reviews in the literature on the effectiveness of direct instruction. The results derived from 328 studies over a 50-year period and almost 4,000 calculated effects and involved a wide range of subjects, settings, comparison groups, and methodological approaches. As well as various academic achievement measures, the study ability measures; affective outcomes; teacher and parent views. And, all of the estimated effects were positive.

McMullen and Madelaine (2014) wrote a literature review where detailed the components of direct instruction, research to support it, and reported attitudes towards it. They, especially, advocated direct instruction against the criticisms it has been drawn to while there is a strong research base to support its effectiveness. They inferred that the criticisms and negative attitudes towards direct instruction likely caused by a mismatch of teaching philosophies and can be attributed to misinformation about the methodology. Moreover, they set out three main practices to improve the attitudes towards direct instruction. These can be summarized as: first, spreading the accurate information about direct instruction; second, active, ongoing support to learn the skills adapted to new methodology during its implementation by schools and staff; third, acquirements of teachers and schools in order to show the effects of their new implemented methodology.

4.3. Studies Suggest Mixed-Approach or Found Inconclusive Findings

Schroeder et al. (2007) conducted a meta-analysis that consisted of research published from 1980 to 2004 on the effect of specific science teaching strategies on student achievement. Studies they have synthesized were required to have been carried out in the United States and must have included effect size or the statistics necessary to calculate an effect size. In the end, sixty-one studies were eligible for the meta-analysis. Since they did not focus on particularly inquiry-based approaches nor traditional didactic approaches. However, inquiry-based strategies were categorized as one of the teaching strategies in the test, while direct instruction was excluded due to the lack of studies with science achievement outcomes. The ranking of teaching strategies can be seen in **Table 1** below.

Table 1 The ranking of teaching strategies, Schroeder et al. (2007)

Teaching Strategy	**Effect Size**
Enhanced Context Strategies	1.48
Collaborative Learning Strategies	0.96
Questioning Strategies	0.74

Inquiry Strategies	0.65
Manipulation Strategies	0.57
Assessment Strategies	0.51
Instructional Technology Strategies	0.48
Enhanced Material Strategies	0.29

Nevertheless, it is essential to note that the studies they analysed within ‘enhanced context strategies` are highly related to the direct instruction. Especially one of the studies where direct instruction and inquiry-based approaches compared, explicitly reported findings in favour of direct instruction. On the other hand, Schroeder et al. (2007) claim that teachers must have competence in order to purposefully employ those strategies to reach particular learning aims. A study from Minner, Levy, and Century (2010) analysed both numerical and text data from 138 studies mainly conducted in the United States (105, 76%) like Schroeder et al. (2007). They synthesize findings from research conducted between 1984 and 2002 to expose the impact of inquiry-based science instruction on student outcomes. Even though they claimed the findings demonstrated a positive trend favouring inquiry-based instructional practices, they noted that that overall higher levels of inquiry intensity do not lead to more positive learning outcomes for students.

One of the studies that actually reported inconclusive findings in terms of the relationship between instructional approaches and student achievement was from N. Lederman, Lederman, Wickman, and Lager-Nyqvist (N. Lederman, Lederman, J., Wickman, P. -O., & Lager-Nyqvist, L., 2007). Lederman et. al. conducted this research with the 8th grade teachers in Chicago and 6th and 7th grades teachers in Stockholm and with approximately 500 students in total. All teachers who participated in the study had 2 weeks of professional development. During these 2 weeks, the teachers taught science subjects using three types of instruction: inquiry-based instruction, direct instruction, and a hybrid method in between inquiry and direct instruction. Briefly, they reported no significant differences in student test scores that may be impacted by a type of instruction. Furthermore, they replicated this study with the same group of teachers. This study resulted in similar findings as well. The authors found no significant differences in post-tests based on the instructional approach. (J. S. Lederman, Lederman, N. G., & Wickman, P.-O. , 2008)

Goh, Kwek, Hogan, and Cheong (2014) presented a technique and applied to the teaching practices ‘data observed Grade 5 and Grade 9 Mathematics classes in Singapore. The findings of the study confirm the PISA 2012 findings on Singapore Mathematics performance and show that there is a strong relationship between the teaching of formal mathematics and student mathematical performance in the PISA tests. In this study, the teaching practices of both Grade 5 and Grade 9 Mathematics lessons were organized around *Knowledge as Truth* and *Instructional Activity (IA): Teacher-Dominated Talk* hubs which exemplify the transmissionist model of teaching. Besides that, the *Doing Mathematics Activity* hub was presenting in the Grade 5 transition network. Although, the findings revealed the effectiveness of direct instruction on statement performance in a high-stake test, (PISA 2011), Goh et al. (2014) attributed these findings to teachers’ and students’ aims to perform well in the high stake’s mathematics examination. They criticized this transmissionist model is insufficient as a teaching and learning model for mathematics and emphasized the necessity of engaging in authentic, content specific mathematical practices.

In a quite recent qualitative study, Heaysman and Tubin (2019), proposed to a mixed approach to teaching. They recommend that innovative teaching practices ought not to be taken as opposed to traditional teaching practices. They challenged the common dichotomy between traditional teaching regarded as limited, teacher centred and innovative teaching which is embraced as enhanced learning by being more engaging. They mentioned the issues of both approaches as well as their positive effects on learning. Eventually, they highlight that a well-regulated combination of traditional and innovative teaching practices may be more effective on student performance.

5. Research Questions and Hypotheses

The following study intends to query the assumption of the inquiry-based approach being more effective on student's science achievement than the traditional didactic approach. The other indicators that potentially have an impact on student science achievement are taken into account and controlled. The study begins with investigating all participated countries in TIMSS 2015, and the sample is drop down to 12 countries: *Chile, Egypt, England, Italy, Japan, Lithuania, New Zealand, Norway, Russia, Singapore, Slovenia, South Africa.* To be able to reach the central aim of the study, further specific research questions are asked:

- To what extent do teachers in the TIMSS 2015 countries use inquiry-based and traditional didactic approaches in their teaching?
- Do the inquiry-based approach and traditional didactic approach practices are significantly related to the science achievement for 8th grade students?
  - Which instructional approach is more effective, that is, related to student achievement?
  - Does the relationship between student achievement and instructional approaches differ among countries different from performance levels?

6. Methodology

This thesis employed Structural Equation Modelling (SEM) study in order to examine the effects of Inquiry-based and traditional didactic approaches on student achievement in 8th grade science students in Trends in International Mathematics and Science Study (TIMSS) 2015. During the preparation of the data and obtaining the descriptive statistics, 'IBM SPSS Statistics 25' software was used. SEM analysis was operated in Mplus version 8.2 (Muthén & Muthén, 2018) This section presents the methodological and statistical procedures carried out in this thesis. Population and sample, instrumentation, data collection, and analysis procedures will be examined.

6.1. Data and Sample

6.1.1. Data

The present study used data from the international TIMSS studies of 2015. TIMSS measures trends in mathematics and science achievement at the fourth and eighth grades in participating countries around the world, while also monitoring curricular implementation and identifying promising instructional practices. TIMSS has assessed mathematics and science since 1995 on a regular 4-year cycle. The main reason for selecting TIMSS 2015 is that it provides the measures of student achievement and teacher questionnaires within a large-scale database. In other words, student data from TIMSS can be aggregated on the teacher level and therefore the relationship between student achievement and teacher responses can lead to a better semblance of actual teaching in classrooms. TIMSS 2015, provides data from 4th and 8^{the} grade students and teachers albeit the countries differentiate for the grades. This study only focusses on the 8th grade students' data. The reason for that, the inquiry-based approach operates at its best in middle school, especially in grades 8–9. (Heaysman & Tubin, 2019)

6.1.2. Sample

In TIMSS 2015, the basic international sample design is a stratified two-stage cluster sample design. The first sampling stage consist of sampling schools from sampling frame which refers to all schools in the country that have students enrolled in the target grade. During the sampling process, a systematic random sampling approach for TIIMSS 2015 has been followed. In the second stage of sampling one or more intact class from the target grade of each participating school were selected (Chapter 3, Sample Design in TIMSS 2015) In the present study, all countries were included in early analysis. In total, 16,959 teachers and 282,204 students from 39 participating countries and regions and 7 benchmarking entities were included at the first stage of the sampling. The student-level sample size ranged from 3,759 to 18,012, lowest in Saudi Arabia, and highest in the United Arab Emirates.

The countries that will be included in the investigation were determined in two steps. The first, an index was created based on the usage of the two instructional approaches in each country. Based on this index six countries included in this thesis. The detailed information presented in the **Results** section. Secondly, in order to examine the potential differences in countries with performance levels, six countries from three different performance levels were chosen for further investigation of their teaching practices. Japan and Singapore were selected as high achieving countries, South Africa and Egypt as low achieving countries, Norway and Italy as medium achieving have been selected from the TIMSS 2015 8^{TH} grade Science Achievement scale. (See **Appendix 1** for science performance in TIMSS 2015) The number of participants (teachers and students) vary among the countries included in the present study. (See **Table 2**)

Table 2 Sample Size for Countries included in the study

Country	Number of Teachers	Number of students
Chile	*194*	4,849
Egypt	*213*	5,711
England	*777*	7,822
Italy	*228*	4,481
Japan	*171*	4,745
Lithuania	*910*	4,347
New Zealand	*211*	8,142
Norway	*333*	4,697
Russia	*761*	4,780
Singapore	*320*	6,116
Slovenia	*572*	4,257
South Africa	*319*	12,514

6.2. Reliability and Validity of TIMSS 2015

Reliability in quantitative research briefly can be referred to as the possibility of replication. That means a scale, or a test will give the same result when the same measurement repeated under constant conditions. (Moser and Kalton, as cited in Taherdoost, 2016) Although, reliability is not, yet, sufficient, it is a vital pre-requisite for validity. Reliability can be assessed in different ways: *internal consistency, inter-rater reliability, test-retest reliability, parallel forms reliability.* Since TIMSS is a 4-year cycle study, most of the measurement items are similar or the same as previous TIMSS studies. Also, considering the number of items used in order to measure the science domain, reliable measurement over time is ensured in TIMSS 2015.

L. Cohen, Manion, and Morrison (2018) describe validity as a demonstration of a particular instrument that measures what intends, purports, and claims to measure. In a quantitative study, validity is defined as the extent to which a concept is accurately measured.(L. Cohen, Manion, & Morrison, 2011) Even though, there are many types of validity, it can be examined around three main types: *content validity, construct validity and criterion validity*. Content validity is "the degree to which items in an instrument reflect the content universe to which the instrument will be generalized."(Straub, Boudreau, & Gefen, 2004, p. 424) During the item development process, TIMSS used a collaborative process by involving subject matter experts, country representatives. This process is run in accordance with the frameworks. They also work closely with the National Research Coordinator in each country and enforce to follow a set of standardized operations procedures. Construct validity is based on the relationship between the theoretical concept and tested measurement. A meaningful relationship between ensures the construct validity. Criterion Related Validity is the extent to which a measure is related to the result. This can be achieved by comparing a measure with another measure which has been proved to be valid. As it is addressed in TIMSS 2015, the test results can be validated by comparing them with student social-economic status which is supposedly related to the academic performance according to literature and test. The number of examples can be increased in TIMSS 2015. Thus, it would be fair to make an inference claiming the criterion-related validity is ensured by TIMSS. Further information regarding reliability and validity of TIMSS 2015 can be found in Mullis, Cotter, Fishbein, and Centurino (2016)

6.3. Variables

In this section, the details of the student achievement and selected variables, theoretically proven to potentially impact student achievement are demonstrated. These variables are the ones that are chosen to measure two latent constructs (TRA; INQ), control variables: socioeconomic status (SES), and student confidence (CON) and student achievement (SciAch). The variables that measure the two latent constructs and SES are independent variables, which is not influenced by any other factor. Student science achievement is, however, a dependent variable that potentially affected by other factors.

The descriptive statistics presented for a set of variables. They usually include mean and standard deviation figures. *Mean* refers to the average value of a group of numbers. *Standard deviation* provides insight into the variation of these groups of values. The mean score is derived by dividing the sum of a group of values by the number of values. The standard deviation (s or σ) is the positive square root of the variance. (Sykes, Gani, & Vally, 2016) Besides these, the minimum and maximum score in these group of values

6.3.1. Teaching items and Parcelling

As previously mentioned, the main reason for choosing TIMSS 2015 as a source of data in the present study is that TIMSS 2015 provides the data from teachers' responses. In these questionnaires, teachers were asked to report the frequency of these teaching activities in their science lessons. Teacher responses on applications of teaching practices made it possible to have more reliable data in terms of the implementation of teaching practices in classrooms. The items that represent the teaching practices have been investigated in a previous study of Gao (2014). (See **Table 3**) Even though, there have been small differences between teachers` questionnaires since the previous study used data from an earlier TIMSS study (2011), the same items still were available in more recent TIMSS 2015. Nevertheless, those items were both theoretically and statistically challenged.

Table 3 Teaching Items which has been used in a study from Gao,2014

Inquiry-Based Instruction
1) Relate the lesson to students' daily lives
2) Use questioning to elicit reasons and explanations
3) Ask students to observe natural phenomena and describe what they see
4) Ask students to design or plan experiments or investigations
5) Ask students to conduct experiments or investigations
6) Ask students to give explanations about something they are studying
7) Ask students to relate what they are learning in science to their daily lives
8) Ask students to do fieldwork outside of class
Traditional Didactic Teaching
9) Summarize what students should have learned from the lesson
10) Ask students to watch me demonstrate an experiment or investigation
11) Ask students to read their textbooks or other resource materials
12) Ask students to memorize facts and principles
13) Ask students to use scientific formulas and laws to solve routine problems
14) Ask students to take a written test or quiz
Affective teaching practices
15) Encourage all students to improve their performance
16) Praise students for good effort
17) Bring interesting materials to class

In TIMSS 2015, when responding these items, teachers were asked to choose one of four levels: 1) "in every lesson or almost every lesson," 2) "in about half the lessons," 3) "in some lessons," and 4) "never." To prepare for further analysis, the answers to each item were recoded to reverse the rank of using these instructional practices so that larger numbers illustrate higher frequency while smaller numbers classify lower frequency.

The first opposition that those items face, based on characterizations of traditional didactic and inquiry-based approaches which are made in the literature review section. Within this it is aimed at whether these items actually represent those teaching approaches. The second item in table 5 that singled out since it was not included in the TIMSS 2015 questionnaire. Also, the items that are located under 'Affective teaching practices' were not included in the present study since it might yield to another argument whether these teaching practices are 'affective'. Besides, adding these items necessarily may not bring contribution in terms of analysis in this study.

Item parcelling first introduced by Cattell in 1956 and since then it has been used in empirical SEM analyses frequently. It refers to aggregating single items into one or more parcels which replace these items as indicators of the latent constructs. (Matsunaga, 2008) In other words, through item parcelling the new variables are computed by taking sums or average across a few items. The use of the parcelling technique in SEM analysis inherently brings some benefits mainly due to its reducing model complexity which refers to a smaller number of indicators of a latent factor. (Nasser & Takahashi as cited in Matsunaga, 2008) Moreover, researchers have noted that use of parcels help reach optimal reliability, avoid violation of normality assumptions (particularly when the individual items are measured with a limited number of response categories) reduce the requirements on sample sizes, reduce influences of individual items' systematic errors on the model estimation, and obtain better model-data fit. In light of these, the 'Correlational Algorithm' method for parcelling applied. According to this method, first bivariate correlations were computed per scale. (See **Appendices 4 and 5**) The teaching items that showed high correlation were examined and the items with higher correlation are paired until there is no unassigned item left.

Table 4 Descriptive Statistics of independent parcelled items

Constructs	Parcelled Variables	Variables	N	Mean	Std. Deviation
Traditional Didactic Approach (TRA)	TRA1	Listen to me explain new science content	4578	3.0163	0.592
		Watch me demonstrate an experiment/ investigation			
	TRA2	Memorize rules, procedures, and facts	4577	2.8977	0.759
		Read their textbooks or other resource materials			
	TRA3	Use scientific formulas and laws to solve routine problems	4574	2.5535	0.585
		Take a written test or quiz			
Inquiry-based Approach (INQ)	INQ1	Relate the lesson to students' daily lives	4605	2.6014	0.534
		Do field work outside of class			
	INQ2	Observe a phenomenon and describe	4604	2.6999	0.633
		Ask students to decide their own problem-solving procedures			
	INQ3	Conduct experiments or investigations	4578	2.4506	0.613
		Interpret data from experiments/ investigations			
		Use evidence from experiments/ investigations to support conclusions			

Cronbach's alpha coefficient was derived in order to evaluate the reliability of such scales. The Cronbach's alpha is often used as a measure internal of consistency for multi-item scales and examines inter-item correlations by measuring the correlation of each item with the sum of all other items. (Cohen et al. as cited in Neuschmidt, 2018) A Cronbach's alpha over than 0.9 is considered as excellent reliability, between 0.7 and 0.9 high reliability, from 0.5 to 0.7 moderate reliability and below 0.5 low reliability. (Hinton, 2014) The Cronbach's alpha for each country presented in ***Appendix 6***, it's also referred in the **Results** section.

6.3.2. Control Variables

In order to examine the effects of instructional practices on academic performance, it is necessary to take other possible predictors of achievement into account as instructional practices are obviously not the only predictors. The studies established that Social Economic Status has the strongest predictor of academic achievement compared to factors such as ethnicity, age, gender which are thought to be associated with achievement. (Byrnes & Miller, 2007; Ma, 2000; Strand, 2014) The assumption of Socio-economic status being the best predictor of academic achievement has excessively evidenced. Even studies that challenged the magnitude of the relationship between SES and academic achievement have exposed the significance of this relation. (Sirin, 2005; White, 1982) Furthermore, Sirin (2005), proposed that researchers should not discuss only the context but must actually measure and evaluate the social and economic context in relation to their special area of interest.

The relationship between self-confidence and academic achievement has been a subject of education research for decades. The findings from earlier studies ranged from a strong negative correlation to a strong positive correlation. (Cheema & Skultety, 2017) According to relatively recent studies student self-confidence is a strong predictor of academic achievement. An empirical study on 15-year-old students from the US illustrates that self-confidence in science is significantly related to academic achievement. (Cheema & Skultety, 2017) Moreover, self-confidence is noted as one of the factors that predict key performance indicators among undergraduate students. (Nicholson, Putwain, Connors, & Hornby-Atkinson, 2013) Findings from a study that using the results from TIMSS 2011 assessment on Korean students resulted that high achieving students are likely to report that they learn things quickly and they do well in science. On the other hand, students who expressed a negative comparison of themselves to others tend to obtain lower academic achievement scores.(House & Telese, 2017) These findings can be based upon Bandura (1997) argument that students who believe that they have the capability to succeed in science tend to show greater interest in their schoolwork, persevere when confronted with difficult problems, and put forth greater effort in completing work. To be able to control the effects of student self-confidence in this study, a scale created by TIMSS was used. In the *Student Confident in science* scale students were scored according to their responses to seven statements. (see **Appendix 2**)

Table 5 Descriptive Statistics of Student Confident in science scale

Countries	Mininum	Maximum	Mean	Std. Deviation
Chile	8.10	11.56	9.8731	.70720
Egypt	9.15	15.01	10.5835	.79717
England	6.30	15.30	9.9131	1.24266
Italy	7.40	12.48	10.3346	.69582
Japan	6.57	9.96	8.6064	.52081
New Zealand	7.78	11.61	9.6471	.64239
Norway	8.30	12.27	10.4967	.71342
Singapore	6.58	12.03	9.6461	.77787
South Africa	6.83	12.01	10.2060	.67568

***Note**: Student Confidence in science scale was not conducted in Russia, Lithuania and Slovenia.*

As well as other studies, in a study on TIMSS 2007 data, Social Economic Status and student self-confidence were found to the strongest positive predictor of student science achievement. (Mohammadpour, Shekarchizadeh, & Kalantarrashidi, 2015)In the light of these findings, SES and student self-confidence were controlled in order to measure whether the teaching approaches have actually impact on student achievement as these variables are available in TIMSS 2015. The number of books in the home is used as a proxy for socioeconomic status in a report from Thomson, Wernert, O'Grady, and Rodrigues (2016) In recent years, however, researchers have emphasized the significance of various home resources as indicators of family SES background (Coleman; Duncan & Brooks-Gunn; Entwisle & Astone as cited in Sirin, 2005). These resources include household possessions such as books, computers, and a study room, as well as the availability of educational services after school and in the summer. (McLoyd, 1998; Eccles, Lord, & Midgley, 1991; Entwisle & Astone as cited Sirin (2005). In the end, as a representative for SES, the present study used the variable 'number of books at home' for which students were asked to choose between five answers which were coded as: (1 = 0–10; 2 = 11–25; 3 = 26–100; 4 = 101–200; 5 = over 200**). Table 6** below the descriptive statistics of 'Number of books' at home scale per country aggregated from student level to teacher level for each country.

Table 6 Descriptive Statistics of 'Number of books et home' scale

Countries	Mininum	Maximum	Mean	Std. Deviation
Chile	1.25	4.22	2.5302	.62662
Egypt	1.29	3.19	2.1609	.34821
England	1.00	5.00	2.8739	.86570
Italy	1.00	4.59	2.9953	.61226
Japan	2.13	4.43	3.1057	.40747
Lithuania	1.20	4.42	2.6889	.53269
New Zealand	1.48	4.75	3.1359	.62037
Norway	2.11	4.54	3.2204	.45292
Russia	1.67	4.04	2.8684	.43286
Singapore	1.50	4.17	2.6904	.55873
Slovenia	1.91	4.40	2.9255	.36993
South Africa	1.16	5.00	1.9421	.47678

To examine the effects of these teaching practices on student achievement multiple regression analysis will be applied, controlling the students' SES, self-confidence.

6.3.3. Students' science achievement

The TIMSS assessments cover a wide range of topics in mathematics and science each includes a large number of mathematics and science items (about 350 to 450) across at the fourth and eighth grade levels, together with sets of questionnaires that gather information on the educational and social contexts for achievement. The science content for TIMSS 2015, 8th grade assessment was defined by four major content domains: biology, chemistry, physics, and earth science. TIMSS 2015 used a matrix-sampling approach that involves packaging the entire assessment pool of mathematics and science items at each grade level into a set of 14 student achievement booklets. Students were given only one of these booklets. TIMSS relies on item response theory (IRT) scaling to describe student achievement and this scaling approach used multiple imputation—or plausible values—methodology to obtain proficiency scores in mathematics and science for all students. In this regard, five plausible were composed for each student.

Plausible values should be not considered as test scores, they rather are imputed values that may be used to estimate population characteristics correctly. They can provide consistent estimates of population characteristics as long as the underlying model is correctly specified. Still, they are not generally unbiased estimates of the proficiencies of the individuals. (Yamamoto & Kulick, 2000) TIMSS 2015 provides a set of five plausible values. In the present study first, plausible value is used as a representative for students' academic achievement. Taking the average of the plausible values still will not yield suitable estimates of individual student scores. (Von Davier, Gonzalez, & Mislevy, 2009) The descriptive statistics of the first plausible value for 12 countries are present in **Table 7** below.

Table 7. Descriptive statistics of Science Achievement in 12 Countries, TIMSS 2015 (1st Plausible Value)

Countries	Min.	Max.	Mean	Std.Deviation
Chile	352.03	594.33	474.2385	59.30270
Egypt	225.10	554.59	380.4096	65.04394
England	311.75	722.65	532.4997	66.41774
Italy	351.58	580.45	499.7228	39.17006
Japan	493.15	666.57	572.4802	29.94519
Lithuania	298.57	639.46	496.5515	48.17681
New Zealand	323.95	673.97	513.5791	65.81271
Norway	408.63	598.77	508.5432	33.69553
Russia	413.54	697.99	546.9284	48.06193
Singapore	340.93	755.00	590.1732	77.62644
Slovenia	486.41	642.17	552.1147	25.97364
South Africa	215.97	673.23	367.1481	82.44787

Since these countries are included in this study in order to represent categories student achievement levels differ considerably. Students' science achievement in Slovenia result in a mean score of 552.11, with the lowest standard deviation among these 12 countries, 25.97. In contrast, student achievement in South Africa shows the standard deviation by 82.44. Also, South Africa has the lowest mean score. Interestingly, the second highest standard deviation, 77.62, after South Africa is observed in Singapore which is at top of science achievement with a mean score of 590.17. Other countries' mean scores for 1st plausible vale and standard deviation ranges between these numbers. Hereby, it is good to mention that TIMSS identified four points on the overall mathematics and science scales to serve as International Benchmarks So, the readers can understand what performance on the overall mathematics and science achievement scales signifies. The TIMSS International Benchmark scores are 625, 550, 475, and 400, which correspond to the Advanced International Benchmark, the High International Benchmark, the Intermediate International Benchmark, and the Low International Benchmark, respectively.

6.4. Analytical considerations

6.4.1. Structural Equation Modelling

To assess the relationship between teaching practices (independent variable) and students' achievement (dependent variable) Structural Equation Modelling (SEM) was chosen for analysis. SEM is a very common statistical modelling technique within behavioural sciences since the early 80s after it was introduced in behavioural and social research in the early 70s. (J. J. Hox & Bechger, 1998) Very briefly, SEM is a combination of factor analysis and regression or path analysis. Path analysis can be view as an extension of multiple regression and it allows us to consider more than one dependent variable at a time and allows variables to be both dependent and independent variables. (Streiner, 2006) Therefore, by extending path analysis, SEM makes it possible to see the relationship between theoretical constructs which are represented by latent factors, and at the same time, theoretical constructs can be treated as independent variables and eventually latent variables as well. Besides SEM allows for the use of multiple measures to represent constructs, it also provides the issue of measure-specific error. Thus, in SEM it is feasible to determine the construct validity of factors. Accordingly, in SEM, the evaluation of the model accurately becomes more complicated. For instance, in order to determine whether the model fits the data, researchers need to evaluate the multiple test statistics and a host of fit indices. (Weston & Gore, 2006) (see **Model evaluation** below)

In SEM, there are 3 symbols: rectangles to represent observed (measured) variables; circles to illustrate the errors and ovals to depict the latent constructs.(Streiner, 2006) In Figure 1, the factors below observed variables represent the measurement errors. They also are often displayed by arrows.

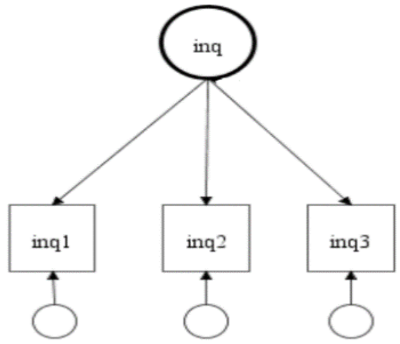

Figure 1 Measurement Model Example

SEM can also be described in terms of measurement and structural models. **Figure 1** illustrates an example of a measurement model from the present study. The measurement model of SEM helps to assure that combination of the observed variables explains the hypothesized latent constructs. (denoted by ellipses). Confirmatory Factor Analysis (CFA) is employed to test the relationship. In this example, the inquiry-based approach (latent variable) represented by three observed variables (parcelled). Instructional approaches (latent variables) can be considered as theoretical constructs represented by items (observed variables) from the teacher questionnaire in TIMSS 2015 since it is not possible to measure these constructs directly. The items which represent the instructional approaches were based on the previous literature and studies on the topic. The *structural model* refers to describing the interrelationships among constructs, both latent and manifest. SEM comprises measurement model (**See Figure 1** above) and structural model (**See Figure 2** below) Thus, a complete Structural Equation Modelling can be composed.

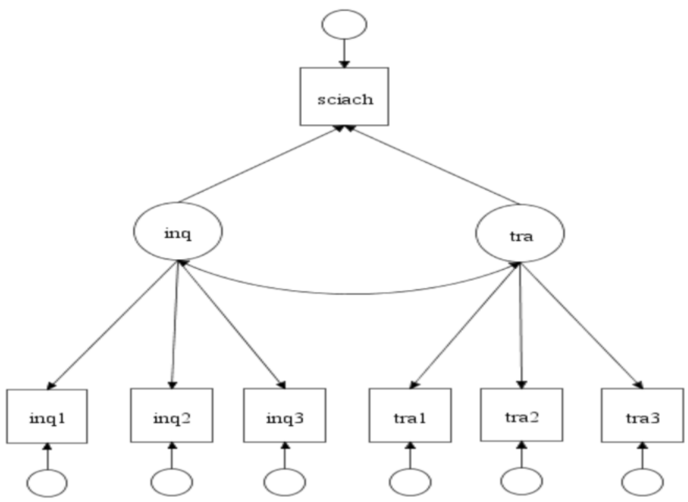

Figure 2 Structural Equation Model Example

Commonly, researchers follow a procedure in SEM model testing. It consists of 5 steps: model specification, identification, estimation, evaluation, and modification. Hoyle; Kaplan; Kline; Schumacker & Lomax as cited inWeston and Gore (2006)In the next sections, this procedure will be followed.

6.4.2. Model Specification

In SEM, model specification is an essential step that needs to be taken before starting the analysis. It consists of evaluating whether the research hypothesis on relationships among the observed and latent variables actually exist or not. It often roots in the theories and the findings from previous studies. (Weston & Gore, 2006) Once a model is specified, the factor loadings and (co)variances can be estimated. **Figure 3** portrays a hypothesized model where the relationships between Inquiry-Based approach (INQ), Traditional Didactic approach (TRA), Students' Socio-economic Status (SES), students' self-confidence (CON) and students' science achievement (SciAch). Before introducing the Students' Socio-economic Status (SES), students' self-confidence (CON) as control variables, a prior analysis was applied only with two latent constructs, the observed variables, and student science achievement. (SciAch). This was in order to observe the possible effects of SES and CON better. In the next chapters of the study, this model referred to as MODEL1. Therefore, the hypothesized model is shown in Figure 2 entitled MODEL2.

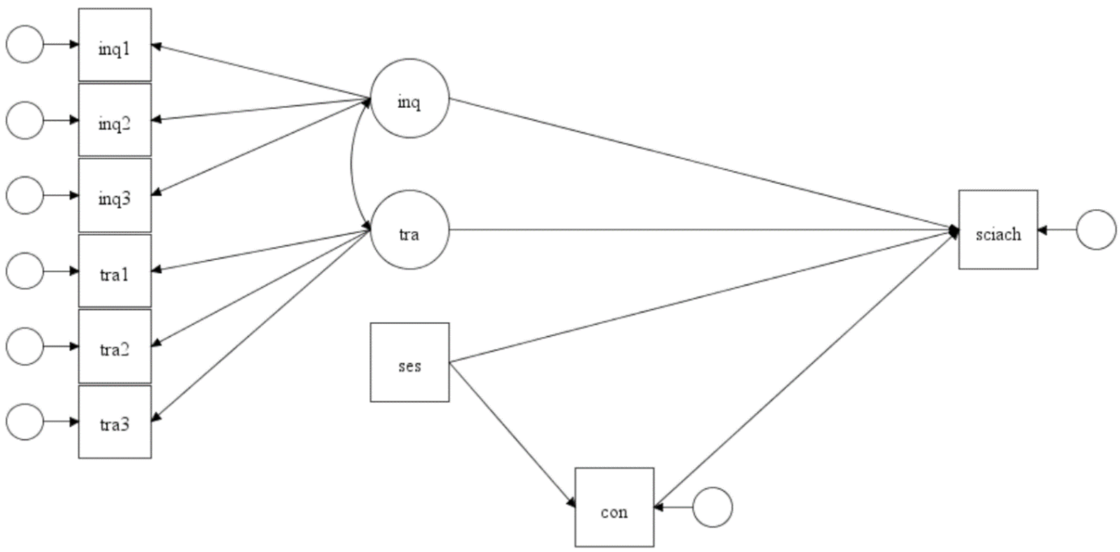

Figure 3 The Hypothesized Model (MODEL2)

As mentioned above, the hypothesized model is consisting of measurement models and a structural model. Inquiry based approach and Traditional didactic approach are measurement models that are to be measured by observed variables. Thus, factor loadings will be estimated for these two latent constructs for 12 countries included in the thesis.

6.4.3. Model Identification

In Structural Equation Modelling, model identification refers to having enough 'known' pieces of information to produce unique estimates of 'unknown' parameters. In SEM, 'knowns' are the variances, covariances, and means of observed variables on the other hand 'unknowns' are the model parameters to be estimated such as factor loadings, factor correlations, measurement errors. There are three possible situations in terms of model identification. Firstly, a model can be *"just-identified"* when the number of parameters -unknowns- to be estimated is the same as the number of known information. (e.g. variances, covariances, and means of observed variable) Secondly, when such a model has more `known' information than the `unknowns', then it is called "*overidentified*". Thirdly and lastly, a model is designated as a "*under-identified*" type of model in which there is not enough known information in order to determine the `parameters-unknowns'. The degrees of freedom demonstrate this difference between unknown and known parameters in the analysis. In the present study, (df)=12 for MODEL 1 and (df)= 24 for MODEL2, consequently, the models showed in Figure 2 are over-identified. For a further detailed explanation and mathematical background on the identification rules see Bollen and Davis (2009).

6.4.4. Model Estimation

After the model is specified and the identification of the model has been made, the following step is to estimate the model. Model estimation provides the estimated values for unknown parameters and the errors associated with estimated values. There are several SEM software programs that researches use to generates these estimates of the unknown parameters. In the present study, Mplus (Muthén & Muthén, 1998-2004) is used for the estimation of the model. Mplus, like many other SEM programs, operates *Maximum Likelihood (ML)* estimator as the default algorithm for missing data patters. ML technique is likely to be used by researchers when the data is normally distributed. The reason for that is that it is robust to moderate violations of the normality assumption.(Weston & Gore, 2006) However, there are other concerns when it comes to determining a type of estimator such as sample size. *"Maximum likelihood robust under a variety of conditions and is likely to produce parameter estimates that are unbiased, consistent, and efficient.*" (Myint Swe, 2013) In light of these pieces of information and considering the features of the given data, Maximum Likelihood estimation was chosen as the estimator type in the analysis through Mplus.

6.4.5. Model evaluation and model fit

Once the unknown parameters in a model estimated, to which degree that the data fits the model must be evaluated. As previously mentioned, the aim of SEM analysis is to estimate the unknown parameters by establishing a model in which associations among measured and latent variables efficiently reflect the observed associations in the data. (Weston & Gore, 2006) Therefore, the smaller difference between the observed and the model-implied variance/covariance matrices is, better the model. The chi-square ($\chi 2$) is one of the prominent fit indices. A value of statistically nonsignificant chi-square ($\chi 2$) suggests the model does fit the sample data absolutely. On the other hand, a significant $\chi 2$ indicates a poor model fit. However, chi-square ($\chi 2$) is not efficient when it comes to data with a large sample size due to detecting unimportant differences statistically significant. Also, it is only testing whether the model is an exact fit for the data. Otherwise, it cannot catch the good-acceptable model fit. (Weston & Gore, 2006) Hence, researchers use additional fit indices in terms of evaluating the goodness of a model fit. Nevertheless, chi-square ($\chi 2$) is still reported by researchers.

There are different types of fit indexes which appear to work well with different samples sizes, types of data, e.g. Normed Fit Index (NFI), Non-Normed Fit Index (NNFI, also known as TLI), Incremental Fit Index (IFI), Comparative Fit Index (CFI), and root mean square error of approximation (RMSEA) Among these indices, the TLI, CFI, and RMSEA are most favourable for researchers. (Schreiber, Nora, Stage, Barlow, & King, 2006) The given statistical software in the present study, Mplus, provides the standardised root mean square residual (SRMR) as well as TLI, CFI, RMSEA, and chi-square ($\chi 2$). In a case of statistically insignificant chi-square ($\chi 2$), researchers use such fit indices in order to interpret whether a model still is an acceptable fit to the observed data.

(CFI), the normal fit index (NFI), and the Tucker-Lewis index (Stockard et al.), also known as the non-normed fit index (NNFI) are called incremental or comparative fit indices. CFI and TLI based on a comparison between the implied model and a null model which suppose there are no covariances among the latent variables. Values for CFI and TLI fall within a range of 0 to 1.0, and values closer to 1.0 stand for a better fit. Recently, CFI ≥ 0.95 is recognised as indicative of good fit (Hu & Bentler, 1999) However, a value greater than 0.90 is still considered as an acceptable fit. The same cut-off values apply for both CFI and TLI. RMSEA is a residual fit index that average difference between the parameter estimates and populations covariance matrix. When there is no difference between the two models in terms of reflecting the observed data, the simpler model will have a lower RMSEA value. Therefore, RMSEA assists in a less complex model. According to Hu and Bentler (1999) for continuous data—RMSEA < .06 is efficient for a good fit or Steiger (as cited in Hooper, Coughlan, & Mullen, 2008) suggests a stringent upper limit of 0.07. On the other hand, Weston & Gore (2006) put forth the cut-off criteria '0.06' which could result in the incorrect rejection of acceptable models when samples sizes are smaller than n = 500. Thus, they further remarked that RMSEA values between .05 and .10 might be considered as an acceptable fit. SRMR index illustrates to what extent observed data and the model differs. Like RMSEA, a value lower than 0.06 for SRMR is acceptable as a good fit. Besides that, Hu and Bentler (1999) estimates (SRMR) < .08 adequate for a good model fit.

6.4.6. Model Modification

After the hypothesized model is identified and specified, a single level SEM analysis to be conducted in Mplus. Since the given data aggregated on the teacher level, a single level model as the default option in SEM was run. TYPE=GENERAL and maximum likelihood estimation are used in ANALYSIS command. ML is the default algorithm for use with missing data patterns under the assumption of normally distributed indicators. (See **Model Estimation** above)

6.5. Multilevel modelling

TIMSS data has a hierarchical structure which would allow for multilevel analysis. Since students are nested within classrooms, they are likely to have more in common than with students from other classrooms. Their teacher for example. This dependency needs to be handled statistically since ordinary one-level regression leans on the assumption that observations are independent (J. Hox, 2003). Besides a statistical advantage, multilevel modelling allows for a simultaneous study of higher and lower levels (e.g., micro and macro). For example, the relation between socioeconomic status and achievement can be different on the individual level and system level (between schools). In this thesis, however, the focus is mainly on differences between countries and what instructional practices teachers use. Teachers are generally independent from each other. When student data has been included in analyses, it has been aggregated to the teacher level. For example, the achievement score at the teacher level is the average achievement of all students in a classroom.

6.6. Ethical Considerations

All TIMSS and TIMSS Advanced 2015 participants were assured that their data would be confidential. Data security and confidentiality were maintained throughout all phases of the study, including data collection, data creation, data dissemination, and data analysis and reporting. (Johansone, 2016) Moreover, the TIMSS 2015 International Database is available in two versions: a public use version and a restricted use version. In the public use version, some variables are removed to minimize the risk of disclosing confidential information. Therefore, the present study does not struggle in terms of ethics as the ethical concerns are covered by IEA.

7. Results

At the first stage, an index, that includes all countries in TIMSS 2015, has been created to determine the usage of the two instructional approaches in each country.(See **table 8**) An overall mean score of instructional approaches was calculated in order to identify countries in terms of teacher's usage of both Inquiry-based and traditional didactic approaches in classrooms. The calculation of the country-wise mean scores was based on the teaching items shown in Table 3.

In Table 8, all participating countries in TIMSS 2015 were ranked according to the distance between the mean scores of these approaches. This distance calculated by extracting the mean score of inquiry-based teaching practices from the mean score for traditional teaching practices. Consequently, the countries that listed on top of the index refer to that the usage of traditional didactic teaching practices is higher than inquiry-based teaching practices in these countries. If teachers in one country use both approaches moderately, the distance will be close to zero. If teachers use inquiry-based teaching practices more, then the distance likely to be negative. Accordingly, the higher is distance is higher the usage of traditional teaching practices.

Table 8 Distance between mean scales (from high to low)

Country ID	N	Minimum	Maximum	Mean	Std. Deviation
Lithuania	860	-1.17	2.00	**0.6870**	0.47211
Armenia	554	-0.83	1.83	**0.6201**	0.45373
Russia	756	-1.67	2.00	**0.6073**	0.51651
Egypt	211	-1.00	2.00	**0.5812**	0.56278
Saudi Arabia	148	-0.50	1.83	**0.5784**	0.46203
Georgia	632	-1.00	1.75	**0.5695**	0.46588
South Korea	215	-0.67	2.00	**0.5620**	0.50210
Hungary	561	-1.50	1.83	**0.5348**	0.54407
Singapore	320	-0.83	2.00	**0.4734**	0.50656
Kazakhstan	788	-1.17	1.67	**0.4598**	0.47367
Iran	250	-0.83	1.83	**0.4216**	0.50821
Taiwan	207	-1.00	1.33	**0.3903**	0.41745
Morocco	705	-1.33	2.17	**0.3874**	0.56741
Japan	169	-1.33	1.83	**0.3844**	0.51960
Jordan	254	-1.17	1.50	**0.3797**	0.47056
Bahrein	187	-1.00	1.83	**0.3504**	0.54200
Lebanon	177	-1.33	1.90	**0.3465**	0.58395
Quebec, Canada	155	-1.00	1.33	**0.3290**	0.48385
South Africa	315	-1.33	1.50	**0.3275**	0.45807
Slovenia	533	-0.83	1.50	**0.3244**	0.43012
Italy	211	-1.33	1.67	**0.2855**	0.53735
Chile	162	-0.83	1.50	**0.2815**	0.44952
Abu Dhabi, UAE	153	-1.17	2.00	**0.2764**	0.57427
Israel	329	-1.17	1.83	**0.2676**	0.50999
Kuwait	184	-1.00	2.00	**0.2663**	0.54595
Qatar	236	-1.33	2.07	**0.2595**	0.54370
Botswana	159	-1.07	1.5	**0.2400**	0.53283
Turkey	218	-1.17	1.67	**0.2362**	0.53375
Malaysia	271	-1.17	1.33	**0.2185**	0.49873
United Arab Emirates	555	-1.83	2.00	**0.2126**	0.61146

Norway	185	-1.00	1.50	**0.2108**	0.39472
Sweden	282	-1.33	1.67	**0.1908**	0.47727
Hong Kong SAR	149	-0.83	1.17	**0.1535**	0.43456
Oman	343	-1.00	1.50	**0.1460**	0.47397
Canada	323	-1.50	1.50	**0.1189**	0.55992
Ireland	358	-1.67	1.50	**0.1106**	0.54102
Buenos Aires, Argentina	57	-1.17	1.17	**0.1082**	0.43926
Malta	299	-1.33	1.83	**0.0933**	0.53184
Thailand	205	-1.50	1.33	**0.0785**	0.48032
Dubai, UAE	215	-1.83	1.50	**0.0498**	0.65017
England	564	-1.33	2.17	**-0.0035**	0.46909
USA	419	-1.83	1.50	**-0.0115**	0.57798
Australia	766	-1.50	1.67	**-0.0188**	0.45960
New Zealand	296	-1.33	1.67	**-0.0436**	0.50666
Ontario, Canada	150	-1.50	1.50	**-0.0707**	0.55590

UAE: United Arab Emirates, USA: United States of America,

It can be seen that countries vary quite substantially with respect to their instructional approach. A general pattern is that countries in the West lean more towards an inquiry-based approach whereas countries in the East are more inclined to use a traditional approach. Overall, the teachers use teaching practices that represent the traditional didactic approach more than inquiry-based teaching practices. There are only 5 countries or regions that showed positive weight for using teaching practices that referred to the inquiry-based approach. Interestingly, all of them are English-speaking countries. This contradicts the assumption that the inquiry-based approach is the dominant instructional approach within classroom practices. On the other hand, this may also be interpreted in favour of the traditional didactic approach, that is still being valid and recognized in science teaching.

7.1. The Results of SEM

In the next step, this index has been used for an in-depth study of countries that adopts three different instructional approaches. The categorisation was made with respect to the weight of the instructional approaches the teachers in these countries use in classroom practice. These levels described as ''High Inquiry- Low traditional'', ''High traditional- Low inquiry'' and ''Moderated approach''. The countries were chosen for ''High inquiry- Low traditional'' were New Zealand and England. Lithuania and Russia represent '' High Traditional- Low inquiry''. Lastly, Chile and Slovenia included in the study in order to exemplify the countries which have a more ''Moderated Approach''.

The present study initially involves a large dataset from 39 countries and 7 benchmarking entities in the early phase of analysis and scales down the scope of analysis to 12 countries. Therefore, to introduce such a model that represent these instructions for 12 countries with a good fit is likely to be challenging. Even so, there are some tendencies within the countries, inevitably some contradictory figures emerge in terms of a verified model for all countries. The method used for parcelling is presented in the **Teaching items and Parcelling** section. Also, **see Table 4** for the descriptive statistics of two constructs after parcelling. Exploratory factor analysis (Toropova, Myrberg, & Johansson) was run to be able to see the multi-dimensionality of these parcelled variables. EFA shows the underlying structure of the measured variables as well as underlying relationships between them. SPSS was used to run this analysis. The inquiry-based approach was comprised of 3 items reported on a 4-point Likert scale that explained 57% of the variance with factor loadings: 0.717, 0,725 and 0.817. On the other hand, the the traditional- didactic approach was comprised of 3 items reported on a 4-point scale that explained 56% of the variance with factor loadings: 0.719, 0.753 and 0.780. Moreover, the figures for Cronbach's alpha were calculated for both instructional approaches in each country. As the indicators shown in **Appendix 6**, the alpha ranges from 0.516 to 0.749. This refers to that data from 12 countries shows moderate reliability. The relatively low figure can be partly attributed to the possible negative effect of a drop-down number of items after parcelling from 6-7 to 3 for both factors.

So far, I have identified differences in instructional practices across countries. Based on this first investigation, I have selected groups of countries for further scrutiny. The latent models of instructional approaches were tested with the aid of EFA and CFA, and the final measurement models were adopted. (See **Table 4**) The factor loadings and model fit for these models are presented in **Appendix 6**. In the next analyses I will focus there is a relationship between instructional approaches and student achievement. In each part of the analysis, there are two estimated models which are **presented** in the methodology section. Briefly, in MODEL1, the relationship between two latent constructs of teaching approaches and student achievement was investigated; in MODEL2, besides this relationship, the effects of SES and student confidence were controlled. Initially, MODEL1 was applied to those 12 countries that were chosen for both categories. However, MODEL 2 is only applied to the countries in which the significant relationship between student achievement and instructional approaches were found in MODEL1.

Prior to sharing the results of structural relationships, the model fit indices of measurement models are exhibited for each country. (See also **Appendix 6** for MODEL1 fit indices for 12 countries). Model fit indices of MODEL2 only shown for dual of countries that at least one of them resulted in a significant relationship between one of the instructional approaches and student achievement. In this sense, the standardized model estimates are also only presented when it is relevant.

7.1.1. Different intensities in using a type of approach

7.1.1.1. Inquiry-based approach found to be used more

According to the index created based on teacher's usage of types of instructional practices in the classroom, New Zealand and England are picked to represent the countries in which the inquiry-based approach found to be used more likely. (see Table 2)

Table 9 Model Fit Indices for New Zealand and England in MODEL2

		MODEL2	
	Good Fit (Acceptable Threshold)	New Zealand	England
χ2 (df)		40.818 (24)	95.699 (24)
CFI	≥.95 (≥.90)	0.974	0.937
RMSEA	≤ 0.08 (≤ 0.10)	0.046	0.062
SRMR	≤ 0.06 (≤ 0.10)	0.045	0.061

*RMSEA=Root Mean Square Error of Approximation
*CFI=Comparative fit index
*SRMR=Standardized Root Mean Square Residual
*χ2 – Chi-square mean, (Df)= degrees of freedom

Table 9 above illustrated the fit indices of New Zealand and England for MODEL2. (See also **Appendix 6** for MODEL1 fit indices) chi-square in both models for both countries is significant. CFI in MODEL1 for New Zealand is very slightly above the figure for an acceptable fit, 0.903. Further, RMSEA and SRMR are under the cut-off value indicating a good model fit. Overall, the scores show an acceptable fit in MODEL1. In MODEL2, New Zealand has a quite high CFI compared to MODEL1, 0.974. Since RMSEA and SRMR scores refer to good fit for New Zealand in MODEL2, MODEL2 fits pretty good to the New Zealand data. The data from England shows an acceptable fit in MODEL2 in terms of CFI, RMSEA and SRMR, unlike showing a poor fit in MODEL1 due to a CFI score slightly lower than '0.9'.

Table 10 Standardised Model Estimates for New Zealand and England

Parameter Estimate		**Standardized of MODEL1**		**Standardized of MODEL2**	
		New Zealand	England	New Zealand	England
Measurement Model Estimates					
Inquiry Based	INQ1	0.534*	0.440*	0.528*	0.453*
	INQ2	0.684*	0.669*	0.687*	0.665*
	INQ3	0.579*	0.479*	0.592*	0.473*
Traditional	TRA1	0.367*	0.411*	0.340*	0.412*
	TRA2	0.504*	0.477*	0.472*	0.477*
	TRA3	0.647*	0.724*	0.686*	0.724*
Structural Model					
INQ ON PV1		0.019	0.313*	0.005	0.128
TRA ON PV1		0.096	-0.045	0.094	-0.013
INQ WITH TRA		0.485*	0.723*	0.498*	0.721*
SES ON PV1		N/A	N/A	0.808*	0.618*
CON ON PV1		N/A	N/A	0.127*	0.258*
SES ON CON		N/A	N/A	0.240*	0.307*

(*) Note. * $p < .05$, represent significant results.

As in the previous countries, the factor loading of each parcelled variable that represents the latent constructs is significant. Also, they are higher than 0.4 with an exception of TRA1 being lower than 0.4 for New Zealand in both models. Even, there is no statistically significant relationship between student science achievement and instructional approaches were found for New Zealand in both models. In MODEL1, a positive relationship between inquiry-based approach and student science achievement is indicated for England with a score of '0.313'. However, when the effects of SES and student confidence are controlled in MODEL2, this relationship is found to be statistically insignificant. Additionally, students' SES and self-confidence in learning science were found to be significantly and positively related to their science performance in both countries.

In the structural model a high correlation between the independent latent constructs is observed in the data from England. (0.721) This figure hints that the teachers use both instructional approaches, and therefore they correlate highly. However, in a regression framework, highly correlated independent variables may cause multicollinearity, which could confuse the interpretation of the parameter estimates. Therefore, a regression analysis was run separately between each latent construct and student achievement as well as controlling the effects of SES and CON. In this analysis, a statistically significant positive effect of the inquiry-based approach on student achievement is found (0.113) whereas the effects of SES by 0.619 and CON by 0.258 are also observed. On the other hand, the traditional didactic approach still did not show any significant effect on student achievement.

7.1.1.2. Traditional approach found to be used more

In the present study, Lithuania and Russia represent the countries that traditional instructional practices are likely to be used more than inquiry-based instructional practices as they respectively take place 1st and 3rd place in the Table 2. Model fit Indices for Lithuania and Russia are shown in the **Appendix 6**. They both have significant Chi-Square. However, they result in very high CFI score, respectively 0.982 and 0.992. Moreover, low RMSEA and SRMR scored are observed for both countries. Therefore, it can be concluded that MODEL1 fit to data from Lithuania and Russian very well.

Table 11 Standardised Model Estimates for Lithuania and Russia

Parameter Estimate		**Standardized**	
		Lithuania	Russia
Measurement Model Estimates			
Inquiry Based	INQ1	0.551*	0.513*
	INQ2	0.668*	0.731*
	INQ3	0.788*	0.571*
Traditional	TRA1	0.686*	0.570*
	TRA2	0.346*	0.467*
	TRA3	0.583*	0.633*
Structural Model			
INQ ON PV1		-0.021	0.080
TRA ON PV1		-0.069	0.058
INQ WITH TRA		0.760*	0.574*
SES ON PV1		N/A	N/A
CON ON PV1		N/A	N/A
SES ON CON		N/A	N/A

(*) Note. * $p < .05$, represent significant results.

The factor loadings for parcelled observable indicators are significant for both countries and almost all of them are higher than 0.5, except TRA2. It is evident that parcelled indicators have an appropriate measurement of latent factors for Lithuania and Russia. On the other hand, in the structural model, the inquiry-based approach and the traditional didactic approach have a non-statistically significant effect on student science achievement in Lithuania and Russia. MODEL2 is not applied to the data from these countries since there are no statistically significant results observed. Still, a high correlation between two latent constructs was found by 0.76 which indicates multicollinearity in the data from Lithuania. Therefore, regression analyses were run in order to particularly see the relationships between the two constructs with student achievement. In these analyses, it is shown that the inquiry-based approach is negatively influencing Lithuanian students' science achievement. (-0.102) Moreover, a similar negative effect of the traditional didactic approach is also found -0.119. In both analyses, student science achievement is found to be explained by SES to the high extent.

7.1.1.3. Moderated approach

Chile and Slovenia are included in the present study in order to analyse whether moderated approach in terms of instructional approaches lead to varied student science achievement. The model fit indices in MODEL1 for Chile and Slovenia can be seen in the **Appendix 6**. Both Chi-Squares are significant. So, CFI, RMSEA and SRMR scores are to analyse to decide whether MODEL1 is at least acceptable fit to the data from Chile and Slovenia. Chile has a relatively higher CFI which is 0.916 and RMSEA, slightly lower than 0.10. On the other hand, Slovenia shows a very high CFI (0.965) and low RMSEA and SRMR scores that is necessary for a good fit.

Table 12 Standardized Model Estimates for Chile and Slovenia

Parameter Estimate		**Standardized**	
		Chile	Slovenia
Measurement Model Estimates			
Inquiry Based	INQ1	0.514*	0.469*
	INQ2	0.806*	0.681*
	INQ3	0.751*	0.665*
Traditional	TRA1	0.647*	0.695*
	TRA2	0.491*	0.240*
	TRA3	0.553*	0.523*
Structural Model			
INQ ON PV1		-0.252	-0.022
TRA ON PV1		0.057	0.049
INQ WITH TRA		0.971*	0.808*
SES ON PV1		N/A	N/A
CON ON PV1		N/A	N/A
SES ON CON		N/A	N/A

(*) Note. * $p < .05$, represent significant results.

The factor loadings of observed parcelled indicators are quite high and statistically significant for both latent constructs for Chile, ranging from 0.491 to 0.806. They are also acceptable measurements for Slovenia, expect TRA2 being quite low, 0.240.

Both latent constructs result in statistically insignificant values in terms of the relationship between the instructional approaches and students' science achievement. This finding eliminates the necessity of controlling SES and student confidence effects. Accordingly, MODEL2 has not been applied to data from Chile and Slovenia. One of the outcomes from MODEL1 was that two latent constructs showing high correlations in both countries. Therefore, due to multicollinearity concerns for both countries, it was decided to run regression analyses for each latent construct separately and student achievement. As a result of these analyses, still, neither of the two latent constructs showed a statistically significant relationship with student achievement.

7.1.2. Comparison within the levels that varies in the weight of the instructional approaches

The only statistically significant result in terms of the relationship between instructional approaches and the students' science achievement is observed in the data from England, in MODEL1. The result suggests that the inquiry-based approach has a positive impact on students' science achievement. (0.313). However, when the effects of SES and student confidence are controlled in MODEL2, this positive correlation between inquiry-based approach and student science achievement is not found on the data from England, whereas SES is highly associated with student science achievement (0.618). In other words, 62 percent of the variance in English students' science achievement can be explained by students' socioeconomic status and 26 percent resulting from the student confidence. Moreover, regression analysis between inquiry-based and student achievement resulted in a positive effect of the inquiry-based approach. But, in New Zealand, another country where the inquiry-based approach is more likely to be used in the classroom practices, there is no such statically significant relationship found between inquiry-based approach and students' science achievement. On top of this, when the same regression analysis applied to the data from Lithuania, the inquiry-based approach is found to be negatively influencing Lithuanian students' science achievement. (-0.102) However, the positive influence of SES and student confidence on students' science achievement are still observed in all countries. Given the analysis, the usage of a type of instructional approach more than another does not lead to this approach having a positive or negative influence on student achievement.

7.1.3. Difference performance levels

7.1.3.1. Teaching Approaches in Countries with low achievements

Egypt and South Africa were chosen to represent the countries with low student achievements. (See **Sampling**) Egypt and South Africa indicated over 0.95 CFI score, respectively 0.967 and 0.958. RMSEA for South Africa is slightly lower than 0.08 which is required for a good fit, as well as, similar SRMR, 0,040 for Egypt and 0.034 for South Africa. Consequently, model fit indices illustrate the MODEL1 fit the data from Egypt and South Africa very well. (See **Appendix 6**)

Table 13 Standardised Model Estimates for Egypt and South Africa

Parameter Estimate		**Standardized**	
		Egypt	South Africa
Measurement Model Estimates		Factor Loadings	Factor Loadings
Inquiry *Based*	INQ1	*0.617	*0.645
	INQ2	*0.638	*0.677
	INQ3	*0.722	*0.790
Traditional	TRA1	*0.460	*0.589
	TRA2	*0.357	*0.495
	TRA3	*0.487	*0.650
Structural Model			
INQ ON PV1		0.378	0.680
TRA ON PV1		-0.301	-0.673
INQ WITH TRA		*0.782	*0.947
SES ON PV1		N/A	N/A
CON ON PV1		N/A	N/A
SES ON CON		N/A	N/A

(*) Note. * p < .05, represent significant results.

The factor loadings of parcelled items that represent the inquiry-based approach and the traditional didactic approach are statistically significant. According to Muthén and Muthén (2002), factor loadings ≥ 0.4 are acceptable, ≥ 0.6 – good. Yet, most of these factor loadings are above 0.6 and the rest are higher than 0.4 with one exception which is 0.357. (See Table 13 above) Conversely, the relationship between both instructional approaches and student achievement for Egypt and

South Africa is statistically insignificant. Therefore, it was not worthy to test MODEL2 for both countries. However, multicollinearity issues appear since the two latent constructs show a high correlation for Egypt and South Africa, respectively, 0,782 and 947. Therefore, the relationship between each latent construct and student achievement investigated separately. Nevertheless, no statistically significant relationship was found in both countries for any of the two latent constructs.

7.1.3.2. Teaching Approaches in Countries with medium achievement

Norway and Italy were included in the study to exemplify the countries that medium place in the science achievement scale from TIMSS 2015. (See **Table 3**). In the Table 10 below, model fit indices from MODEL1 and MODEL2 are shown for Norway and Italy. The Chi Square values are significant in both models for both countries. Therefore, analysis of CFI, RMSEA and SRMR values is essential. The CFI scores for Italy in MODEL1 and MODEL2 are respectively 0.936 and 0.965. RMSEA is lower than 0.8 in both models for Italy as well as SRMR scores lower than 0.6. From the analysis, it can be concluded that both MODEL1 and MODEL2 fit Italy data very well. On the other hand, MODEL1 result in a poor CFI score for Norway, being below 0.90 although RMSEA, SRMR refers to an acceptable fit. Still, compared to MODEL1, in MODEL2, Norway has a better model fit for with a relatively higher CFI, 0.924.

Table 14 Model fit indices for Norway and Italy

		MODEL2	
	Good Fit (Acceptable Threshold)	Norway	Italy
χ2 (df)		43.014 (24)	38.963 (24)
CFI	≥.95 (≥.90)	0.924	0.965
RMSEA	≤ 0.08 (≤ 0.10)	0.061	0.052
SRMR	≤ 0.06 (≤ 0.10)	0.070	0.048

*RMSEA=Root Mean Square Error of Approximation
*CFI=Comparative fit index
*SRMR=Standardized Root Mean Square Residual
*χ2 – Chi-square mean, (Df)= degrees of freedom

First of all, the factor loadings of parcelled items for Italy and Norway in both models, are statistically significant. Besides, they are quite high for Italy in both models and therefore it can be said that parcelled observable indicators have an appropriate measurement of latent factors for Italy. However, TRA1 is below 0.4 in both models for Norway, and that is violating the acceptable fit. Furthermore, the estimates for Norway that illustrates the relationship between the latent factors and student achievement for statistically insignificant in both models.

In MODEL1, the estimate that presents the relationship between student achievement and the traditional didactic approach is '-0.242' and is statistically significant. Moreover, In MODEL2, the traditional didactic approach was found to be negatively related to student science achievement even though it loosens up a little from ' -0.242' to '0.210'. In MODEL2, SES and student confidence are controlled in order to extract the effects of these parameters which likely impact student achievement. In the present analysis as a matter of fact, the effect of SES on student achievement has been evidenced in Italy and Norway data. (See **Table 15** below

Table 15 Standardised Model Estimates for Norway and Italy

Parameter Estimate		Standardized of MODEL1		Standardized of MODEL2	
		Norway	Italy	Norway	Italy
Measurement Model Estimates					
Inquiry Based	INQ1	0.484*	0.609*	0.482*	0.604*
	INQ2	0.665*	0.696*	0.663*	0.698*
	INQ3	0.718*	0.535*	0.721*	0.538*
Traditional	TRA1	0.275*	0.584*	0.278*	0.599*
	TRA2	0.520*	0.620*	0.515*	0.621*
	TRA3	0.728*	0.718*	0.730*	0.701*
Structural Model					
INQ ON PV1		0.064	0.069	0.004	0.089
TRA ON PV1		-0.104	-0.242*	-0.067	-0.210*
INQ WITH TRA		0.501	0.582*	0.502*	0.586*
SES ON PV1		N/A	N/A	0.582*	0.739*
CON ON PV1		N/A	N/A	0.161*	0.028
SES ON CON		N/A	N/A	0.113	0.254*

(*) Note. * $p < .05$, represent significant results.

7.1.3.3. Teaching Approaches in Countries with High achievement

Singapore and Japan are selected to represent the countries which performed high in science achievement in TIMSS 2015. Singapore placed first and Japan second in science achievement scale. (See **Appendix 1**) The model fit indices of MODEL1 for Singapore and Japan are shown in **Appendix 6**. Chi-Square for both of the countries is significant. CFI for Singapore is very slightly lower than the acceptable figure, '0.896'. Also, RMSEA is just in the limit for an acceptable fit whilst SRMR is enough for a good fit as it is below '0.6'. On the other hand, Japan results in a 0.963 CFI score which refers to a good fit. Besides, RMSEA and SRMR scores for the data from Japan are below the limits that are necessary for a good model fit, respectively 0.058 and 0.054.

Table 16 Standardised Model Estimates for Singapore and Japan

Parameter Estimate		Standardized	
		Singapore	Japan
Measurement Model Estimates			
Inquiry Based	INQ1	0.570*	0.495*
	INQ2	0.646*	0.784*
	INQ3	0.706*	0.735*
Traditional	TRA1	0.501*	0.560*
	TRA2	0.631*	0.615*
	TRA3	0.715*	0.646*
Structural Model			
INQ ON PV1		0.048	-0.106
TRA ON PV1		-0.074	0.010
INQ WITH TRA		0.576*	0.518*
SES ON PV1		N/A	N/A
CON ON PV1		N/A	N/A
SES ON CON		N/A	N/A

(*) Note. * $p < .05$, represent significant results.

Table 16 presents the standardized estimates of MODEL1 for Singapore and Japan. The factor loadings of the parcelled variables for both instructional approaches are statistically significant and are above the acceptable figure '0.4' with the lowest score '0.495'. Nevertheless, the figures of the structural model do not show any statistically significant relationship between latent constructs and student achievement in Singapore and Japan. The correlations between latent constructs in both countries are relatively expectable measures. So, multicollinearity is not observed in these countries unlike some of the instances above.

7.1.4. Comparison within the different performance levels

The statistical analysis of the relationship between the two instructional approaches and student science achievement for countries with different performance levels is demonstrated by the figures above. The only statistically significant relationship was observed in the data from Italy. The estimates imply that the traditional didactic approach has a negative impact on science achievement for Italian 8^{th} grade students (Estimate = - 0.282) On the other hand, there is no such statistically significant relationship was found on the data from Norway. Hence, a conclusion stating that the traditional didactic approach has a negative impact on 8^{th} grade student science achievement in countries with medium achievement cannot be made. Moreover, as there is no statistically significant relationship was found in low performing countries and in high performing countries, it is not possible to make an inference on whether these two instructional approaches show various effects on different performance levels.

8. Discussions

In this section, besides discussing the results from the present study, I will address what these results translate to in a broader context. Finally, the limitations of this study will be mentioned in terms of both the data and the analysis.

This study aimed at examining the extent of the inquiry-based and traditional didactic approaches across countries in TIMSS 2015 and finding out whether these approaches have an impact on students' science achievement. The findings of this thesis conclude that neither the inquiry-based approach nor the traditional didactic approach has a statistically significant relationship with students' science achievement. There were two instances out of 12 countries where such a relationship was observed in different models. However, when the validity of these instances taken into account in terms of rationality and statistics, it is seen that these found relationships are not meaningful to able to make inference in a large context. These findings are drawing a different picture than most of the studies mentioned in the literature review section. It may go beyond questioning the superiority of a type of instructional approach to another and discuss that instructional approaches do not have an impact on student science achievement as expected. This seems like a reasonable conclusion considering the observation of the positive impact of both SES and student confidence, in the present study. Nevertheless, the number of studies that found no significant effect of instructional approaches on student achievement are not many. The most similar findings to the present study were reported from N. Lederman, Lederman, Wickman, and Lager-Nyqvist (2007) and in a follow-up study they conducted in 2008. In both studies, Lederman et.al. found no impact of instructional approaches in the post-tests.

The studies which argue the inquiry-based approach is more effective on science achievement than the traditional-didactic approach or any other, conceptualize the inquiry-based approach differently. There are various definitions, perceptions, and applications of the inquiry-based approach. For instance, Abrams et al. (2007) introduces the levels of inquiry which was adapted from Schwab (1962) and Colburn (2000a). In brief, these levels based on the guidance given by the teacher. Level 3 means almost no guidance by the teacher, while level 0 designates the minimum level of inquiry. Even the researchers who support the argument of the inquiry-based approach being more effective agree on that no guidance is not ideal way in science teaching and teacher guidance is essential. (Blanchard et al., 2010; Furtak et al., 2012; Lazonder & Harmsen, 2016). Considering these together with the findings of the present study, these frameworks do not oppose to key elements of traditional forms of education, such as direct instruction.

I also would like to discuss shortly what the index is shown in **Table 8** in which all countries in TIMSS 2015 sorted based on overall mean scores that were calculated according to the usage of two instructional approaches, might tell us. Especially, comparing this index with TIMSS 2015 Science Achievement Distribution (See **Appendix 1**) gives a portray that can be easily understood. Singapore that showed the highest science achievement in TIMSS 2015 is ranked as 9th in the given index. This means that, even though it is not at the top, a teacher in a science class in Singapore predominantly uses teaching practices related to the traditional didactic approach more than the ones that can be referred to the inquiry-based approach. If we look at Japan, as it follows Singapore by science achievement, we see that it is ranked at 14 in the given index. This number corresponds to the less dominant usage of traditional-didactic teaching practices. Another interesting country to look at is Russia by being close to the top of both rankings. This might imply that the teachers in Russia use a more traditional-didactic approach in science lessons and thus, 8th grade students in Russia perform very well. The instances that will speak for the favour of the traditional-didactic approach in the same regard can be derived from this comparison of the tables. However, there are also some examples that will contradict this outcome. For example, such countries Egypt and Saudi Arabia show poor science achievement in TIMSS 2015. These countries take place nearly on top of the index on the usage of the instructional approaches. This can be easily interpreted as Egypt and Saudi Arabia follow the traditional didactic approach, and this leads to poor science performance. Therefore, such conclusions that can be made both favours of the inquiry-based approach or traditional didactic approach are not reliable. The findings of the present study briefly reject the conclusions that claim the superiority of one of two approaches to another in terms of being effective on students' science achievement.

8.1. Cognitive Load Theory

The main reason for the choice the Cognitive Load Theory as the theoretical framework in this thesis was that any instructional approach would not disregard capabilities or limitations of the human cognition or at least, the ones that disregard would fail. Given the theoretical framework, CLT, one would anticipate that traditional didactic approach positively associated with the student science achievement. Moreover, according to CLT, the assumption of inquiry-based approach being more effective on student science achievement should have failed. However, the findings of the present study do not support this hypothesis. The only statistically relationship found between instructional approaches and student science achievement in data from Italy is a negative association. On the other hand, the inquiry-based approach is not found to be positively influencing student science achievement in any of the 12 countries either. Hence a claim that rejects the credibility of traditional didactic approach in science teaching would be unjustified. In short, the study's findings suggest that there is continuing need to clarify the relative merits of inquiry-based versus instruction approaches in science teaching.

Several reasons can be given in response to why discourses of CLT do not match with the findings of this study and why they should be challenged. Firstly, criticism towards minimal guidance during a learning process relies on the absence of necessary teacher intervention during learning. When the teacher is seen as the only source of knowledge then this assumption may be somehow more valid. However, this hypothesis neglects the dimension of interaction among students. Because, such assistance can be also given by more capable peers as suggested by Vygotsky. (as cited in Yilmaz, 2011) Okada and Simon (1997) reported that peer interaction results in a better performance of pairs than single students when forming a hypothesis.(as cited in M. Cohen, 2008) Therefore, this peer effect needs to be taken into account when considering the implications of the inquiry based approach. According to CLT, when the learners are left to discover a complex environment or a concept independently then they will end up with a heavy working memory which is undesired. Then, one may ask "Are the learners actually left alone in the inquiry-based teaching?" The answer to that question already argues the proposed hypothesis of CLT.

Secondly, in a classroom setting, thinking of teachers using a type of approach entirely is not very realistic. Looking at the results of this thesis also proves that these two instructional approaches are highly used. Especially, teachers from participated countries in TIMSS 2015 reported that they largely use the teaching items of the traditional didactic approach. (See **Table 8**) Therefore, distinguishing when or where exactly they use these teaching items from both instructional approaches is unclear. Accordingly, the potential effects of these approaches become hard to measure explicitly. This conclusion also can be supported by the fact that two instructional approaches showed multicollinearity in many countries analysed in this thesis. Consequently, the studies that investigate the effects of these two instructional approaches in a so-called laboratory setting where they form the instruction and distinctly apply them possibly give more accurate insights on these effects. Thus, the discourses of CLT can be better discussed. In fact, there are studies that have done that. Yet, then, the compatibility of the findings of such studies to actual science classrooms is questionable. Because, this brings arguments around whether these instructional approaches can likewise be applied or not, in consideration to some constraints e.g. money, time, school resources, classroom environments etc. As a result, this leads to a well-known theory and practice gap. That is how well the hypothesis of CLT operates in practice.

The third reason for CLT struggling to explain the findings of this study can be attributed to the features of science education. Science education inherently is suitable for the inquiry-based approach. In the inquiry science teaching students are expected to design and conduct an experiment or to observe a natural phenomenon and then reflect on them. Moreover, they can do all these in collaboration with their fellows. Such an engagement is not that much feasible in any other subject. That's why thinking of a field that is more appropriate for inquiry than a science subject is rather hard. Nevertheless, NRC (1996, 2000) acknowledged that not all science concepts can or should be taught using inquiry. (Barrow, 2006) This put forwards the necessity of using other instructional approaches e.g. as traditional didactic approach as CLT suggests. However, this view might not be sufficient to back up CLT for its argument in response to science education.

9. Conclusion

The argument on the effectiveness of instructional approaches is a long-standing matter. The literature over the past decades indicated that the traditional didactic approach encountered some drawbacks. Together with the rapidly changing world, as we hear about an advance of a new technology coming into our lives, the view, and the settings of education are changing. (B. K. Khalaf, 2018) This resulted in shifts in the implementation of learning models from teacher-centred to student-centred. The inquiry-based approach is prominent among these learning models.

The present study contributed to the analysis of the effectiveness of traditional didactic and inquiry-based approaches in terms of student science achievement. Even though the present study does not conclude that one of two instructional approaches leads to better academic performance in science education, it rejects the assumption that the inquiry-based approach is more than traditional didactic teaching. Additionally, the present study argues the need of questioning the assumption of instructional approaches having strong impact students' learning since, no remarkable relationship was reported between instructional approaches and student achievement, while, indicators like socio-economic status and student confidence, as in the previous research, were found to be effective on student achievement to some extent.

These approaches can yield better outcomes in different contexts. Besides the effects of these approaches on student achievement, their benefits, and disadvantages in different aspects need careful consideration. For instance, researchers suggest that inquiry-based approaches support students to acquire critical thinking skills.(Scott et al., 2018) In this sense, the implementation of inquiry-based approaches to science curricula can be encouraged in order to develop students' critical thinking skills. However, this should not be considered as an opposition to key elements of traditional didactic approaches. Ignoring traditional didactic approach in science teaching contradicts fundamental learning principles of human cognition. Moreover, the controversy studies on the application of the inquiry-based approach and criticisms that it is drawn in this regard cannot be overlooked.

10.Limitations

As this study involved very large data consisting of 12 countries from all over the world, there were some limitations regarding the examination of specific countries. In the present study, the socio-cultural differences among countries are disregarded. Such cultural learning theories that hypothesize the learning can differ among the racial and ethnic groups and the learning needs and styles interrelated with the cultural structure of one group were not taken in the scope of this thesis.

In this thesis, TIMSS 2015 data aggregated from student level teacher level to be able to use teachers' responses as the source for classroom practices and link them with the student data. TIMSS reports notes that teacher samples are not necessarily are representative as the student data is. Toropova et al. (2020), however, noted that teacher data also corresponds well to census data for teachers.

11. References

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., . . . Tuan, H. L. (2004). Inquiry in science education: International perspectives. *Science Education, 88*(3), 397-419. doi:10.1002/sce.10118

Abrams, E., Southerland, S. A., & Evans, C. A. (2007). Inquiry in the classroom: Necessary components of a useful definition. In E. Abrams, S. A. Southerland, & P. Silva (Eds.), *Inquiry in the science classroom: Realities and opportunities.* Greenwich: CT: Information Age Publishing.

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an Inquiry-based Approach known as the Science Writing Heuristic to Traditional Science Teaching Practices: Are there differences? *International Journal of Science Education, 29*(14), 1745-1765. doi:10.1080/09500690601075629

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does Discovery-Based Instruction Enhance Learning? *Journal of Educational Psychology, 103*(1), 1-18. doi:10.1037/a0021017

Anderson, R. D. (2002). Reforming Science Teaching: What Research Says About Inquiry. *Journal of Science Teacher Education, 13*(1), 1-12. doi:10.1023/A:1015171124982

Bandura, A. (1997). *Self-efficacy : the exercise of control.* Basingstoke: Basingstoke : W. H. Freeman.

Barrow, L. H. (2006). A Brief History of Inquiry: From Dewey to Standards. *Journal of Science Teacher Education, 17*(3), 265-278. doi:10.1007/s10972-006-9008-5

Bevins, S., & Price, G. (2016). Reconceptualising inquiry in science education. *International Journal of Science Education, 38*(1), 17-29. doi:10.1080/09500693.2015.1124300

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347-364. doi:10.1007/BF00138871

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education, 94*(4), 577-616. doi:10.1002/sce.20390

Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 498-522.

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity–propensity analysis. *Contemporary Educational Psychology, 32*(4), 599-629. doi:10.1016/j.cedpsych.2006.09.002

Cheema, J. R., & Skultety, L. S. (2017). Self-efficacy and literacy: a paired difference approach to estimation of over-/under-confidence in mathematics- and science-related tasks. *Educational Psychology, 37*(6), 652-665. doi:10.1080/01443410.2015.1127329

Cohen, L., Manion, L., & Morrison, K. (2011). *Validity and reliability*: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (Eighth edition ed.): Abingdon : Routledge.

Cohen, M. (2008). The Effect of Direct Instruction versus Discovery Learning on the Understanding of Science Lessons by Second Grade Students. *NERA Conference Proceedings 2008.*

Colburn, A. (2000a). An inquiry primer. *Science Scope, 23*(6), 42.

Colburn, A. (2000b). What Teacher Educators Need to Know about Inquiry-Based Instruction.

Condon, M., & Wichowsky, A. (2018). Developing Citizen-Scientists: Effects of an Inquiry-Based Science Curriculum on STEM and Civic Engagement. *The Elementary School Journal, 119*(2), 196-222. doi:10.1086/700316

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97-140. doi:10.1080/10888691.2018.1537791

Dean Jr, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384-397. doi:10.1002/sce.20194

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective. *Performance Improvement Quarterly, 26*(2), 43-71. doi:10.1002/piq.21143

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research, 82*(3), 300-329. doi:10.3102/0034654312457206

Gao, S. (2014). Relationship between Science Teaching Practices and Students' Achievement in Singapore, Chinese Taipei, and the US: An Analysis Using TIMSS 2011 Data. *Frontiers of Education in China, 9*(4), 519-551. doi:10.3868/s110-003-014-0043-x

Goh, W., Kwek, D., Hogan, D., & Cheong, S. (2014). Complex network analysis of teaching practices. *EPJ Data Science, 3*(1), 1-16. doi:10.1140/epjds/s13688-014-0034-9

Hattie, J. (2009). *Visible learning : a synthesis of over 800 meta-analyses relating to achievement*. London: London : Routledge.

Heaysman, O., & Tubin, D. (2019). Content teaching: innovative and traditional practices. *Educational Studies, 45*(3), 342-356. doi:10.1080/03055698.2018.1446334

Hinton, P. R. (2014). *SPSS explained* (2nd edition. ed.). Hove: Hove : Routledge.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). The Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

House, J. D., & Telese, J. A. (2017). Confidence in science and achievement outcomes of fourth-grade students in Korea: results from the times 2011 assessment. *Education, 137*(4), 389.

Hox, J. (2003). Multilevel Analysis: Techniques and Applications. *http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=019198/(100)*. doi:10.4324/9780203852279

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Johansone, I. (2016). Survey Operations Procedures in TIMSS 2015. In *Methods and Procedures in TIMSS 2015 (pp. 6.1-6.22)* (M. O. Martin, I. V. S. Mullis & M. Hooper ed.). Retrieved from http://timss.bc.edu/publications/timss/2015-methods/chapter-6.html

Kaymakamoglu, S. E. (2018). Teachers' Beliefs, Perceived Practice and Actual Classroom Practice in Relation to Traditional (Teacher-Centered) and Constructivist (Learner-Centered) Teaching (Note 1). *Journal of Education and Learning, 7*(Web), 29-37.

Khalaf, B. (2018). Traditional and Inquiry-Based Learning Pedagogy: A Systematic Critical Review. *International Journal of Instruction, 11*, 545-564. doi:10.12973/iji.2018.11434a

Khalaf, B. K. (2018). Traditional and Inquiry-Based Learning Pedagogy: A Systematic Critical Review. *International Journal of Instruction, 11*, 545-564. doi:10.12973/iji.2018.11434a

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The Equivalence of Learning Paths in Early Science Instruction: Effects of Direct Instruction and Discovery Learning. *Psychological Science, 15*(10), 661-667. doi:10.1111/j.0956-7976.2004.00737.x

Krahenbuhl, K. S. (2016). Student-centered Education and Constructivism: Challenges, Concerns, and Clarity for Teachers. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 89*(3), 97-105. doi:10.1080/00098655.2016.1191311

Kubiatko, M. (2016). Is The Inquiry-Based Science Education The Best? *Problems of Education in the 21st Century, 73*, 4-5.

Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Review of Educational Research, 86*(3), 681-718. doi:10.3102/0034654315627366

Lederman, J. S., Lederman, N. G., & Wickman, P.-O. . (2008). *An international, systematic investigation of the relative effects of inquiry and direct instruction: A replication study.* . Paper presented at the The annual conference of the National Association for Research in Science Teaching, Baltimore, MD.

Lederman, N., Lederman, J., Wickman, P. -O., & Lager-Nyqvist, L. (2007). *Development of a valid and reliable protocol for the assessment of early childhood students' conceptions of nature of science and scientific inquiry.* Paper presented at the the National Association for Research in Science Teaching, New Orleans, LA.

Ma, X. (2000). Socioeconomic Gaps in Academic Achievement Within Schools: Are They Consistent Across Subject Areas? *Educational Research and Evaluation, 6*(4), 337-355. doi:10.1076/edre.6.4.337.6935

Magliaro, S., Lockee, B., & Burton, J. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research and Development, 53*(4), 41-55. doi:10.1007/BF02504684

Matsunaga, M. (2008). Item Parceling in Structural Equation Modeling: A Primer. *Communication Methods and Measures, 2*(4), 260-293. doi:10.1080/19312450802458935

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist, 59*(1), 14-19. doi:10.1037/0003-066X.59.1.14

McMullen, F., & Madelaine, A. (2014). Why is there so much resistance to Direct Instruction? *Australian Journal of Learning Difficulties, 19*(2), 137-151. doi:10.1080/19404158.2014.962065

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474-496. doi:10.1002/tea.20347

Mohammadpour, E., Shekarchizadeh, A., & Kalantarrashidi, S. A. (2015). Multilevel Modeling of Science Achievement in the TIMSS Participating Countries. *The Journal of Educational Research, 108*(6), 449-464. doi:10.1080/00220671.2014.917254

Mullis, I. V. S., Cotter, K. E., Fishbein, B. G., & Centurino, V. A. S. (2016). Developing the TIMSS 2015 achievement items. In *Methods and Procedures in TIMSS 2015* (M. O. Martin, I. V. S. Mullis, & M. Hooper ed., pp. (pp.1.1-1.22)).

Muthén, L., & Muthén, B. (2018). Mplus users guide and Mplus version 8.2. Retrieved January 10, 2019. In.

Myint Swe, K. (2013). *Application of Structural Equation Modeling in Educational Research and Practice*: Germany: Springer Verlag.

Neuschmidt, O. (2018). *Using TIMSS 2015 data to compare educational effectiveness-enhancing factors in the countries of the Gulf Cooperation Council* (PhD.). University of Hamburg,

Nicholson, L., Putwain, D., Connors, L., & Hornby-Atkinson, P. (2013). The key to successful achievement as an undergraduate student: confidence and realistic expectations? *Studies in Higher Education, 38*(2), 285-298. doi:10.1080/03075079.2011.585710

Rowe, K. (2006). Effective teaching practices for students with and without learning difficulties: Issues and implications surrounding key findings and recommendations from the national inquiry into the teaching of literacy. *Australian Journal of Learning Disabilities, 11*(3), 99-115. doi:10.1080/19404150609546813

Rundgren, C.-J. (2018). Implementation of inquiry-based science education in different countries: some reflections. *Cultural Studies of Science Education, 13*(2), 607-615. doi:10.1007/s11422-016-9787-8

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*(6), 323-338. doi:10.3200/JOER.99.6.323-338

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the

United States. *Journal of Research in Science Teaching, 44*(10), 1436-1460. doi:10.1002/tea.20212
Schwab, J. J., & Brandwein, P. F. (1962). The Teaching of science: The teaching of science as enquiry. *Cambridge: Harvard University Press.*
Scott, D., Smith, C., Chu, M.-W., & Friesen, S. (2018). Examining the Efficacy of Inquiry-based Approaches to Education. *Alberta Journal of Educational Research, 64*.
Shabani, K., Khatib, M., & Ebadi, S. (2010). Vygotsky's Zone of Proximal Development: Instructional Implications and Teachers' Professional Development. *English language teaching, 3*(4), 237-248.
Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. In (Vol. 75, pp. 417-453). Thousand Oaks, CA.
Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. *Review of Educational Research, 88*(4), 479-507. doi:10.3102/0034654317751919
Strand, S. (2014). *Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it' for the white working class* (Vol. 29).
Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information systems, 13*(1), 24.
Streiner, D. L. (2006). Building a Better Model: An Introduction to Structural Equation Modelling. *The Canadian Journal of Psychiatry, 51*(5), 317-324. doi:10.1177/070674370605100507
Suchman, J. R. (1966). CHAPTER 11 - A MODEL FOR THE ANALYSIS OF INQUIRY. In H. J. Klausmeier & C. W. Harris (Eds.), *Analyses of Concept Learning* (pp. 177-187): Academic Press.
Sweller, J. (2008). Human Cognitive Architecture. In *Handbook of research on educational communications and technology* (Vol. 35, pp. 369-381). Oxford, UK: British Journal of Educational Technology.
Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-296. doi:10.1023/A:1022193728205
Sykes, L. M., Gani, F., & Vally, Z. (2016). Statistical terms Part 1: The meaning of the MEAN, and other statistical terms commonly used in medical research. *South African Dental Journal, 71*, 274-278. Retrieved from http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0011-85162016000600009&nrm=iso
Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management, 5*, 28-36. doi:10.2139/ssrn.3205040
Thomson, S., Wernert, N., O'Grady, E., & Rodrigues, S. (2016). *TIMSS 2015 : A first look at Australia's results.* Retrieved from Melbourne:
Toropova, A., Myrberg, E., & Johansson, S. (2020). Teacher job satisfaction: the importance of school working conditions and teacher characteristics. *Educational Review*, 1-27. doi:10.1080/00131911.2019.1705247
Valcke, M. (2002). Cognitive load: updating the theory? *Learning and Instruction, 12*(1), 147-154. doi:https://doi.org/10.1016/S0959-4752(01)00022-6
Vogel-Walcutt, J., Gebrim, J., Bowers, C., Carper, T. M., & Nicholson, D. (2011). Cognitive load theory vs. constructivist approaches: which best leads to efficient, deep learning? *Journal of Computer Assisted Learning, 27*(2), 133-145. doi:10.1111/j.1365-2729.2010.00381.x
Von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). *What are plausible values and why are they useful* (Vol. 2).
Weston, R., & Gore, P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist, 34*(5), 719-751. doi:10.1177/0011000006286345
White, K. (1982). *The Relation Between Socioeconomic Status and Academic Achievement* (Vol. 91).

Yamamoto, K., & Kulick, E. (2000). Scaling Methodology and Procedures for the TIMSS Mathematics and Science Scales. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 237-263). Chestnut Hill: MA: Boston College.
Yilmaz, K. (2011). The Cognitive Perspective on Learning: Its Theoretical Underpinnings and Implications for Classroom Practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 84*(5), 204-212. doi:10.1080/00098655.2011.568989

12.Appendices

Appendix 1 TIMSS 2015 Science Achievement Distribution

Country	**Average Scale Score**
1. Singapore	597 (3.2)
2. Japan	571 (1.8)
3. Chinese Taipei	569 (2.1)
4. Korea, Rep. of	556 (2.2)
5. Slovenia	551 (2.4)
6. Hong Kong SAR	546 (3.9)
7. Russian Federation	544 (4.2)
8. England	537 (3.8)
9. Kazakhstan	533 (4.4)
10. Ireland	530 (2.8)
11. United States	530 (2.8)
12. Hungary	527 (3.4)
13. Canada	526 (2.2)
14. Sweden	522 (3.4)
15. Lithuania	519 (2.8)
16. New Zealand	513 (3.1)
17. Australia	512 (2.7)
18. Norway (9)	509 (2.8)
19. Israel	507 (3.9)
TIMSS Scale CenterPoint	500
20. Italy	499 (2.4)
21. Turkey	493 (4.0)
22. Malta	481 (1.6)
23. United Arab Emirates	477 (2.3)
24. Malaysia	471 (4.1)
25. Bahrain	466 (2.2)
26. Qatar	457 (3.0)
27. Iran, Islamic Rep. of	456 (4.0)
28. Thailand	456 (4.2)
29. Oman	455 (2.7)
30. Chile	454 (3.1)
31. Georgia	443 (3.1)
32. Jordan	426 (3.4)
33. Kuwait	411 (5.2)
34. Lebanon	398 (5.3)
35. Saudi Arabia	396 (4.5)
36. Morocco	393 (2.5)
37. Botswana (9)	392 (2.7)
38. Egypt	371 (4.3)
39. South Africa (9)	358 (5.6)
Benchmarking Participants	
Quebec, Canada	530 (4.4)

Dubai, UAE	525 (2.0)
Ontario, Canada	524 (2.5)
Florida, US	508 (6.0)
Norway (8)	489 (2.4)
Abu Dhabi, UAE	454 (5.6)
Buenos Aires, Argentina	386 (4.2)

13. *( ) Standard errors appear in parentheses.*

Appendix 2 Student Confident in science scale

How much do you agree with these statements about science?

	Agree a lot	Agree a little	Disagree a little	Disagree a lot
1) I usually do well in science	○	○	○	○
2) Science is harder for me than for many of my classmates*	○	○	○	○
3) I am just not good at science*	○	○	○	○
4) I learn things quickly in science	○	○	○	○
5) My teacher tells me I am good at science	○	○	○	○
6) Science is harder for me than any other subject*	○	○	○	○
7) Science makes me confused*	○	○	○	○

* Reverse coded

Very Confident in Science | 10.2 | Confident in Science | 8.2 | Not Confident in Science

Appendix 3 Number of books at home variable in TIMSS 2015

TIMSS 2015 Science 4th Grade

Exhibit 4.1: Home Resources for Learning

Reported by Parents, except Number of Books and Home Study Supports Reported by Students

Students were scored according to their own and their parents' responses concerning the availability of five resources on the *Home Resources for Learning* scale. Students with **Many Resources** had a score of at least 11.9, which is the point on the scale corresponding to students reporting they had more than 100 books in the home and both of the home study supports, and parents reporting that they had more than 25 children's books in the home, that at least one parent had finished university, and that at least one parent had a professional occupation, on average. Students with **Few Resources** had a score no higher than 7.4, which is the scale point corresponding to students reporting that they had 25 or fewer books in the home and neither of the home study supports, and parents reporting that they had 10 or fewer children's books in the home, that neither parent had gone beyond upper-secondary education, and that neither parent was a small business owner or had a clerical or professional occupation, on average. All other students were assigned to the **Some Resources** category.

Number of books in the home (students):
1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

Number of children's books in the home (parents):
1) 0-10
2) 11-25
3) 26-50
4) 51-100
5) More than 100

Number of home study supports (students):
1) None
2) Internet connection or own room
3) Both

Highest level of education of either parent (parents):
1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher

Highest level of occupation of either parent (parents):
1) Has never worked outside home for pay, general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator)
2) Clerical (clerk or service or sales worker)
3) Small business owner
4) Professional (corporate manager or senior official, professional, or technician or associate professional)

Many Resources | 11.9 | Some Resources | 7.4 | Few Resources

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

IEA TIMSS & PIRLS International Study Center Lynch School of Education, Boston College

Appendix 4 The bivariate correlations of traditional didactic teaching practices

Correlations

		Listen teacher to explain new content	Watch teacher to demonstrate an experiment	Read their textbooks or other resource materials	Have students memorize facts and principles	Use scientific formulas and laws to solve routine problems	Take a written test or quiz
Listen teacher to explain new content	Pearson Correlation	1	.220**	.188**	.278**	.156**	.103**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	4566	4553	4557	4551	4547	4553
Watch teacher to demonstrate an experiment	Pearson Correlation	.220**	1	.129**	.231**	.346**	.202**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	4553	4565	4556	4550	4547	4553
Read their textbooks or other resource materials	Pearson Correlation	.188**	.129**	1	.467**	.206**	.180**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	4557	4556	4573	4562	4558	4565
Have students memorize facts and principles	Pearson Correlation	.278**	.231**	.467**	1	.408**	.221**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	4551	4550	4562	4566	4553	4560
Use scientific formulas and laws to solve routine problems	Pearson Correlation	.156**	.346**	.206**	.408**	1	.220**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	4547	4547	4558	4553	4562	4557
Take a written test or quiz	Pearson Correlation	.103**	.202**	.180**	.221**	.220**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	4553	4553	4565	4560	4557	4569

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 5 The bivariate correlations of traditional didactic teaching practices

Correlations

		Relating to students` daily lives	Ask students to decide their own prosol procedures	Observe a phenomena and describe	Conduct Experiments or Investigations	Interpret data from experiments or investigations	Use evidence from experiments or investigations to support conclusions	Do field work outside of class
Relating to students` daily lives	Pearson Correlation	1	.247**	.234**	.085**	.125**	.138**	.152**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	4582	4577	4544	4547	4544	4544	4540
Ask students to decide their own prosol procedures	Pearson Correlation	.247**	1	.284**	.190**	.255**	.225**	.261**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	4577	4582	4545	4548	4545	4545	4540
Observe a phenomena and describe	Pearson Correlation	.234**	.284**	1	.319**	.359**	.343**	.269**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	4544	4545	4567	4559	4557	4554	4548
Conduct Experiments or Investigations	Pearson Correlation	.085**	.190**	.319**	1	.663**	.588**	.225**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	4547	4548	4559	4569	4561	4558	4552
Interpret data from experiments or investigations	Pearson Correlation	.125**	.255**	.359**	.663**	1	.711**	.251**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	4544	4545	4557	4561	4567	4558	4550
Use evidence from experiments or investigations to support conclusions	Pearson Correlation	.138**	.225**	.343**	.588**	.711**	1	.240**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	4544	4545	4554	4558	4558	4567	4554
Do field work outside of class	Pearson Correlation	.152**	.261**	.269**	.225**	.251**	.240**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	4540	4540	4548	4552	4550	4554	4563

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 6 Descriptive Statistics and Reliability coefficient After Parcelling- Cronbach's Alpha

	Valid Cases	**TRA After Parcelling**	**INQ After Parcelling**
Chile	194	0.597	0.733
Egypt	213	0.516	0.695
England	564	0.546	0.526
Italy	228	0.663	0.641
Japan	171	0.634	0.706
Lithuania	910	0.538	0.707
New Zealand	333	0.519	0.631
Norway	211	0.531	0.645
Russia	761	0.561	0.627
Singapore	320	0.630	0.676
Slovenia	572	0.541	0.665
South Africa	319	0.614	0.749

Appendix 7 MODEL1 Fit Indices for 12 countries

	χ2 (df)	**CFI**	**RMSEA**	**SRMR**
Good Fit (Acceptable Threshold)		≥.95 (≥.90)	≤ 0.08 (≤ 0.10)	≤ 0.06 (≤ 0.10)
Chile	34.088 (12)	0.916	0.097	0.054
Egypt	17.625 (12)	0.967	0.047	0.040
England	62.307 (12)	0.884	0.073	0.042
Italy	26.389 (12)	0.936	0.073	0.049
Japan	18.978 (12)	0.963	0.058	0.054
Lithuania	29.321 (12)	0.982	0.040	0.023
New Zealand	31.179(12)	0.903	0.069	0.045
Norway	34.851 (12)	0.848	0.095	0.071
Russia	16.484 (12)	0.992	0.022	0.020
Singapore	49.330 (12)	0.896	0.099	0.054
Slovenia	27.982 (12)	0.965	0.048	0.026
South Africa	32.860 (12)	0.958	0.074	0.034

*RMSEA=Root Mean Square Error of Approximation
*CFI=Comparative fit index
*SRMR=Standardized Root Mean Square Residual
*χ2 – Chi-square mean, (Df)= degrees of freedom

Appendix 8

```
TITLE:        Multilevel SEM_ MODEL2
DATA:         FILE IS  Parcelled_Teacher_Level_ENG.dat;
VARIABLE: NAMES ARE IDCNTRY
            SciAch PV2 PV3 PV4 P5 SES Ed_lvl_mom
            Ed_lvl_dad CON CON2
            INQ1 INQ2 INQ3
            TRA1 TRA2 TRA3;
USEVARIABLES ARE
           SciAch SES CON
           INQ1 INQ2 INQ3
           TRA1 TRA2 TRA3;
  Missing are all (99);
ANALYSIS: type = general;
MODEL: INQ by INQ1 INQ2 INQ3;
          TRA by TRA1 TRA2 TRA3;

             SciAch ON INQ TRA SES CON;
             CON on SES;
OUTPUT: standardized modindices;
           sampstat;
           stdyx;
```