The Impact of Explore-Instruct Teaching Approach on University Students' Scientific Reasoning at The United Arab Emirates.

أثر التدريس بطريقة الاستكشاف - ثم الإرشاد على المنطق العلمي لطلاب جامعيين في دولة الامارات

by

FATIMA AHMED ABAZAR

A thesis submitted in fulfilment

of the requirements for the degree of

DOCTOR OF EDUCATION

at

The British University in Dubai

November 2021

The Impact of Explore-Instruct Teaching Approach on University Students' Scientific Reasoning at The United Arab Emirates.

أثر التدريس بطريقة الاستكشاف - ثم الإرشاد على المنطق العلمي لطلاب جامعيين في دولة الامارات.

by

FATIMA AHMED ABAZAR

A thesis submitted to the Faculty of Education.

In fulfilment of the requirements for the degree of

DOCTOR OF EDUCATION

at

The British University in Dubai

November 2021

Thesis Supervisor

Professor Sufian A. Forawi

Approved for award:

_____________________________ _________________________

Name Name

Designation Designation

_____________________________ _________________________

Name Name

Designation Designation

Date: ____________

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

____

FATIMA AHMED ABAZAR

COPYRIGHT AND INFORMATION TO USER

ABSTRACT

*Background :*Minds of people living in the present and future centuries should differ from those of the ones lived before. It is therefore the concern of societies looking for glory, to enhance their people cognitive abilities of problem solving and critical thinking to face contemporary challenges. In the scientific domain, Scientific Reasoning (SR) is one of the critical thinking dimensions in problem solving. Therefore, in order for societies, to enhance the quality of their human resources, there is an urge to raise the scientific reasoning abilities of their personnel. A wise choice of a tool to raise SR would be the formal education. Therefore, the pressing question would be whether exist any teaching strategy would assist in reaching the former goal. Features of such teaching strategy may be figured out by studying pedagogies and cognition. Explore-instruct teaching strategy has been chosen here to explore its impact in developing scientific reasoning ability and compare it to the Instruct-Solve traditional strategy.

Purpose: The purpose of this research was to study the impact of explore-instruct teaching approach on college students' scientific reasoning development in the United Arab Emirates, and to explore the perspectives of participating students who experienced this approach.

Methods: The study took place in a private university at the United Arab Emirates (UAE). The participants were students studying General Physics (I). The study compared the development in SR in the control group which receives traditional physics lectures via "Instruct-solve" and the experimental group that received the same lectures via "explore-instruct" teaching approach. Scores of students in Lawson Classroom Scientific Reasoning Test were compared for the two groups prior and post the course in a four-month semester. The perspectives of students experienced the explore-instruct approach were revealed through e-mails interviews which included open-ended questions and one last close-ended question.

Results: The study quantitatively proved that both teaching strategies had no significant effect on the Proportional Reasoning (PPR) of the participants, but they resulted in a normalized gain in the Control of Variables (CoV) and Correlational Reasoning (CRR) dimensions of scientific reasoning. But Explore-Instruct teaching approach made a higher normalized gain in CoV than the traditional approach of Instruct-Solve. Qualitative data collected via email interviews to study perspectives of students; revealed preference for the Explore-instruct teaching approach by some students when there is no work to be submitted for assessment. But students, in general, preferred the traditional approach when lab reports or any post-session work to be completed and submitted for assessment.

Implications/Contributions: Higher education in general, and scientific and technological education, have a powerful role in qualifying personnel leading to developing economies. It has been therefore widely argued that the complexity of challenges encountered by the 21st society increased with the advance in technology, the education system is required to focus on improving students' problem-solving cognitive skills. Scientific reasoning is one of the essential cognitive skills whose deficiency often leads to judgment errors when facing an issue. Studying the impact of the Explore-Instruct teaching approach on scientific reasoning would provide teachers, curriculum developers, educational quality assurance personnel, and decision-makers with evidence of improvement of CoV in a period of 4 months only, via simple flip in order of periods of the problem solving or conducting an experiment by the students and the instructor's period of lecturing. This will accordingly assist educators in making their choices.

***Keywords*:** Explore-Instruct, Scientific Reasoning, Control of Variables, Teaching approach.

الخلاصة

خلفية الدراسة: يجب أن تختلف عقول الأشخاص الذين يعيشون في القرون الحالية والمستقبلية عن عقول الذين عاشوا من قبل. لذلك فإن من هم المجتمعات التي تبحث عن المجد ، تعزيز القدرات المعرفية لشعوبها في حل المشكلات والتفكير النقدي لمواجهة التحديات المعاصرة. في المجال العلمي ، يعتبر التفكير العلمي (SR) أحد أبعاد التفكير الناقد في حل المشكلات. لذلك ، لكي تعزز المجتمعات جودة مواردها البشرية ، هناك حاجة إلى رفع قدرات التفكير العلمي لموظفيها. الاختيار الحكيم لأداة رفع مستوى التفكير العلمي و بالتالي الأداء الاجتماعي سيكون هو التعليم الرسمي. لذلك ، فإن السؤال الملح هو ما إذا كانت أي استراتيجية تعليمية موجودة من شأنها أن تساعد في الوصول إلى الهدف السابق. يمكن التعرف على ميزات استراتيجية التدريس هذه من خلال دراسة طرق التدريس والإدراك. تم هنا اختيار إستراتيجية التدريس بالاستكشاف -ثم التدريس لاستكشاف تأثيرها في تطوير القدرة على التفكير العلمي ومقارنتها باستراتيجية التدريس-ثم التطبيقالتقليدية.

غرض الدراسة: كان الغرض من هذا البحث هو دراسة تأثير نهج التدريس الاستكشافي على تنمية التفكير العلمي لطلاب الجامعات في الإمارات العربية المتحدة ، واستكشاف وجهات نظر الطلاب المشاركين الذين جربوا هذا النهج.

منهج الدراسة:. تمت الدراسة في جامعة خاصة في دولة الإمارات العربية المتحدة. كان المشاركون طلابا يدرسون الفيزياء العامة (1). قارنت الدراسة التطور في الاستدلال العلمي في المجموعة الضابطة التي تتلقى محاضرات فيزياء تقليدية من خلال طريقة التدريس بالإرشاد-ثم التطبيق والمجموعة التجريبية التي تلقت نفس المحاضرات من خلال طريقة التدريس بالاستكشاف-ثم الإرشاد. تمت مقارنة عشرات الطلاب في اختبار لاوسون الصفي للاستدلال العلمي للمجموعتين قبل الت وبعدها في فصل دراسي مدته أربعة أشهر. تم الكشف عن وجهات نظر الطلاب الذين اختبروا منهج الاستكشاف والتوجيه من خلال المقابلات عبر البريد الإلكتروني التي تضمنت أسئلة مفتوحة وسؤال أخير مغلق.

نتائج الدراسة: أثبتت الدراسة كميًا أن كلا استراتيجيتي التدريس لم يكن لهما تأثير كبير على التفكير التناسبي (PPR) للمشاركين ، لكنهما أسفرتا عن مكاسب طبيعية في أبعاد التحكم في المتغيرات (CoV) والاستدلال الترابطي (CRR) للتفكير العلمي. لكن نهج التدريس طريقة التدريس بالاستكشاف- ثم الإرشاد حقق مكاسب طبيعية أعلى في بُعد التحكم في المتغيرات(CoV)مقارنة بالنهج التقليدي وهو الإرشاد- ثم التطبيق. البيانات النوعية التي تم جمعها عبر مقابلات البريد الإلكتروني لدراسة وجهات نظر الطلاب ؛ كشف عن تفضيل بعض الطلاب لنهج التدريس بالاستكشاف- ثم الإرشاد عندما لا يكون هناك عمل لتقديمه للتقييم. لكن الطلاب ، بشكل عام ، فضلوا النهج التقليدي عند ضرورةاستكمال التقارير المعملية أو أي عمل بعد الصف وإرساله للتقييم.

استنتاجات ومساهمات الدراسة: يلعب التعليم العالي بشكل عام والتعليم العلمي والتكنولوجي دورًا قويًا في تأهيل الكوادر المؤدية إلى الاقتصادات النامية. لذلك فقد قيل على نطاق واسع أن تعقيد التحديات التي يواجهها المجتمع الحادي والعشرون زاد مع التقدم في التكنولوجيا ، والنظام التعليمي مطلوب للتركيز على تحسين المهارات المعرفية لدى الطلاب في حل المشكلات. التفكير العلمي هو أحد المهارات المعرفية الأساسية التي يؤدي نقصها غالبًا إلى أخطاء في الحكم عند مواجهة مشكلة ما. إن دراسة تأثير نهج التدريس بالاستكشاف- ثم الإرشاد على التفكير العلمي من شأنه أن يزود المعلمين، ومطوري المناهج، وموظفي ضمان الجودة التعليمية، وصناع القرار بدليل على تحسين بُعد التحكم في المتغيرات (CoV) في فترة 4 أشهر فقط ، عن طريق قلب بسيط بترتيب الفترات، حل المشكلة أو إجراء تجربة من قبل الطلاب وفترة المحاضرة للأستاذ. سيساعد هذا ، وفقًا لذلك، التربويين في اتخاذ قراراتهم.

الكلمات المفتاحية: الاستكشاف-ثم االإرشاد ، التفكير العلمي، الإرشاد – ثم التطبيق، طريقة التدريس.

DEDICATION

I dedicate this thesis to my father (Ahmed Abazar), my mother (Seddiga Awad Al-Ballal), my sisters (Tasneem, Rabia and Abrar), my brother (Awab), my husband (Ayman Ibrahim El Sheikh), my daughters (Dan, Rawan and Rudainah), my son Mohammed , my nephews Eyad and Hassan Ahmed Mukhtar, and my niece Aleen Osama Tayfour, as an appreciation to their constant love and support.

ACKNOWLEDGMENTS

First, I am extremely thankful to Allah for giving me the patience and the strength all the way through my postgraduate journey. Furthermore, I wish to express my appreciation to my Director of Studies, Prof. Sufian Forawi for his great support and encouragement. His advice on both research and education career have been invaluable. It was a great privilege to get the chance of contributing with a chapter called “Critical and Creative thinking in Science Education” in his edited book “Science and Mathematics Education in Multicultural Contexts: *New Direction in Teaching and Learning*”, in addition to finishing this piece of work under his supervision.

I would like to acknowledge the remarkable support from my parents, sisters, brother, my husband, and my children. The trust you put on me, have empowered me throughout this long difficult course of my life, I am so grateful for the amazing care and support. I owe you an apology for the sacrifices you made to help me complete this work successfully.

I also acknowledge the support of my students and the assistance of my colleagues; Dr. Ahmed Al Rahil (statistician), Dr. Makhtar Sarr (statistician), Dr. Haitham Farouk (Physicist), Dr. Abdullah El-Mahi (English instructor), Ms. Muna Hamad (English instructor). Prof. Ali El-tigani Al-Mahi (Archaeologist)

To my family and friends, your inspiration and motivational words have carried me throughout this process.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Categ. Post/ Categ.SR-PRE	Category of SRA post to the intervention
Categ. Pre/ Categ.SR-PRE	Category of SRA prior to the intervention
CMR	Combinational Reasoning
COV	Control of Variables.
CRR	Correlational Reasoning
CVS	Control of Variables Strategy
ED	Experimenting and Discussion.
EI	Explore-Instruct teaching approach.
IBL	Inquiry-Based Learning
LCTSR	Lawson Classroom Test of Scientific Reasoning
N-Gain	Normalized Gain.
PPR	Proportional Reasoning
Pre-test%	Pre-instruction LCTSR percentage score
RPQ	Read Present and Question
SRA	Scientific Reasoning Ability
SSI	Socio-Scientific Issue.
STEM	Science Technology and Math
TM	Traditional Method

CHAPTER 1

INTRODUCTION

It is generally acknowledged that there is a correlation between the quality of higher education of any nation and its economic development. It is also admitted that science and technology are essential for enhancing the quality of life and raising economic chances (Oluwafemi & Laseinde, 2020, Krumm and Kimmie, 2013). Therefore, higher education in general, and particularly scientific and technological education, have a powerful role in developing economy. One of the main indicators of the quality of higher education and science graduates, and a cognitive skill that has been declared by several researchers an essential cognitive skill supporting the Critical Thinking (CT) and scientific literacy, is the Scientific Reasoning (SR) skill (Faust and Meehl, (1984); Pyper (2012); Lawson &Worsnop (1992); She & Liao (2010); Morris et al., (2012), etc.). Scientific Reasoning (SR) is the way that the person uses an evidence to come at a conclusion and is the reasoning we use when taking contentions seriously in science, whether it be physical sciences (like physics and biology) or social sciences (like psychology and economics) (Lack and Rousseau, 2016). Scientific reasoning is generally defined from the research point of view (Zimmerman, 2005) as the term "includes the thinking and reasoning skills involved in inquiry, experimentation, evidence evaluation, inference, and argumentation that support the formation and modification of concepts and theories about the natural and social world" (Bao, 2009, p.1).Scientific reasoning was also defined from the science literacy point of view (Brown et al. 2010,) as the term represents the cognitive skills essential to understand and evaluate scientific information, that usually involve understanding and evaluating causal, statistical and theoretical hypotheses (Bao, 2009, p.1).

In practicing scientific reasoning, Lack and Rousseau (2016) shed a light on an essential cognitive skill that facilitates scientific reasoning. It is the capability of changing minds about particular matter and realizing that we are maintaining a false belief, and to modify. It is found to be a precious essential human skill. Incorrect ideas can confuse debates, it can also lead to worse choices than we might make in any other case. In general, it limits our possibilities to make the maximum of our reasoning situations and decision making. Yet, the ability to modify our believes, is a skill that needs exercising. Usually, it is taken for granted and in a way that we forget about exercising and improving it. If we allow defective assumptions or unsound reasoning to grow to be rooted in our minds, we may then be worse at changing our believes flexibly with time. This is, in brief, why do we need the skills of Scientific reasoning and critical thinking and why do we need to exercise those skills during our life (Lack and Rousseau, 2016). Consequently, appear here an urgent need for a whole education system at which the students practice scientific reasoning supported by conceptual change. Such system will raise this skill and at the same time improve performance in science field.

Though, despite eminent introduction of scientific fields in the higher education (HE) sector, the undergraduate's failure rates at these institutions remain high(Kuhn & Modrek, 2017) and improvement of reasoning skills remain unsatisfactory (Ding, Wei & Liu 2016).In a daily life pattern something that is observed and may be related to this problem is that people discussion so often seem superficial, having people arguing past each other rather than with one another, Kuhn and Modrek (2017) think that causes may be not only dialogical but also individual and logical. They considered in their study whether there exist mistakes in reasoning that would be specifically of unfavourable consequences on argumentative debates. specifically, they considered implications of two mistakes: explanation as a substitute for evidence and neglecting the possibility of multiple causes responsible of final outcomes. Kuhn and Modrek (2017) displayed in their research those errors to be ordinary to occur in a cross section of adults, and also in samples of community college undergraduates and also in young children, with minimal age-related development. in addition, they occur, although less frequently, among a sample of exceptionally educated adults, and Kuhn and Modrek (2017) finally demonstrated the effect of these mistakes in discussions and debates of college-educated adults. Khun and Modrek (2017) point ultimately to proof that those individual reasoning mis-practice is probably addressable via education.

Reasoning is a cognitive ability connected to cognition, and cognition routing has been discussed by the artificial intelligence researcher Qadir, (2016) in the cognitive radio networks (CRNs). Cognitive radio networks are nodes supplied with cognitive radios (CRs) which are programmed to “sense, learn, and react to changes in network conditions” (Qadir, 2016). The idea of adapting to changes was supported by incorporating cognitive networking with artificial intelligence —that entails understanding and reasoning the radio environment —to make optimal decisions autonomously. The term Dzobo (2020) stated that "artificial intelligence" goes back to the last twentieth century, i.e., more specifically to the year 1956, when it was used for the first time at a conference held at Dartmouth College in the United States of America. Since that historical era, the gradual and partial development of artificial intelligence, especially in the field of neuroscience, began through neuroscientists who attempted to study and comprehend the human brain, when those scientists sought to create intelligent machines that have the ability to perform complex tasks such as those performed by the human race(Dzobo et al., 2020).These machines have a wide range of applications such as in public safety systems, intelligent transport systems, cooperative networks, smart grid communications, femtocells, dynamic spectrum access.

Cognitive networking usually includes models of cognition and learning that have been set for cognitive networking’s. Such cognitive networks will observe and analyse its surrounding and set up and judge to fulfil the policy goals. Also, it is to constraints and enact on the determined policy. Significantly, such networks learn from past situations and from the surrounding environment to enhance performance with time (Qadir, 2016). This should motivate the education systems to enhance the scientific reasoning of students to deal with and be able to develop artificial intelligence (AI) based techniques into the design of future machines.

In this regard, the concepts of “artificial intelligence” and “wireless communication networking” have been merged with the aim of operating cognitive radio networks through a multiple integration of engineering disciplines. The use of such artificial intelligence concepts represents the cornerstone of cognitive radio networks and thus distinguishes them from other wireless networks in terms of their superior ability to build, learn and adopt knowledge bases. The process of integrating these concepts accurately and tightly is made through building and developing a unified knowledge-based network that has a great ability to identify the environment in which it operates, and through an effective artificial mind that possesses the superior ability in the decision-making process, knowledge inference, performance development, monitoring, adaptation to the surrounding conditions and the ability to deal with future projections(Gavrilovska et al., 2013).

In the field of artificial intelligence, the two key attributes and characteristics of perception and reasoning are actually realized, noting that these two characteristics are essential in the context of human intelligence, as people use them to solve the problems they face in various walks of life. Perception and reasoning capabilities are usually achieved through the work of logic programming and machine learning, and the two features have already been developed separately throughout the history of the development of artificial intelligence sciences. Logic programming is based on exploiting familiarity with the symbolic domain and then correcting incorrectly perceived facts in order to improve and develop machine learning, while the machine learning program benefits from the awareness of initial logical facts through the data available to it. Thus, machine learning has become something possible, as machines can recognize numbers and solve various mathematical and accounting problems and equations with a faster and more accurate capacity and adapt them to the modern logical thinking model, which may bypass deep learning models(Dai et al., 2019).

Thinking things through in a quick reflective critical way became more essential for individuals and urge making efforts in the cognitive domain He et al. (2010). However, what is needed as per (Krumm & Kimmie, 2013) is transforming the learner's conscious. It is assumed that it will enable these learners to completely own the learning process via inquiring, reflective and critical thinking.

Likewise, jumping to conclusions, cognition includes both planning and learning. With planning being the procedure of preparing the suitable activity at specific circumstances to optimize performance. As for learning, it is the way of collecting experience from effects left by the past activities. He et al. (2010). Recent studies also found a strong correlation between scientific creativity and scientific reasoning (Sternberg et al., 2020)

A critically well-versed teaching approach is claimed by Krumm and Kimmie (2013) to provide the perfect opportunity to reduce with time many calamities, and support reasoning to achieve good decision making. Studies focus on decision making because people need to decide routinely during the day and these small decisions are the ones accumulate to form the big decisions and they are smaller parts of the total outcome, but science and scientific approach is often taught as a set of standardized techniques to be memorized and undertaken(Krumm and Kimmie, 2013),and as per Boudreaux and others (2008),If control of variables Instruction occurs at all in a college science course, it tends to be carried out via this pattern. Students are usually unable to go over the reasoning steps necessary in contexts that differ from the examples used during instruction in the classes. Through their investigation, Krumm and Kimmie found that even students who are smart and have a high level of intelligence in areas of mathematical operations in courses such as physics that contain problems in calculus and integration have had difficulties in linking the experimental results and the specific conclusions and inferences that they have reached, while usually science courses provide an appropriate opportunity to develop and reinforce logical thinking and draw conclusions based on available evidence. Boudreaux et al., 2008)

It is a natural sequence of thinking that the person often follows in decision making but giving it the formal mould helps in fixing and developing that pattern of thinking process. It will also be easier to consciously notice if a step has been missing or performed without intention or with no understanding. A classic 5-steps process that may be found helpful as a formal framework of thinking to decide effectively was suggested by Adair (2020). The person expects that logically these five thinking steps need to be followed in strict sequence, but the human mind is not a manmade machine. The notes may be connected in different sequences and different mental chords in the human mind. Making thinking not necessarily a tidy process, but it needs to be made with a sense of order.

1.1. Context of the research Study

Context is the thing that characterize the essence of thought procedures (Rivaz & Saiz 2016). Outside the context, everything is possible ,but nothing is real. The daily faced problems occur at a given time and in a particular space, and this means that they are not an abstraction, they are true facts. However, this contextualization tends to be ignored in teaching. The facts are only seen “on paper” such that it is almost impossible to observe them or see them properly in real life contexts.

In the higher education context, SR is expected by Pyper (2012) to be a limiting factor of the conceptual understanding process in college students. Lawson &Worsnop (1992) found that college students with higher reasoning abilities are less likely to hold pre-instruction non-scientific beliefs about life than the students with the lower reasoning abilities. And these students with higher abilities of scientific reasoning are cognitively more flexible to change their alternative non-scientific beliefs if they have any, and more liable to experience conceptual change. This finding was also quantitatively proved by She and Liao (2010) for grade eight students in Taiwan. She and Liao carried out a two-factorial experimental design study to investigate the effect of the instruction approach, and the effect of the level of the scientific reasoning of the students, on their achievement and scores in three tests which were provided by the researchers to measure their scientific reasoning, concept construction and conceptual change. And they found that; students with higher abilities of scientific reasoning are cognitively more flexible to change their alternative non-scientific beliefs, and more liable to experience conceptual change.

According to She and Liao, we can clearly notice that the term "misconception" is often used in context of various fields of literature, as it is used to describe the reason for the inferences pertaining to incorrect answer patterns rather than looking at the student's answers from an empirical perspective. It makes sense that if students believe in their well-established but false patterns of thought and apply those concepts and thoughts constantly to answer related questions; it is likely that their answers will be wrong as a result of following a specific and consistent pattern of misconception. However, She and Liao it is not necessarily true when the student gives wrong answers that the reason for them is the result of wrong ideas and concepts, but the wrong answer may be due to other factors. That is, wrong answers are sometimes referred to as patterns and products of ideas and responses that resemble and look like misconceptions. Criticism of the interpretation of the concept “misconception” was mainly directed through the identification of the student’s coherent model. This is according to She and Liao because the theories that were used for the explanation of this term were found to be incorrect and were not universally valid.

The “misconception” interpretation criticism according to She and Liao is based on two specific grounds: first, when the students were asked about their concepts and theories that they adopt in their response to questions and answer patterns, their replies were personal, fragmentary, and inconsistent, indicating apparent lack of coherent theories they had. second, it has also become very clear that through the students' responses and answers to the questions that the answer is largely related to the context of the question, and that the answer may change as soon as even a small change occurs to the context of the question (She & Liao, 2010). Emphasizing this point, Kaiser, Jonides, and Alexander (1986) have discovered, for example, that when a question regarding objects moving along a curved path, many students answered that “water that comes out of a bent hose is in the form of a straight line”. And the least answers were conveyed as “less than a ball would come out of a curved tube in a straight line”. However, this criticism can be partially addressed regarding interpretations of the misconception of questions that some experts consider similar may be perceived by students as not alike, and as a result of this significant difference, specific patterns in students' answers cannot be expected. In addition, there are many questions that students answer incorrectly all the time due to the fact that they have incorrect ideas and beliefs for these important questions. Also, in some cases, students provide wrong answers that give the impression that the student does not have a strong and coherent valid theory capable of applying it to the context of the question, even in light of the small changes that occur to the content and structure of the question (Kaiser, Jonides& Alexander 1986). With those challenges of misconceptions and difficulty to transfer scientific concepts to real life, the UAE faced the challenge of not having enough undergraduates joining science and mathematics-intensive streams in high school (< 5%) and choosing then to study STEM disciplines at university (< 30%) (Hitt et al., 2014).

1.2. Statement of the Problem

Despite eminent introduction of scientific fields in the higher education (HE) sector, the undergraduate's failure rates at these institutions remain high. This has progressed problems in many countries such as hunger, disease, unemployment, conflict, and environmental destruction which are growing at a frightening rate as per a study by Krumm & Kimmie (2013).

The researcher's experience in the career of physics teaching in the HE sectors at the middle east, formed a major driver for this current study. The researcher rarely found a high school or college student would prefer to study physics. Most students were joining her classes with a prior perception of a difficult , complex, abstract course. Students often felt that physics is not a relevant subject and has been studied only to complete a requirement of graduation in scientific colleges. So, students will be so tensed while taking the physics course trying not to fail the course and to negatively affect their GPA. Students' tension reflected negatively on instructors' performance as per the researcher observations. Instructors will be teaching to cover certain content knowledge while making sure that the students are able to pass the exam and ignoring cognitive skills enhancement. That made physics classes to take the form of theoretical lectures followed by pre-prepared recipes for practical experiments conducted in laboratories. Eventually statistics and studies showed gaps between what outcomes is the HE sector supposed to provide the employment market with, and the actual outcomes of HE sectors not only in the middle east, but all over the world as per the following literature.

The problems in the undergraduate sector were well tested in many contexts, and the South African context was one context at which these growing problems resulted from serious gaps in the three processes of teaching, learning and application of science, and that is especially clear at physics (Krumm & Kimmie, 2013)

Another study was carried out in the Chinese context. It studied meticulous 1,637 undergraduates throughout several long periods (Ding, Wei and Mollohan, 2016). The students were from distinctive college levels, and diverse fields. Consequently, the study revealed that there was little correlation between HE learning and critical thinking. This study made known the present state of HE institutes in China, and recommended academics and researchers over the world to investigate this issue in their own countries.

One of the many reasons causing the previously mentioned gaps in the HE sectors, is the problem of individuals' implicitly jumping to conclusions while thinking (Alfaro-LeFevre, 2017). That often occurs when any individual tries to keep up with today's scientific and technological fast rhythm changing world. Alfaro-LeFevre, (2017) described this problem as "Ready, fire, aim" problem (instead of "Ready, aim, fire"). Indeed, this problem may lead to failure or cause disastrous results (Alfaro-LeFevre, 2017. p. 97). So, training individuals of the society to think things through in critical manner became essential to cope with a fast pace. Yet, the 21st century learners are not to be like machines programmed to perform tasks with set of instructions. Even the 21st century machines and wireless networks are not going to jump to conclusions anymore. The Artificial Intelligence (AI) development is the reason for that machine transformation. For human reasoning, and in particular the scientific reasoning. With regard to scientific and logical thinking, Kashyap (2021) has presented his theoretical framework for this kind of important human thinking and reasoning patterns in life through intuition by formalizing a wide range of assumptions, definitions, models, and basic components in order to accept intelligence and its determinism as a proven fact in the evolution and development of life for both the individual and society. This does not mean that there is no relationship or correlation between events and outcomes, or that there is nothing to understand once we as readers or students do not understand something or do not see a connection between the events. As a result, this study raises key questions in such cases and situations when things are not completely clear and understood. That is, a set of clear questions and proposals related to increasing intelligence and linking it to our intuition and actual experiences in life. Therefore, this type of studies is constantly being accredited in academic and scientific institutions around the world. More specifically, the topic of this study revolves around the theme of human intelligence, as is the case with the concept of artificial intelligence, which is applicable in machines. In this regard, it must be noted that there is a direct relationship between intelligence and curiosity, the more we stop curiosity or decreases it, the lower the level of intelligence. Therefore, the level of intelligence occurs and increases due to curiosity to know things (Kashyap, 2021). As also has been pointed out by some authors, Boudreaux and others (2008), on the topic of national standards for science education that the main goal of education is for man to develop a firm and concrete understanding of science as a life-long process, and for science students to learn how to learn and get developed through increasing knowledge of what they do not know. In order to achieve this important and fundamental educational value, teachers must assist students and raise their educational skills in order to enable students to reach the conclusions and draw inferences by themselves via analysing the data available to them. Accordingly, the results that can be accessed through this study can help and assist to develop human capabilities in the society in which they live, so that the citizen is a conscious and intellectual person who can reach specific facts by analysing a specific political or social reality from the data available to them about that a specific reality. Thus, the participation of a real educated and responsible citizen, for example in the political sphere, in matters of voting and choosing a specific candidate, is to deal with them responsibly and consciously, without exercising any allegations and fabrication in political discourse by the candidates who are willing to take up public office (Boudreaux et al., 2008).

Originally, the term "decision" comes from a Latin verb meaning "to cut". This phrase is associated with words related to cutting and slitting (Adair 2019).The term "cut off" means in the context of the decision-making process, especially in the area of budgetary decisions, in terms of the merits and demerits of various aspects of action (Adair 2019). It refers to the initial activity of thinking such as when one gets to weigh up the cons and pros, and has to come up with the action, such as getting out your check-book, followed by talking about delivery deadlines, etc. It is a natural sequence of thinking that the person often follows in making decision. but giving it the formal mould help in fixing and developing that pattern of thinking process. it will also be easier to consciously notice if a step has been missing or performed without intention or with no understanding. A classic 5-steps process that may be found helpful as a formal framework of thinking to decide effectively was suggested by (Adair 2019). The person expects that logically these five steps need to be followed in strict sequence. But the human mind is not an artificial machine. The notes may be connected in different sequences and different mental chords. So thinking is not necessarily a tidy process, but it needs to be made with a sense of order. Studies focus on decision making because people need to decide routinely during the day and these small decisions are the ones accumulate to form the big decisions and they are smaller parts of the total outcome(Adair, 2019).It is worth identifying the point called by Adair, the Point of No Return (PNR), at that point of the process of thinking , it costs you more to change your mind or to turn back than to continue with a decision which you know now not to be a perfect one. In many situations that require decision-making, the person or official in charge has a chance to manoeuvre before making the final decision, so that the decision-maker has little time to change their mind. When making a decision, the subconscious and the deep mind checks the decision again, either by accepting and feeling satisfied, or by making one feels the need to review the decision before it is too late, or completely make one change their mind about the topic related to the decision-making situation (Adair 2019)

Faust and Meehl back in 1984, reviewed human cognitive limitations of scientific reasoning and demonstrated that these limitations often lead to judgment errors. He stated in the following,

> “It is my belief, and I freely admit that it rests on faith, that through recognition and effort we will be able to extend our reasoning capabilities. What is to be gained by such an achievement is knowledge for its own sake and a greater capacity to predict and control the events around us. When considering our capacity to use such powers in a consistently just and humane fashion, however, my faith breaks down and I must acknowledge grave concerns. Extending our cognitive capacities may make us smarter, but it will not necessarily make us wiser or bring us closer to ethical perfection. It is my hope that, in pursuing the apple, the latter goal will not be sacrificed.” (P. 165)

Addressing by that the importance of both SR and moral reasoning. How can the educational system then, enhance such an important cognitive skill? the relation between problem solving and quantitative reasoning skills studied in (Tanişli and Dur, 2018). The results of Tanişli and Dur study have shown that exposing students to complex issue circumstances when conducting any session impact the improvement of their quantitative reasoning. It is in this way vital to open students to problems that are rich in quantitative connections that urge students to consider numbers autonomously. Considering the significance of the reasoning procedure as opposed to the outcome, a classroom atmosphere in which students can express their ideas openly ought to be given, and students ought to almost certainly examine "what's going on there?" In critical thinking, it ought to be guaranteed that students focus on quantities and characterize quantities and quantitative connections in the issue before the words (more, times, and so forth.). In this procedure, students ought to be urged to utilize visual portrayals, for example, charts, tables, figures. Moreover, the study of (Tanişli and Dur, 2018) recommended to think about the role of the course textbooks and if they bolster the quantitative reasoning skill.

As per Drummond and Fischhoff (2017), non-scientists are affected by scientific innovations and discoveries in their daily choices. Studies on communication reveal that social media, on the minimum, could lead to a "spiral of silence," in which people often end up sharing opinions only if they expect that their audiences are ready to agree with them. It is true that the Internet, and social media, can destroy the people's best efforts and intentions of critical thinking.(Lack & Rousseau, 2016).

SR is usually practiced whenever the person is faced by a situation where scientific evidence needs to be assessed. That is clear when political debates take place, or deciding on taking a pharmaceutical but not after reading its package insert, or when deciding on activities depend on climate conditions, and also in expecting the economic effects of new healthcare law (Drummond & Fischhoff, 2017) In order to know to what degree may individuals own and apply this cognitive ability and because little is known about the skills which needed by the non-scientists to peruse and assess scientific evidence; Drummond and Fischhoff (2017) built up an individual-difference scale of SR skills (The skills expected to assess the elements which determine the quality of scientific discoveries).Based on research in public comprehension of science, behavioural decision and cognitive developmental psychology, Studying the individual-difference of SR of the American population revealed broad variation in scientific knowledge (Drummond and Fischhoff 2015), and sizeable portion of the population hold convictions inconsistent with scientific evidence. The participants with the higher SRS scores hold convictions consistent with the scientific accord on conceivably debatable issues, such as political and religious convictions, education , and scores on two broadly utilized measures of scientific literacy. They also performed better in a task to evaluate scientific data. Drummond and Fischhoff 's (2017) study recommend that the SR skills highly contribute to the person responses to scientific arguments and scientific results. In analysing the normalized gains in the Force Concept Inventory (FCI), Coletta & Steinert (2020) emphasized considering LCTSR or the Scientific Ability Test (SAT) average scores, because these scores have a high correlation with normalized gain in FCI, so, students' abilities impact on the achieved gains may be greater than the impact of the specific pedagogy used(Coletta & Steinert, 2020)

Studies such as the one by Lagubeau, Tecpan& Hernández (2020) may form real steps in exploring for solution to the stated problem. It was a pilot study of active learning applied in introductory physics course in a public university at Chile. The model was based on research, it depended on a literature review for careful selection and construction of activities reliable to the levels of students' reasoning . The level of SR found to positively correlate to student success. Lagubeau, Tecpan & Hernándezb compared the failure rate post the intervention in experimental group with active learning, and control group with traditional lectures. They detected a significant reduction in failure rate for experimental group's students, even though they do not yet own formal SR. These students being the majority, Lagubeau, Tecpan& Hernández concluded that in the context of a developing country, applying active learning is mainly appropriate to the first year of higher education. It suits the characteristics of this context's student and typical size of classroom, resulting in a reduction of academic failure and improvement of learning, and at the same time the active learning approach is financially sound.

In general, studies such as (Effendy et al., 2018) indicated that the approach of science teaching used by the teacher is one of the major factors that can affect the development of students' SR. the study also stated that the focus of the science teaching in schooling years before the higher education is on the content knowledge. So, by maximally involving them in the scientific approach, the maximum ability of students' SR would be possible. There is a remarkable deficiency associated with the practical education. That is, experimental practical education professors and specialists have not yet developed this science to use the highest levels of scientific methods that can successfully stimulate and encourage the growth and development of sound scientific thinking that can be viable and applied appropriately on the ground (Effendy et al., 2018).

Speaking about the value of a good SR skills for health professionals as an example; and according to several researchers (e.g.Bawazir, 2014; Haig, 2008) it is vital to Make sense of what's beyond their sense. So, for health sciences students it's crucial to embed scientific reasoning skills earlier during the undergraduate level. It has been indicated also that critical thinking which found to correlate with scientific reasoning abilities; is a key skill for clinical reasoning of health professionals (Bawazir, 2014; Haig, 2008). But Moore and Rubbo, (2012) indicated that; the attaining of significant gain in SR entails a reconsideration of the traditional approach of physics teaching for non-STEM majors. In line with Moore and Rubbo (2012) vision, Lawson (2004) also suggested that the traditional teaching approach should be revisited and replaced with an approach that is similar to science practice (Lawson, 2004). It was also found that, the necessity for explicit instruction of scientific reasoning (Brandon & Sibbratthie, 1996, Bao et al., 2009) and the taught topics sequencing are two factors evident for its influence on developing the ability of students to reason scientifically (Bao et al., 2009).

One of the studied approaches for its positive impact on the critical reasoning is the one suggested by Golanbari and Garlikov (2008) in the engineering field, which is called "the Socratic Method of teaching by systematic questioning". Their conference paper reports on the thoughtful and amiable use of the Socratic Method of teaching by systematic questioning - replacing teaching by telling - to foster and emphasize critical reasoning skills of students at the University of the Pacific (USA). The careful use of the Socratic Method (in combination with active learning exercises and traditional lectures) in several topics in engineering, statistics courses, projects and teaching laboratories got the students actively involved, and became excited about the material being taught and the projects and they better master course content and they learned to reason and think more clearly, relevantly, rationally, logically, and ethically. The use and employment of the Socratic curriculum in the above-mentioned educational regions, as mentioned by Golanbari and Garlykov, has greatly contributed to the development of students' basic intellectual capabilities and skills, such as the ability to make inferences and deductions through data collection, data analysis, formulation of questions and assumptions and reaching logical conclusions and inferences. American Society for Engineering Education

Explore-instruct approach is another non-traditional science teaching approach, it resulted from several researchers' efforts, such as Shwartz and Bransford (1998), Mayer (2004) and Lorch et al. (2010), who began to investigate possible ways of integrating problems' exploration and teacher's instruction to get benefits of both, rather than using them separately as contrasting instructional designs. And researchers in education and psychology who advocated that proposal, ended up with this explore-instruct teaching approach (e.g., Schwartz et al., 2011 and Kapur, 2012). In this approach students are being asked to explore a problem, case or phenomena, before being instructed, this teaching approach has been investigated by Loehr, Fyfe and Rittle-Johnson (2014) for its effectiveness in mathematics problem solving, and the students experienced it showed superior learning compared to the ones with the instruct-solve approach. Therefore, considering the student to be the novice mind, and the professional graduate to be the mature expert mind; the goal of this study will be to investigate the suitability of explore-instruct teaching approach in accomplishing successful transition from novice scientific thinker level – base evaluation of reasoning on general common sense as per (Brandon & Sibbratthie, 1996) – to expert thinker level who reason depending on scientific basement.

Dancy & Henderson (2010) found that, in spite of the success of the "inform/convince/distribute" strategy at producing initial pedagogy knowledge and generating interest in change, it did not result in major changes in actual classroom practice. There is a need for a model which accounts for the complications of an actual classroom change. Based on Dancy & Henderson's (2010) results, the needed model should focus on the high levels of reforms currently being made and focus also on external challenges faced by the faculty when trying to implement research-based strategies. But due to Dancy & Henderson (2010), change promotion in instructional practices is poorly understood and complicated. It needs a careful research-based attention similar to the one that has been paied to the development of effective pedagogies and curricula (Dancy & Henderson, 2010).

Thus, the current study carried out by the researcher to study the impact of shifting from traditional (Instruct-solve) to new (explore-instruct) teaching approach on college students' scientific reasoning. It has been carried out for university students in the United Arab Emirates.

1.3. Purpose and Questions of the Research Study

The purpose of the proposed study is first, to investigate the impact of explore-instruct teaching approach on college students' scientific reasoning (SR) ability. Second, to explore the perspectives of participating students who experienced EI regarding the impact of this teaching approach on the scientific reasoning ability.

The impact of explore-instruct approach is being compared to the impact of the traditional teaching approach called by Paul (1993, p.304) "Mother Robin teaching" and being referred to as "Instruct-solve" approach by (Loehr, Fyfe and Rittle-Johnson, 2014).

The instructional design model according to Obizoba (2015) consists of: analysis, design, development, implementation, and evaluation stages. This research can be considered a part of the analysis stage in the instructional design model. The process which was declared by Leonard (2002) to enable the educational system to efficiently exploit the available resources, such as instructors, educating technologies and laboratories' tools leading to the best possible outcomes. When taking decisions related to science curriculum in the higher education institutes locally at the United Arab Emirates UAE and internationally, the current study may offer supporting information for educational policy makers. So, the study aims to:

- Compare the impact of the new Explore-instruct teaching approach with the traditional Instruct-solve teaching approach by comparing the gain in SR of the students experiencing each teaching approach.
- Explore participating college students' perceptions regarding the EIP.
- Offer recommendations and supporting data for instructors and educational policy makers.

The research has been proposed to answer the following questions:

RQ 1. Is there a difference in college students' (SRA) development if taught by explore-instruct teaching approach compared to the instruct-solve approach (traditional approach) in Physics?

RQ 2. What are the perspectives of the college students regarding the explore-instruct approach after experiencing it?

1.4. Outline of the thesis

The study includes after this introduction a *literature review* chapter with a suggested theoretical framework, and related previous studies in topics of: Science Education in the college level, some Instructional approaches start with exploration, inquiry and problems solving, followed by articles in definition, elements, aspects, and styles of Scientific Reasoning (SR), articles about the topic of SR versus common sense, topic of Science reasoning and teaching approaches, and finally the SR assessment.

The *methodology* chapter then shows the context of the study and the utilized instrumentation for the data collection and data analysis. That chapter also presented the ethical considerations followed prior to and while conducting the research.

The fourth chapter is the one of the *data analysis and results*. That chapter started with demographics' description of the studied sample, followed by presenting the results of the processed quantitative data. Starting with the reliability of LCTSR to the study . Internal validity of the collected data was also presented, followed by descriptive statistics of the study variables before and after the intervention. In addition to the inferential statistics to study the impact of the intervention, and finally the quantitative results were summarized. The qualitative data was then presented, and email interviews were analysed into themes presented and summarized at the end of this chapter.

Discussion of the results was the fifth chapter, in addition to discussing the data, limitations, recommendations and future implications of the study were presented and finally the conclusion of the study was presented. Bibliography list of *references and Appendices* are then at the end of this work.

CHAPTER 2

THEORETICAL FRAMEWORK & LITERATURE REVIEW

2.1 Theoretical Framework

Reasoning is a good synonym for critical thinking because according to Alfaro-LeFevre (2013) it implies deliberate, careful thought. Everyday reasoning is inductive reasoning. An example of that, drawing inferences about the weather, or anticipating the taste of a meal or any other form of an uncertain probabilistic, approximate reasoning. but the creative reasoning that leads to a discovery is still considered a mystery as per expanding philosophical perceptions of reasoning by Nersessian (2008)leaded him to preliminary conception that supports un-equating the reasoning with logic. that is in addition to the beliefs of many Philosophers that the basic units in dealing with theories are most often models and not propositional networks or trusted systems. but how are these models facilitating reasoning about phenomena? Nersessian (2008)believes that modelling practices themselves are parts of genuine reasoning, and many researchers believe that creating new conceptual representations which occur as a result of good reasoning, happens mainly through model building, manipulation, assessment, and adaptation. That is clear in contemporary studies in pursuit, discovery, and application research in the sciences , a signature feature is the different types of modelling, that is in addition to many philosophers agreed that the basic units in working with theories for scientists are most often models, not propositional networks or axiomatic systems (as cited in Nersessian (2008)). modelling usually comes first in building a theory as shown in histories of the sciences, secondly comes formal expression and abstraction in proverbs and laws of theories. Nersessian (2008) was not concerned much like other studies about the relations that represent and link between models and targets, or about other pragmatism issues; She was instead concerned about how can models facilitate and figure in reasoning about any phenomena. To prove that practicing modelling comprise genuine has required that Nersessian develop philosophical concepts of reasoning. Creative reasoning that is model based cannot be considered a simple recipe, is not always useful to produce solutions, in addition to the fact that its most standard application can produce incorrect solutions. To approve these model-based reasoning practices as reasoning process; Nersessian have accordingly needed to confront the deeply rooted philosophical concept that sees reasoning equal to logic. Even Though it is possible to get the same results of the process of model-based reasoning when means of logic is used instead but that possibility is available only after the creative drive has been completed, so still the creative work or the discovery will not be explained by this equating of reasoning with logic, and so leaving these processes a mystery. Nersessian (2008) agreed with philosophers like Popper and Reichenbach that there is no “classical logic of discovery” , but she disputed the equating of reasoning with logic. The concepts of reasoning which philosophy has been adopting are too confined and upon adopting it emerged the misguided view that reasoned processes cannot come up with discoveries according to Nersessian (2008). Traditional philosophical versions of SR see it as logic-based: applying inductive or deductive procedures to sets of proposals. Nersessian indicated that Deductive reasoning is something based on deductive inference that represents the true standard for reaching real conclusions. However, this good concept of logical reasoning does not apply to much scientific practice. In fact, the "deductive hypothesis" involves generating the hypothesis and then testing the inferential results associated with it. This hypothesis points out to an important truth, which is the possibility of availing variances and some differences, and the necessity of acceptance of the truth of science's error, especially with regard to setting initial hypotheses and formulating preliminary introductions. However, logical thinking has genuinely contributed to the development of the concept of safety for induction and established the logical position on the issue of the inductive nature of scientific inference, which should be explained in broad terms, as shown in the current Bayes' calculations, starting from the possible introductions and ending with the use of sound inductive logic, which ultimately leads to the maximum possible conclusions as per Nersessian. The thing that remains puzzling about this issue is how to determine the prior probabilities of a given hypothesis. Creative inference is often called "abduction." But in these cases, the nature of the deductive processes of inference known as abduction is largely indeterminate. The deductive processes of inference provide an effective means of determining the nature of inflationary thinking in relation to inference, abduction. In addition, it helps in the issue of inductive reasoning calculations by providing a real basis on which to determine the previous possibilities of hypotheses that result from optimized analogue reasoning, such as the interpretation of Maxwell's hypothesis that electromagnetic waves usually propagate at the speed of light. However, this model may be loosely described at the present time as it represents a system with interacting parts and with representations of those interactions. Model-based inference can also be studied by using conceptual, physical, mathematical, and computer models, or by some combinations of these models._In this context, the conceptual models are dealt with by Nersessian. These conceptual models are imaginary systems and patterns that are designed with the goal of serving as behavioural, functional, or structural analogues to the target phenomena. It is also considered that these models are dynamic, especially in future situations and conditions that can be identified by simulating the model in a mental image. These various forms, analogue, visual and simulated, are widely used to solve scientific and ordinary problems alike. These models are also used in a regular or innovative and creative way. However, the uses of these models do not differ in terms of their type, but they are in a continuum according to the historical cognitive calculation. And the use of one side of the model leads to the importance of studying the other side, because a deep examination of the model requires a deep study of the whole range. For instance, when using the specified model in a practical and creative context, often the different representational methods are exploited in one loop in order to solve the problems. As for the historical model, it is explained in detail how Maxwell has created visual representations of imaginary analogue models that represent aspects of a new conceptual system, which is known as the "electromagnetic field". These models have been visualized and portrayed on the idea that they are dynamic models - as moving in time - or in terms of visual perceptions of them being illustrated in clear text that leads the reader to how to move them mentally. Reasoning has significantly helped build and manipulate the model in deriving and developing theoretical hypotheses, mathematical equations, and experimental predictions. In this regard, it can be claimed that by understanding the existing thinking framework of these models that represent a new scientific model, how this has led to the development of intellectual work carried out by model-based thinking across the spectrum (Nersessian, 2008)

Lawson (2004) has provided an explanation of how knowledge developed, as he agreed in principle with Piaget in the process of intellectual development that requires self-organization by practicing thinking, perseverance and working to comprehend the outputs of that process and then understand its various divisions and procedures. Piaget (1976) has also added that the process of speculative abstraction, as he calls it, has strengthened and developed this issue of comprehension. Contemplative abstraction has well developed and moved from the stage of using automatic and spontaneous verbal communication patterns to the stage of using clear and explicit rules of discourse by using verbal expressions to correct and direct behaviour. The reflexive abstraction process usually occurs when individuals are asked to express contradictory reactions (using verbal discourse markers such as: if then, but, any) and the resulting state of mental imbalance in the matter of thinking about what these people do first and then what they present as well as their arguments during their interaction in conversational situation with others later. So, we find that the underlying reflexive abstraction is the presence of a contradiction in the physical environment and what others say of words. The value and benefit of reflexive abstraction is that it benefits the person and gains them more knowledge, awareness and perception, as well as the acquisition of stronger skills regarding the procedures used in the development of knowledge, and thus expressive knowledge and the procedures for such knowledge are divided (Lawson, 2004).

2.1.1 Adaptive Control of Thought (ACT) theory

In 1976, Dr. Anderson espoused a theory that represent the human cognition (Leonard, 2002). The cognitive architecture by ACT is the base of the below proposed framework of the study. Dr. Anderson utilized his theory in two main directions of research; 1. The enhancement of the ACT itself, 2. The application of the theory to create artificial learning systems for computers programming and mathematics. Anderson calls application no.2 ACT-R. The theory states the existence of three types of memory for the human mind. The declarative memory which is responsible for storing ideas and facts in a semantic structure, and it is a long- term memory. Another memory is the procedural memory which is also a long-term memory that takes the production of the declarative memory and make logical inferences about it. Working memory is the third type of memory and it is a short-term memory. The working memory is responsible for the conscious thought, retrieving declarative knowledge, carries out action sequences found in procedural memory, and it also receives the new information collected from the environment to form new sequences (Anderson, 2009).

2.1.2 Phase sequence theory

The phase sequence theory was set by Donald O.Hebb in 1946 (Weitz et al., 2011) showing that every new-born is born with random neuron interconnections. Along with the development of the mind the neural networks will be organized. Hebb defined the cell assembly to be a set of neurons linked together and it will be activated when experiencing an environmental event. The cell assemblies' sets also interconnect to form phase sequences. Cell assemblies that follow each other consequently in time form a phase sequence of thoughts that follow a logical order.

2.1.3 Cognitive load theory

The Cognitive load theory discussed by Sweller (1988), Paas, Renkl & Sweller (2003) defines three types of cognitive loads, these cognitive loads identify the capacity to process new information by the learners' working memory during the instruction stage. Intrinsic load; is the load results from the unfamiliarity of the material or the complexity of it. Extraneous load: is the load that depends on the quality of the instruction materials and tools, so better quality of instruction decreases this load and makes more capacity available for the learning process reducing the third load which is responsible for the learning (German load). The effects of a cognitive load provided the final, logical explanation and justification for the cognitive load theory. When studying and examining the effects of cognitive load, one should not lose sight of the issue of theoretical constructs that has led to the emergence of these effects. The cognitive structure of the cognitive load theory should also be used for the purpose of evaluating the potential effects of any educational intervention. The main goal of these instructions, according to this cognitive structure, is to increase the cognitive and perceptual vessel usable in the long-term memory, stored in the information store. In the long term this saved memory helps in achieving the systematic work based on the ecological linkage. This long-term memory helps also the person by actively participating in difficult or impossible events and activities. The principle of "borrowing and organizing," which is the issue of obtaining information from other people, is the best way to acquire knowledge. However, if it is not possible to retain the knowledge in the long-term preserved memory, in this case the information available in the working memory, which is often limited in terms of duration and capacity, should be processed, especially when new information is dealt with similar to the principle of narrow change. New information can be obtained during the problem-solving process according to randomness as the principle of formation, but this issue requires working memory resources, and this is not available for learning. In addition, educational procedures that do not meet the ambition to increase knowledge at the level of long-term memory, at the same time while reducing any unnecessary load on working memory, are likely to be ineffective. In this situation, we must not ignore human perceptual knowledge when we design instructions, on the understanding that they will not be additional or optional (Sweller 1988; Paas, Renkl & Sweller, 2003).

2.1.4 Abductive theory of scientific method (ATOM) by Haig, (2005)

According to ATOM by Haig, (2005), scientific research runs as follow: "guided by evolving research problems comprised of packages of empirical, conceptual, and methodological constraints, sets of data are analysed to detect robust empirical regularities, or phenomena. Once detected, these phenomena are explained by abductively inferring the existence of underlying causal mechanisms. Here, abductive inference involves reasoning from claims about phenomena, understood as presumed effects, to their theoretical explanation in terms of underlying causal mechanisms" (Haig, 2008, p. 1019).

2.1.5 Proposed framework for the current study

The framework set by the researcher for this study (Figure 2), is combining the four previously mentioned theories. The student's mind is to be considered the one with the cognitive architecture described by the ACT theory. When experiencing the traditional way of Instruct–solve (Gray squares'& Gray arrows' sequence in Fig. (2) 1-7 follow the arrows' direction) the student's working memory will try to upload the new information in the declarative memory. If needed later the working memory will try to retrieve the declarative knowledge, and logically match it and save it in the procedural memory when needed to apply it in practical situations or perform problem solving. When experiencing the Explore-Instruct approach (blue sequence 1-9 follow the blue arrows' direction), the sequence will start at the working memory call for procedural information during an inquiry or experiment performing (Explore phase).

- In case of previously stored procedural knowledge, the working memory will solve the problem or execute the experiment utilizing the prior procedural knowledge (4) and if new information produced out of that sequence it will also be restored as a procedural knowledge.
- If prior knowledge was not fully found in the procedural memory, working memory will utilize the available procedural knowledge and try to match the new received information during the instruction phase (5-6-7), to solve and perform the inquiry activity, more cell assemblies of logical matching will be formed because of such cognitive sequences.

Connecting what preceded with the cognitive load theory, the process of forming more procedural memory sequences by the student during the exploring phase, and prior to the instruction; should increase the capacity for procedural learning (reduce German load) during the instruction phase.

While experiencing the Explore-Instruct approach, more phase sequences that start with procedural memory cell-assemblies formation are expected by the researcher to occur prior to declarative cell-assemblies (Phase sequence theory), the thing that makes the explore-instruct approach cognitively closer to the process of scientific way of thinking. According to the abductive theory of scientific method (ATOM) by Haig, (2005) “phenomena serve the important function of prompting the Search for their own understanding by constructing relevant explanatory theories”, that is consistent with the explore-instruct, where the observation or natural phenomena is noticed or experienced first (explore) and then the mind analyses it to reach a solution or an explanation, ending up by elaboration and discussion of theories during the (instruction phase) lecture. And empirical data by Fabby and Koenig (2015) suggested that the lower the students’ reasoning abilities, the more dependence are them on basic recall of facts (declarative data) and simple procedures in solving problems.

Based on the previously proposed theoretical framework, which is supported by the cognitive learning theories (students’ understanding),and with the assistant of constructivist theories' teaching approaches (Barz & Achimaș-Cadariu 2016); The researcher expects that; as the teaching approach of explore-instruct make the students to experience more phase sequences start with logical matching; SR mechanisms may then precede other cognitive processes when experiencing daily life situations.

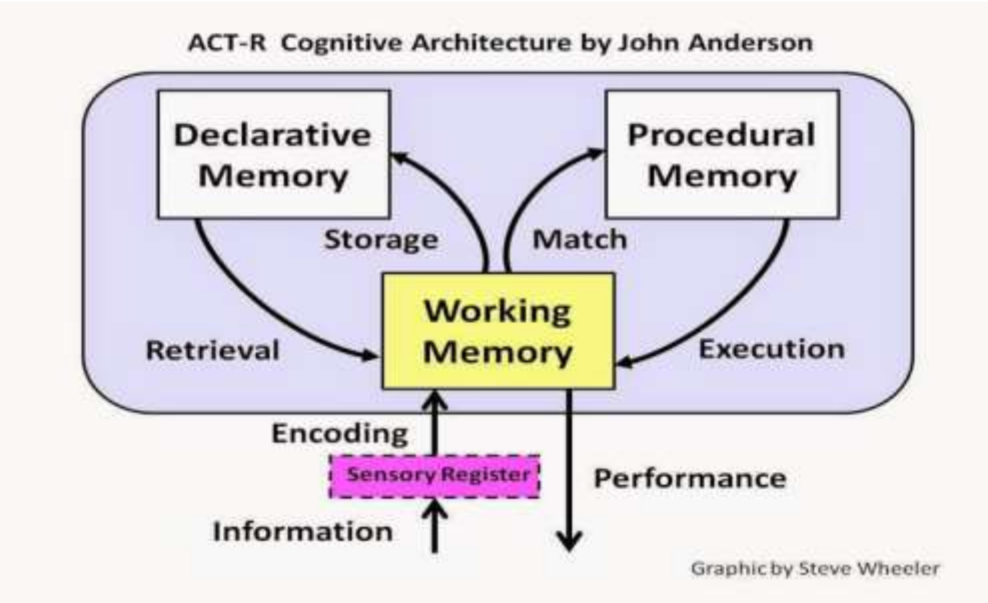


Figure 1. Anderson Cognitive Theory Architecture

2.1.5.1 Development of SR and nature of its synthesis

Lawson (2004) has given and articulated a good description of the development of scientific thinking, the nature of its synthesis, and the importance of its pivotal role in the acquisition of scientific knowledge. This scientific thinking is viewed as a systemic scientific institution for generating hypothetical ideas and inferences that are based on premises of research, development and testing of alternative explanations. The construction of logical hypotheses and the testing of explanations are usually done by using a number of thinking styles, both major and minor. Inference is a complex process, especially when it is at its highest level, because scientific explanations generally include an assumption of entities that are imperceptible. It also requires the availability of sub-arguments used in testing these hypotheses, in order to link the hypothesis under test with its results that are extracted from the available data. In fact, science is characterized by great and high accuracy and interpretations of tests in nature are considered the correct way in which the brain processes information automatically, whether by visual recognition of objects or by testing descriptive and causal hypotheses on daily basis, or by testing advanced theories. The main point in this topic relates to building complex HD arguments is that if concepts are not divided and / or subtypes of inference are not used adequately, any attempt to build and maintain these arguments within the framework of working memory and use them for the purpose of drawing conclusions and building concepts, will thus be lost. As a result, teachers should realize the importance of what students bring with them in relation to the different stages of their intellectual development. Add to this the importance of apprehending declarative knowledge of scientific thinking. The importance of meaningful and active education lies in practicing various forms of science, where students face challenges represented in having buzzling notes and then begin to participate in explaining their contexts and testing hypotheses - and here some of their ideas may conflict with the evidence and arguments presented by other students. (Lawson, 2004).

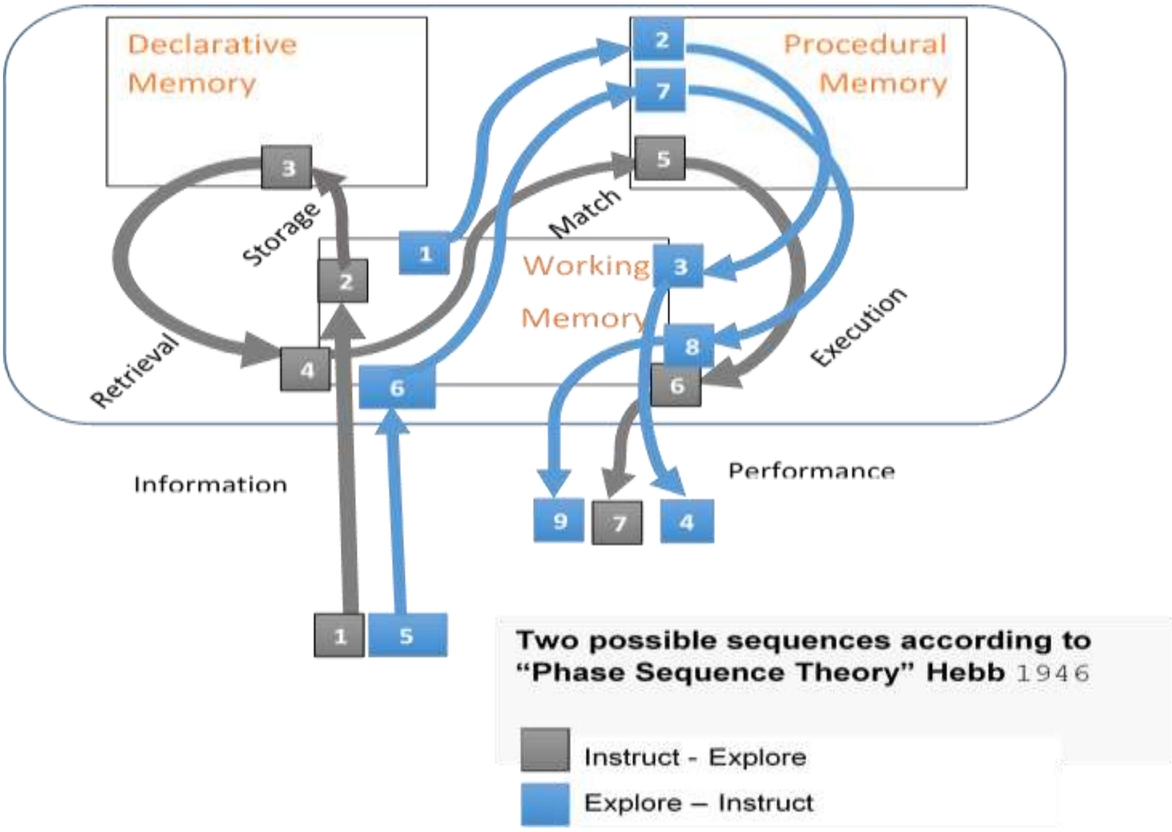


Figure 2. Theoretical framework for the proposed study.

2.1.5.2 Reasoning Mechanisms in the Artificial Intelligence field

The cognitive frameworks and the inferences associated with them were discussed by Qadir (2016), The mechanisms by which the obtained data and the learned knowledge are being related, are called the inference mechanisms. These inference mechanisms happen as, and represented by reasoning mechanisms, which are typical part of the cognitive procedures. Developing cognitive networking in the artificial as well as in the natural intelligence face the challenge of developing reasoning. Reasoning mechanisms are showed by Qadir (2016) to be different from one context to the other. So, reasoning procedures in any scientific field will not fit the cognitive networking field. In the radio cognitive networking, it is possible to simulate a form of the perceptual radio architecture identifying many different perceptions. In fact, the different cognitive structures include a group of different approaches within the cognitive cycle, which confirms their profound impact on the human thinking process. Generally, cognitive templates are classified as either basic or stemming from the unified theory of the cognitive process, and the basic elements that make it up may be communicative, symbolic, or mixed. Also, for example, the cognitive frameworks that emerged from the unified theory of perception could be simple. The processes of Orient Control, Decision Making (OODA), Criticism, Exploration, Comparison and Adaptation (CECA) may all be complex, and this is embodied, for example, in SOAR, Storm, and ACT-R. The implementation of any specific knowledge framework is directly reflected in the logic associated with the interior framework. The OODA framework, for example, depends on the nature of the feedback loop to model adaptations related to changing environmental conditions. The logic in the decision-making process includes identifying all the available changes in the configuration of devices, and then determining the best option to meet the challenges surrounding the new situation, and then making necessary changes to reconfigure the devices within the framework of a constant feedback loop. Originally, this cognitive framework was developed by the US Department of Defence with the aim of describing the methodology used by US combat pilots during wars and air combat. This framework is also applicable to interactive situations. It also extends CECA's framework for OODA to appropriately describe the proactive decision-making process. In this context, logic depends on social cognitive awareness, which is a combination of multiple entities working to solve complex problems. In addition, this framework does not rely on interactive external observations, but rather focuses primarily on proactive situations that are targeted towards a specific goal. All OODA and CECA frameworks are applicable to all frameworks of different cognitive radio networks. Normally, the SOAR framework consists of a complex and robust package of programs specifically designed to approximate rational behaviour. It also extends the SOAR framework to the development of cognitive architectures that are synthesized from a biological perspective (BICA) and at the same time suited to cognitive radio network applications. Finally, theories are framed within the ACT-R framework according to the way human cognition functions, allowing users to represent tasks and measure time to perform tasks accurately and dynamically. This matter has potential application in the decision-making process within cognitive radio networks. Generally, this section has focused on inferences with their various possible types, methods and perceptions within the framework of specific practical applications for learning and thinking within the context of artificial intelligence (Qadir, 2016).

2.2 Literature review

Curricula and handbooks for teaching science subjects taught by faculty members of colleges, and other higher education institutions that teach a wide range of natural sciences subjects, have been developed and updated, including the courses offered by traditional departments that teach biology, chemistry, physics, earth and space sciences, and also in new departments of health sciences, Computer, Engineering, and Information Technology (Mintzes, 2006).Contemporary Educational research even provided models of teaching and learning which are more than conventional lecture-laboratory design and offered rationales for these new practices in the college classes. Experienced faculties studied and wrote about different areas in the college science teaching such as active learning, innovative teaching methods, impacting learning, attitudes and motivations including science anxiety, the utilization of technology, and special challenges that emerges in college science teaching, challenges like effectively teaching culturally diverse classes or teaching students of determination (Mintzes, 2006), and recently the challenge of effectively teaching remotely online because of pandemic circumstances.

2.2.1 Science Education in the college level

When teaching life sciences to undergraduates; one would think of a teaching approach that prepare them for future career and future learning. In a society which is invaded with rapidly growing technology; the traditional teaching methods need to be revisited to evaluate its suitability to deal with the modern society compared to alternative approaches that proposed based on available learning theories (Spector et al. 2013, p. 7). Lewis Wolpert (1994) in his book showed that achievements of early technology is not science, it is problem-solving based on experience and common sense. And science did nothing to aid technology until the 19th century. And it has always been dependent on technology for apparatus and ideas. Wolpert presented the case of the Japanese industry which made people think about its achieved success and found that it was mainly based on its ability to apply science. Emphasizing the importance of science education in today's rapidly growing world in all different aspects of life, it was indicated by Wolpert that the goal of science education is not only limited to helping students acquire basic concepts related to each of these sciences, but rather goes beyond that limit to prepare students and raise their intellectual abilities and skills in order to accomplish more complex tasks in addition to providing them with the basic means to solve problems through data collection and analysis using advanced hypotheses and scientific research tools such as induction and deduction as effective tools to reach logical scientific results. Thus, we find that the study of science education has attracted scientists and science students to use scientific thinking in order to solve the real problems facing humanity in various walks of life Wolpert (1994). And as per Yuliantiand others (2020) one of the major goals of advanced education is to prepare students well and appropriately with the necessary skills and abilities required in the future. This is because advanced education offers an excellent opportunity to use logical thinking based on drawing inferences through case studies. Moreover, this advanced education can stimulate students' motivation permanently and continuously to invoke logical scientific thinking in discussing various issues. In particular, if students choose science-based colleges especially with the highest classification that require their students to pay great attention to the promotion of their scientific thinking, which these colleges assume on all their students. In this context, the Lawson Test that closely linked to scientific thinking (LCTSR) was used on a group of Chinese students estimated at 1,637 students who are enrolled in long-term studies and affiliated with different and varied fields and colleges of study and have high academic levels. The result of this test was that it was discovered that the logical thinking performed on these students showed little variation over the four long periods of teaching these university students, and that there is little relationship between college-level learning and the level of scientific thinking of these students. Unfortunately, this test illustrates the fact that the current environment of higher education institutions does not develop students' innovative intellectual and creative abilities and skills. Therefore, the result of this test urges all those interested in developing higher education in the context of developing logical scientific thinking around the world to investigate this problem and work immediately to solve and promote the level of higher education institutions in their countries, given that the development of scientific thinking represents the cornerstone of the teaching and scientific research process (Yulianti et al., 2020).

So, thinking about applying science to life in the UAE context, educators should consider understanding science in real life context and consequently relating real life phenomena to science while teaching (Bao et. al., 2009). Boudreaux et al. (2008) has criticized the teaching of science courses as being taught in a traditional manner as mere a set of standard techniques that students should only memorize and follow as such. This traditional science teaching style according to Boudreaux and others, tends to have control over variables. In this case, students' intellectual and creative abilities are stifling and restricted and thus they are unable to comprehend and follow the steps in the work of logical thinking process in order to deal with circumstances that differ from the examples presented in the course.

While depending on a scientific approach in learning, the area of dominance (affective), converts the substantive teaching material or substance so that students "know why." Then the (psychomotor), or the skill area, takes the conversion of the teaching material to enable the students to "know how". The (cognitive) domain of knowledge uses the transformation of teaching material to enable the student to "know what." The final outcome will be the progress and sense of balance between the competence to be a human being having the knowledge and skills to live appropriately (hard skills) and a good human (soft skills) with aspects of good attitude. (Effendy et al., 2018, p.505)

In the recent century, an emphasize has been put on the students' understanding of scientific reasoning and the nature of science. That found to be necessary to educate qualified people in the society. That is viable via application of effective science teaching in the classrooms. And the process of teaching science subjects effectively can only take place as per Yuksel (2019) through students getting acquainted with the nature of the scientific thinking models of scientists and using these models themselves in order to comprehend the nature of the sciences they study. In this regard, the subject teachers of these sciences should provide a good explanation to their students about the basic scientific thinking patterns, and how to use these patterns of thinking while dealing with their respective subjects. The scientific thinking approach is employed to reach logical and acceptable answers to scientific questions, as well as to use these patterns in explaining the observable and unobservable contexts. Thus, the use of scientific thinking patterns by students will inevitably provide a better understanding of the nature of science (Yuksel, 2019). Therefore, in order for students to be able to properly use the different scientific thinking patterns during their study of the nature of science, the curricula must be reviewed in a manner consistent with achieving this goal, bearing in mind the impossibility of excluding the computational and mathematical thinking (Dobhansky, 1973). It is of great importance, in this context, when reviewing the curricula of science subjects that special priority must be given to the theories and analogue models because they represent the cornerstone of these sciences, i.e., they explain the cognitive nature of these sciences. It is also necessary to review and address the evolutionary calculations of the origins and types of the universe, the stars, and other species, in order to familiarize students with one of the most important types of evolutionary scientific thinking models, an example of this in the case of self-development observed by Dubhansky in biology, by indicating that there is nothing logical in this science is only in light of evolution (Dobhansky, 1973). The second vital issue that should be addressed regarding the need for students to understand and use the patterns of scientific thinking is that this thinking provides a rationale in the teaching of school sciences and thus we can bypass the dominance of focusing on the content of the teaching material that it is at the expense of developing scientific thinking, noting that teaching the content of the course described in the curriculum is considered the basic organizational framework in defining the curriculum (Kind & Osborne, 2017). Moreover, Kind and Osborne, (2017) indicated that, in today's world that scientific knowledge is in a state of continuous increase, which places an additional burden on school curricula. However, in fact, the primary goal in developing school curricula must be for the curriculum to be a means of introducing students to the six main patterns of scientific thinking and the need to use them in solving problems facing students inside or outside classrooms, so that students can use scientific thinking as an educational and cultural logic to understand the nature of science.(Kind & Osborne, 2017)

The recommended characteristics and skills needed by the 21st science graduate, are all correlated to the challenges and problems which will be encountered in the contemporary life and society (Spector et al. 2013). It has been widely argued that the complexity of those problems encountered by the 21st society has increased because of the advance in technology, and due to globalization and the rapid change of circumstances surrounding these problems. That requires the education system to focus on improving problems solving skills of the students and to consider them as a central objective to be accomplished (Spector et al. 2013). Major question that needs to be answered then, is what are the sources of complexity in most of those problems? According to Michael Spector (Spector et al. 2013) some sources are related to the level of challenge of the problem itself, because modern science widens the distance between science and common sense (Redekop, B., 2009) and we are not good in making sense of things which are very small (the atom and the cell) or very big (the earth and the universe) that the new technology deals with. So, the advance technology made the problems people encountering more challenging.

Other sources of problems complexity are related to the person involved in solving those problems, sources such as the person's prior knowledge and background, also the person's distracting matters if there are any, and the limitations which are encountered during human reasoning when delayed results or non-linear relationships are included in the problems as per (Dörner, 1996). Prior knowledge was tested as an intervening factor between Epistemological commitment and Scientific Reasoning when the last two checked for having relationship among college students in a study by (Zeineddin & Abd-El-Khalick, 2010). Results of this study proved the believe that epistemological commitment delineate the process of reasoning and showed that the higher the epistemological commitment the higher is the quality of reasoning, though Prior knowledge only impacted reasoning when epistemological commitment was week. Thinking of a combination of the formerly mentioned limitations in human reasoning, and the increased complexity of the modern problems; one would recognize the need for educational pedagogies with characteristics, to enhance the problem-solving process in contemporary society and contribute to academic and everyday life success. One of the primary goals of the educational process is to develop students' thinking abilities. The study of Lawson (1985) examines and reviews the role of science and developmental psychology education programs that seek to assess the validity of official thinking and its relationship to the educational field, by examining Piaget's theory. In this regard, there is a question that arises that needs to be answered by us, namely, should one of the main goals of school education be to produce formal thinkers? To answer this question, there are other six secondary questions as per Lawson that present themselves in the following context: First, what is the role of biological maturity in developing formal thinking? Second, are the formal functions and tasks of Piaget theory valid and reliable? Third, has the issue of formal thinking represent a general and unified pattern of intellectual performance? Fourth, what is the effect of the presence or absence of formal thinking on the issue of students’ academic achievement? Fifth, is it possible to teach formal thinking to students? Sixth: What is the functional or structural nature of advanced formal thinking? Finally, the study has reached a general conclusion that although Piaget's theoretical framework review and what emerged from it leave a number of methodological and theoretical problems that have not been solved, the study has significantly contributed to creating a solid background on which to base the achievement of qualitative progress in the educational process that is centred around scientific reasoning, whereby we can thus develop in all aspects of life, more specifically through thinking, we can constitute the doctrine of morality (Lawson, 1985). In fact, the development of thinking in general, and in particular scientific thinking, and the capacities of argumentation are among the most important goals of teaching science subjects. Scientific thinking and argumentation skills are applied through the participation of students in activities that require the integration of inquiries with scientific methods in scientific social contexts (Putri et al., 2020). In this regard, the Inquiry Levels model is considered one of the effective tools and methods in the emergence and development of the scientific environment for students based on thinking and dialectical abilities. The inquiry levels model adopted by Putri, and others research is the learning method by exploring, inquiring and interactive explanation. Also, Putri and colleagues found that in any stage of the levels of inquiry there are five educational structures, known specifically as observation, manipulation, generalization, verification, and application respectively. The aim of the study is to identify the capabilities of primary school students in using scientific thinking and controversial abilities in solving problems by applying the inquiry levels model based on the SSI perspective in the material of global warming pertaining to science education. The approach used in the lesson was the weak trial method in addition to the pre and post-test design. The surveyed sample consists of seventh grade students in the primary school. The tool used in the test was scientific thinking and the test of scientific arguments. The results of Putri and others' study have shown an increase of 0.43 for changes in students 'scientific thinking skills, as well as an increase of 0.39 percentage for changes in students' dialectical abilities. The results of that study have also indicated that there was a significant difference between the results of the pre and post-test, according to the analysis of the t-test for the double sample (Putri, Siahaan and Hernani 2020).

Although the levels of scientific thinking among higher education students are higher in the different fields and disciplines of science and engineering compared to other university majors, there is a relative stability in the level of thinking skills of students that was measured by the LCTSR pattern, and this is in approximately the four levels of higher education, which it has become a recurring pattern for all student majors at the undergraduate levels. It is also worth mentioning, in the context of this topic, that there are results of studies conducted in China indicating that higher education does not have a noticeable effect on the development of students' level of scientific thinking, regardless of the students' specializations and the university or college to which they belong (Ding et al., 2016). The acquisition of laboratory capabilities and skills is one of the key tools for developing scientific and research thinking skills among science students in general and an essential part in providing knowledge for students as in Ortega and others (2019), as the laboratory experiments conducted by these students help in understanding the uncertainty in the issue of experimental measurements. Despite the importance of developing the laboratory capabilities of science students to comprehend the nature of the applied dimension of these disciplines, science students, even when they are in their final university years, are still immature in dealing with laboratory experiments from the perspective of employing sound scientific thinking pattern, despite the fact that scientific thinking is an issue that is natural and innate in humans. One of the reasons for the lack of understanding and development of laboratory capabilities may be attributed to the way in which the methods of approaching these laboratory exercises are designed, as laboratory experiments should ensure that the theoretical aspect that was previously explained in the classroom is understood. To bridge the gap between theoretical teaching and laboratory experiments, Ortega and others propose in their study a method of learning, namely the Inquiry-based Learning (IBL) approach. Under this approach, students are expected to work on the application of FTIR method as well as to use special thermodynamic concepts in order to study the equilibrium process of formation of H-linked benzoic acid dimers in solvents with different polarity (Ortega et al., 2019).

Unfortunately, it was found that the development in the issue of scientific thinking occurs slowly, as the focus in teaching science subjects is still limited to the cognitive aspect that relies on memorization and comprehension of the subject matter only, instead of paying attention to the development of reasoning and analytical skills that are considered fundamental to scientific thinking (Kind & Osborne, 2017; Piekny & Maehler, 2013). In addition, laboratory experiments have not been used properly and effectively in the development of scientific thinking among students, as students spend more time in collecting data and following up on the implementation of procedures instead of taking advantage of that time in analysing, discussing, and drawing conclusions from collecting such data. There is also another problem associated with the use of inquiry as a mode of teaching, as it is not relied on its philosophical and theoretical foundation on which it is established (Erlina et al., 2018). The inquiry model is originally used by teachers to build the knowledgebase in students, and also aimed to contribute to the development of free thinking in these students. However, Erlina and others found, that free thinking takes the inquiry away from achieving its basic goal, and then leads it to the testing of hypotheses instead of playing its major role in setting hypotheses (Erlina et al., 2018).

In order for the United Arab Emirates to maintain a competitive advantage in the higher education sector as an attractive educational hub for students from various countries of the region and to be able to achieve the sustainability of its educational institutions, some important educational issues should be addressed. A pressing issue in this regard is to focus on the question of quality assurance in education. That is, higher education institutions must focus first and foremost on the quality of the educational service provided to their students, rather than focusing on the number of enrolments and graduates from these institutions as well as increasing the number of their entities. To restrict the rapid growth in the number of new private academic institutions, research should be conducted on the needs of the labour marketplace before approval of any new academic institution to work in the field of education. This is very important because the private education sector is the main resource and supply for the higher education sector in the UAE (Ashour & Fatima, 2016). Thus, restrictions should be put in place for the private educational sector related to ensuring the quality and keeping pace with educational programs that comply with international standards in education. The quality academic service provided by this sector can definitely assist our colleges and universities to meet the requirements of the labour market through qualified graduates in both the government and private sectors (Ashour & Fatima, 2016). If these issues are addressed properly, the UAE can achieve advanced positions in the ranking of international reports in the field of competitiveness among the different universities of the world, and most importantly, the UAE can better invest in human capital through quality education, which represents the backbone in achieving sustainable development, which represents the strategic goal to the state (Ashour & Fatima, 2016).

Howard et al. (1995) have argued that providing professional advice by experts due to their accumulation of expertise to businesses regarding the selection of employees to fill in vacancies is no longer sufficient at the present-day business world to ensure the success of these businesses, despite the correlation between employees' adaptability and the success of institutions. Thus, the immediate objective in researching this topic in the study of Lepine and colleagues is to attempt to achieve some development in the issue of adaptability in relation to organizational success. Lepine and others' paper examines the relationships between three known individual differences, namely general cognitive ability, conscience, and openness to experience, and then the paper continues to investigate the process of performing decision-making before and after an unexpected change occurs in the stages of achieving this task. Aligned with the objective of Lepine and others' study, some researchers point out that the performance of decision-making prior to the change represents the performance in the context of a steady task, and the performance of decision-making after the change is an indicator of the type of adaptive capacity required in modern businesses. Lepine and others' study actively sought to provide an effective and visionary insight on the significance of adaptability, as they made an effort whether it is possible to predict the ability to adapt in the issue of decision-making performance by using the types of individual differences when new employees are selected. It was concluded that openness, and conscience can indeed be used to predict adaptability. In fact, Lepine and others found that there is more power to predict adaptability compared to decision-making performance before an unexpected change occurs. The surprising thing is that Lepine and colleague found that there is a negative relationship between adaptability and conscientiousness. Unfortunately, this reverse relationship constitutes an obstacle to maintain the adaptability for a business that is established on a systematic criterion. However, the results of this study can only be generalized after being tested on a global level. On the whole, Lepine and others wish their study will act as a catalyst in further research on how to manage effectively human resources in the dynamic environments that operate in today's world (Lepine et al., 2000).

2.2.1.1 The Nature of physics as a foundation subject in collage.

Physics content is seen by students as an abstract complex subject, and it is logically sophisticated (Bao et. al., 2009). Redish (1994) stated the requirements of Physics that makes it seen difficult by the students:

> "Physics as a discipline requires learners to employ a variety of modalities (methods of understanding) and to translate from one to the other-words, tables of numbers, graphs, equations, diagrams, maps. Physics requires the ability to use algebra and geometry, and to go from the specific to the general and back. This makes learning physics particularly difficult for many students. One of our goals should be to have our students understand this, be able to identify their own strengths and weaknesses, and while building on the former, strengthen the latter" (p.801)

Redish also describe an analogy used by students to learn Physics called "the dead leaves model" where he stated:

> "…it is as if physics were a collection of equations on fallen leaves. One might hold $S = \frac{1}{2}gt^2$, another $F = ma$ and a third $F = -kx$. The only thing one needs to do when solving a problem is to flip through one's collection of leaves, until one finds the appropriate equation" (p. 799)

But as per Rusli (2012); and due to the nature of the contemporary knowledge community which based upon science; The basic physics lectures may lend themselves usefully for developing the scientific thinking of the students. That is expected because according to Schuster and others (2007) during teaching Physics problem solving, the learners apply the physics' principals by passing over the worked-out solutions and that reflects the richness of the learner's thinking when tackling real-life issues. And as an experimental science, Physics surrendered itself to the natural enhancement of the hypothetic-deductive reasoning; this reasoning is widely used in science practice and it involves utilizing a hypothesis to explain observed facts, designing an experiment and predict its outcomes using the proposed hypothesis, carrying the experiment out and then compare the outcomes with the predictions and accordingly accept or reject the hypothesis (Etkina et al., 2007).

2.2.2 Instructional approaches start with exploration and inquiry.

Instructional strategies are the techniques used by the teachers to enhance students' independence, and strategic learning (Alberta learning, 2002). When students select the appropriate strategies independently and utilize them efficiently to achieve assignments or meet objectives, these strategies become learning strategies. Effective learning and instruction strategies can accommodate a variety of students' differences and can also be used throughout subject areas and grade levels (Alberta learning, 2002).

Think–pair–share is an exploration cooperative learning strategy. In that strategy, the teacher poses question or a topic. Students individually think about the posed question for a certain period of time, often one to three minutes. students then pair with partners to discuss the problem, and accordingly clarify their thoughts. Then, each pair may share their answers with the whole class. (Alberta learning, 2002)

As per Fyfe, DeCaro, and Rittle-Johnson (2014), good amount of evidence is in favour of delaying the instruction for a better performance. And the approaches of teaching which depend on cognitive theories are found to be more enhancing to scientific reasoning skills than the ones that relate to the behaviourism or the constructivism (Leonard, 2002). And according to Lawson (2004), The key implication is that teachers should teach using methods that enable students to develop the essential reasoning abilities if they hope to assist their students to become scientifically literate. In brief, instruction should not "fit" students' recent cognitive developmental levels only, but it should also trigger students to improve to higher levels. The evidence has shown that the best way to achieve this, is to teach science the same way science is performed, as a process of critical thinking and inquiry in which the assumptions are generated freely and tested rigorously. Three learning approaches from the cognitive school which delay the instruction phase were described by Yilmaz (2011), these are inquiry learning, discovery learning and problem –based learning (PBL).

2.2.2.1 Inquiry-based learning (IBL)

According to Gillani (2010) IBL is a type of learning that incites a teaching approach that depends on Piaget's theory of cognitive development. It appears to be similar to scientific exploration method. The main goal of this teaching method is to develop the higher-order thinking skill (Gillani, 2010). The students are being engaged in a process of finding a solution for a problem or investigating an issue. Combinational reasoning, propositional reasoning and hypothetical-deductive reasoning are three types of reasoning support this teaching method (Gillani, 2010). Combinational reasoning requires examining different issues simultaneously to solve a problem. Propositional reasoning involves an examination of assumption to solve problems. Hypothetical-deductive reasoning entails a hypothesis to be considered to address a problem. Gillani (2003) also indicated five phases for IBL: "1. Phase One: Puzzlement or intellectual confrontation by presenting students with the problem to create a state of disequilibrium in their mind.2. Phase Two: Students will hypothesize a reason for the puzzlement.3. Phase Three: Students will gather new information in regard to the hypothesis. Then they isolate relevant information and organize it based on some core concept or theme.4. Phase Four: Students analyse the data they have gathered and organized, and they postulate a possible answer for the hypothesis, which explains the original puzzlement.5. Phase Five: Students test their hypothesis as a possible answer" (Gillani 2003, pp. 60–61)

Leonard (2002) also showed that IBL is based upon five principles: 1. a driving question to be provided, 2. Group work to solve the question, 3. Learners present a work that shows their understanding of the issue, 4. Learners extend their contact to include scientist exploring the same issue, 5. Communication of information to be shared by the group via electronic network. In total, Lazonder and Wiskerke-Drost, (2015) indicated that the science teachers may effectively utilize inquiry-based methodologies to help students learn topics of science. However, that can effectively be achieved only if children or students are effectively guided, and the guided methods studied in Lazonder and Wiskerke-Drost (2015) research proved better to unguided inquiry-based methods. And both IBL methods of Control of Variables Strategy (CVS) and the inquiry method, were found to be similarly effective, the choice is up to the personal preferences or up to the practical considerations. So, the instructor can either offer the inquiry first and support it by dividing the task in convenient subtasks to help the students focus on the variables one by one, or the instructor may teach the students the fundamentals of the CVS before providing the inquiry task. Both teaching methods were found to enable the students to obtain better conclusions and consequently gain a better understanding of the topic. Added to that, both methods also assist students to practice their scientific reasoning skills and strengthen it, the thing that has been considered by lots of countries world-wide to be a key competency (Lazonder & Wiskerke-Drost, 2015).

2.2.2.2 Discovery learning

The discovery learning model is one of the constructivism theories models which indicates that students build their knowledge via an environmental interaction or through an experimental activity. The constructivist theories are more effective in developing the conceptual understanding and SR of people than the other learning theories (Jensen & Lawson, 2011, Wuriyudani et al., 2018). Discovery learning involves conscious reasoning to find a new knowledge that students have not been learning it. This learning style may encourage students to process data or hypotheses actively and competently, and to summarize problems (Piaget, 1973).

The report of “The Vision and Change” has put emphasis on teaching science as science practicing. Learning of science should not only focus on the content, but also focus on the scientific process of thinking (*Vision and Change in Undergraduate Biology Education: A Call To Action.* 2011). Performing Problem solving from the perspective of a discovery-learning is how learners acquire knowledge of principles under focus in classrooms (Leonard, D. 2002). So, the cognitive skill of reasoning is a major skill underpinning problem solving, and its importance extends far beyond a particular discipline to be the backbone of the quantitative research paradigm in any area (Sharp et al., 2011). This cognitive skill is selected to be the dependent variable in the proposed study.

2.2.2.3 Problem-based learning (PBL)

PBL entails presenting students with, open-ended, ill structure or real-life problem that has many possible right solutions and asking them to find solutions to the problem. opposing the traditional teaching method that teaches facts first and then present the problem, this approach begins by introducing the problem on the basis of the students’ existing knowledge and teaches in a relevant context the facts and skills (Yilmaz, 2011). According to Bransford & Schwartz (1998); it is challenging for the teachers to share with the novices (students) the experts’ knowledge that is delivered in the lectures. But they found it to be easier when sharing the same knowledge with an expert. So, the idea beyond allowing the students to experience the problem first, is also to raise the level of the students’ expertise, which allow for more precise acquiring of information, and consequently ground the bases for enhancing other activities such as generating questions, elaborating and learning (Bransford & Schwartz, 1998). Problem solving from a perspective of organizational learning is also related to efficient decision making (Leonard, 2002). Dunker (1945) who is considered a key researcher in problem solving; defined the problem as “a situation in which a desired goal cannot be attained by direct application of known operators, and so there has to be recourse to thinking” (Holyoak & Morrison, 2005, p.345). So, it was concluded from Dunker definition that human practice problem solving through carrying on different deductive or inductive reasoning activities and decision making (Holyoak & Morrison, 2005).

2.2.3 Explore-Instruct teaching approach

In the current study, a teaching approach which is to be utilized for teaching both major and non-major students is being looked for. So, the interest is toward raising cognitive skills of reasoning more than enhancing students' acquiring of very specialized content, which may be an objective when teaching major students. Also, regarding the spent learning time, the targeted non-major students are to learn the basics of the subject in a limited time of one semester (3.5 months). Therefore, due to the non-physics major context; the researcher looked for instructional designs which fit the case, and at the same time a teaching strategy that makes the student more active while promoting creative scientific thinking. Ersoy and Başer (2014) concluded that not like the conventional teaching strategies, PBL procedures has a commitment to the improvement of students' creative thinking aptitudes and they build the expertise of the person's creative thinking which is one of the higher-order thinking abilities. This achieved point is so vital as particularly these days there is an urgent need for people having the capacity to think creatively. Therefore, the proposed pedagogy may be of the active learning ones such as; cooperative discovery, IBL or PBL. And based on the theoretical framework demonstrated in figure (2); the explore-instruct teaching approach which is considered a cooperative IBL and PBL strategy has been chosen here to be studied for its impact on SR of the college students. The explore-instruct approach is a result of several researchers' efforts like Shwartz and Bransford (1998), Mayer (2004) and Lorch et al. (2010), who began to investigate possible ways of integrating "exploring problems" and "instruction" to get the best benefits of each, rather than using them separately as contrasting instructional strategies. Eventually researchers in education and psychology who followed that proposal, ended up with this explore-instruct teaching approach (Schwartz et al., 2011; Kapur, 2012). The classroom standards essential for good-functioning Physics Education Research-based instruction are shown by Dancy and Henderson (2010) to be: hands-on activities, student-student interaction, equity in teacher-student interactions, in addition to sense making over answer making (Dancy & Henderson, 2010)

Diagram of Explore-instruct strategy is presented in **Figure 3** to show the following components of EI teaching strategy that had been studied in the current study:

1- Students **Explore** questions, problems or issues provided by the instructor and aligned with on or more learning outcome of the lesson.
2- Students Propose ideas for Solutions, and Instructor, Guide students' trials and thinking.
3- Instructor**,** **Instruct** topic and demonstrate applications, and student discuss and apply problem-solving of problems attended prior to the instruction.

   Order of the gears' rotation is important hear. Starting with rotating the exploration gear, urges the second and accordingly the third gear motion.

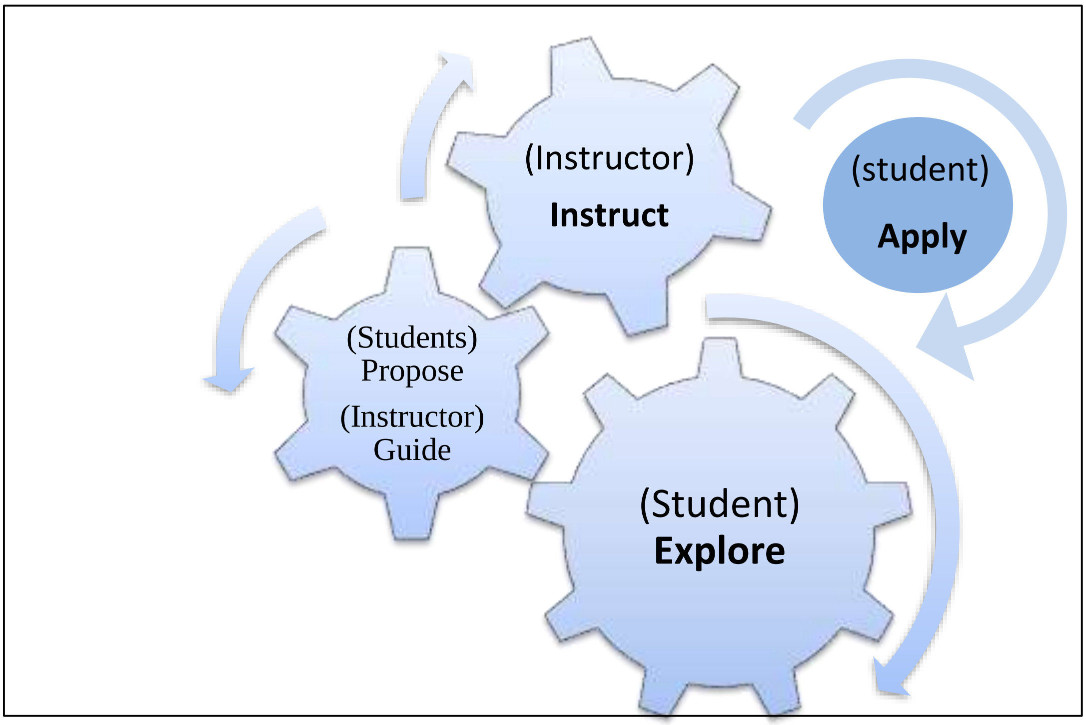


Figure 3*:* Explore-Instruct strategy

A brief physics sample lesson (Vectors Addition and Subtraction) adopting the EI strategy is proposed by the researcher and shown below in

table ***1***.

The template of the lesson plan was adapted from (Ndihokubwayo et al., 2021) and the original lesson plan was by Dr. Zanj K. Avery at the College of Engineering, Computer Science, and Technology , San José State University . The researcher adapted it to follow EI teaching strategy. For lesson plans that have laboratory experiments related to them, the laboratory experiments to be planned before the theoretical class, and laboratory reports to be turned in after the related theoretical class.

The Prerequisite knowledge of this lesson as per Avery's plan:

1. How to draw, label, and interpret an x-y-z coordinate system.
   a. Basics of number lines
   b. Plotting points in a two-dimensional coordinate system
   c. Plotting points in a three-dimensional coordinate system.
2. How to calculate the lengths of the sides of a right triangle using the Pythagorean Theorem.
3. How to calculate the angles of a right triangle.

Table 1: Explore-Instruct Physics Lesson Plan.

Term: ………….. **Date:** …………... **Subject**: Physics.

Course: Introductory Physics I for science and engineering. **Unit:** Vector and Scalar Quantities.

Time frame: 90 min. **Class size**: 25 students

Learning Materials:

- Calculator
- Notebook
- Graph paper
- Pencils
- (x3) 12 inch/30 cm rulers
- Protractor

Lesson title: Vector addition and subtraction

	Concept	Activities	Problem or Question	Learning Objectives
Explore		**a) Take quick quiz**	**(MC Questions 1-5)** ***5 min.***	

	a. Vectors may be added graphically using the tip-to-tail method b. Vectors may be added graphically using the parallelogram method.	b) Using graph paper, begin with an x-y (rectangular) coordinate system, practice adding vectors using their components in accordance with vector addition problem solving strategy 1.3. Use the following video as a reference: c) For added practice, do examples 1.7 and 1.8 (Young and Friedman) on your own. Compare your results with textbook solution.	Q1) The figure above shows three ropes tied together in a knot. One of your friends pulls on a rope with a force of 3 Newtons and other pulls on a second rope with a force of 5 Newtons. How hard and in what direction must you pull on the third rope to keep the knot from moving? You can use the graph paper. 10-15 min.	1. Add and subtract vectors graphically.
Guid	Differentiate and guide - Type of Special Educational Needs to be catered for in this lesson and number of learners in each category: Slow to understand Physics concepts, weak mathematical background, etc.			
Explore	a. The base and height of a right triangle represents the x- and y-components of a vector and that the hypotenuse of a right triangle represents the resultant vector of these two components. b. Components are not vectors- they are quantities/numbers that simplify vector analysis by allowing us to use trigonometry to add vectors.		**Q2)** A racing car is accelerating at 20 m/s^2 70 North of West. Find the acceleration of the car in the north direction and in the west direction. ***10-15 min.***	2. Apply knowledge of the x-y coordinate system and properties of right triangles to decompose a vector into its x-and y-components.
Guid		*Instructor*: Note down the difficulties that groups face and individuals' capabilities to learn	You can use the graph paper and properties of right triangles.	

		what type of guidance is needed.		
Explore	Given the side of the triangle parallel to the x- and y-axes, we can use the distance formula to find the length of the hypotenuse which is the distance between two points on an x-y coordinate system.		A plane flies with a velocity of 52 m/s east through a 12 m/s cross wind blowing the plane south. Find the magnitude and direction (relative to due north) of the resultant velocity at which it travels. ***10-15 min.***	3. Recognize that the Pythagorean Theorem is a mathematical model that defines the relationship between the lengths of the sides of a right triangle and can be used to find the length of one side of a right triangle given the lengths of the other two sides. 4. Add two vectors at right angles. 5. Apply knowledge of how to use the lengths of the sides of a right triangle to find its angles.
Guide		*Instructor*: Note down the difficulties that groups face and individuals' capabilities to learn which groups can present in the next session.		

Instruction	All previous concepts related to all LOs of the lesson.	Present topic to cover all LOs of the lesson and connect it to problems been tried at the beginning by the students. Focus on difficulties faced by students and pitfalls.	Use Examples of simple questions different from the ones attended by students at the beginning of the session to cover each LO. ***20-25 min.***	Students recognize LOs 1-5. 6. Recognize unit vector notation and practice using unit vectors in accordance with vector addition problem solving strategy.
Application	All previous concepts related to all LOs of the lesson.	Students: solve the pre-instruction questions. Instructor: Give HW and practice work sheets to be solved outside the class.	Questions been posed by the instructor at beginning of the lesson: Q1, Q2, Q3. ***15-20 min.*** HW and practice work sheets. ***Outside the class time.***	Apply knowledge of right triangles and their related trigonometric functions to: 7. Perform vector calculations using components (Adding vectors using their components). 8. Perform a simple vector addition in three dimensions 9. Find the x- and y-components of a vector given its magnitude and angle. 10. Perform vector calculations using unit vectors. (Adding vectors using unit vectors).

2.2.4 Definition, elements, aspects, and styles of Scientific Reasoning

A cognitive skill that has been declared by several researchers an essential cognitive skill supporting the scientific literacy and Critical Thinking (CT) is the Scientific Reasoning (SR) skill (Faust and Meehl, 1984, Pyper (2012); Lawson & Worsnop (1992); She and Liao (2010); Morris et al., 2012 etc.). An actual theory of scientific reasoning was proposed by Simon and Newell (1971). Scientific reasoning, according to Simon and Newell, is a problem-solving method that involves a problem space that includes the original state, the target state, and all possible states in between, as well as operators, which are the acts that can be taken to shift from one state to the next. Referring to the problem space theory, understanding scientific reasoning requires looking at the types of representations people have and the behaviour they take to get from one state to another (Simon & Newell, 1970). From the perspective of science literacy (Hazen and Trefil, 1991 & Giere, 1984) "Scientific reasoning represents the cognitive skills necessary to understand and evaluate scientific information, which often involve understanding and evaluating theoretical, statistical, and causal hypotheses." (Bao, 2009 p.1)

From the *research point of view, i*t is broadly defined by Zimmerman (2005) as the term; "includes the thinking and reasoning skills involved in inquiry, experimentation, evidence evaluation, inference, and argumentation that support the formation and modification of concepts and theories about the natural and social world" (Bao, 2009, p.1). In 2007, Zimmerman stressed on that, SR includes "the skills involved in inquiry, experimentation, evidence evaluation, and inference that are done in the service of conceptual change and scientific understanding" (Zimmerman, 2007, p. 172). It includes an extend of cognitive and metacognitive abilities, and it is considered an aggregate and recurrent procedure that requires coordinating between the theories and the proofs (Kuhn, 2011). The objective of this patterned cyclic procedure is knowledge production or an existing knowledge alteration. Scientific reasoning includes the capacity to create, test, and review hypotheses and theories and to think about the cognitive process itself (Zimmerman, 2007). In this way, SR and development of epistemic cognition are closely related, which can be characterized as cognitive forms centred on epistemic issues. And SR procedures in the knowledge seeking and proofs evaluation processes are believed to be drove by epistemic cognition (Schiefer et al. , 2019).

From *the science literacy* point of view, SR was defined (Brown et al. 2010) as the term represents the cognitive skills essential to understand and evaluate scientific information, that usually involve understanding and evaluating causal, statistical, and theoretical hypotheses (Bao, 2009, p.1). But human reasoning extends far beyond the scientific disciplines to support decision making in everyday life (Faust &Meehl, 1984; Bhat, 2019). Alfaro-Lefevre, R. (2017), indicated that reasoning is a good synonym for critical thinking, because it implies deliberate careful thinking. This reasoning is guided by faith, culture, ethics and other factors that generally support it (Faust &Meehl, 1984;Lloyd 2013; Alfaro-lefevre, R., 2017).

The *two types of reasoning* according to many educational scholars; (e.g.,Holyoak and Morrison, 2005; Lawson, 2002) are deductive or formal reasoning, which is the process of retrieving conclusions which are essentially correct if the premises are correct, it depends on testing hypothesis and supported by logic. And the other type is *inductive*, which is the type of reasoning with premises that are not conclusive to relate the results to it, and it rely on inferences due to analogy for example or beliefs and not necessarily due to logic (Johnson & Christensen, 2014). The reasoning skills according to Holyoak and Morrison (2005), involve the ability to explore the problems, form and test hypothesis, isolate variables and manipulate them, and skills of observing and analysing the effects.

The elements of reasoning according to Nosich (2005) are the ones illustrated in figure (3) below. Each time we reason something the 8 elements are always present. The alternatives which are the different choices available during the reasoning are also always present, and in addition also to the context at which the reasoning is taking place in (p.47). In Appendix IV there is an illustration of the intellectual standards to be applied to these elements of reasoning in order to develop intellectual traits according to model of critical thinking by Paul and Elder (2019). These standards include, clarity, accuracy, relevance, logicalness, etc. which if had not been applied to the elements of the reasoning, no efficient progression in intellectual traits to be occurred.

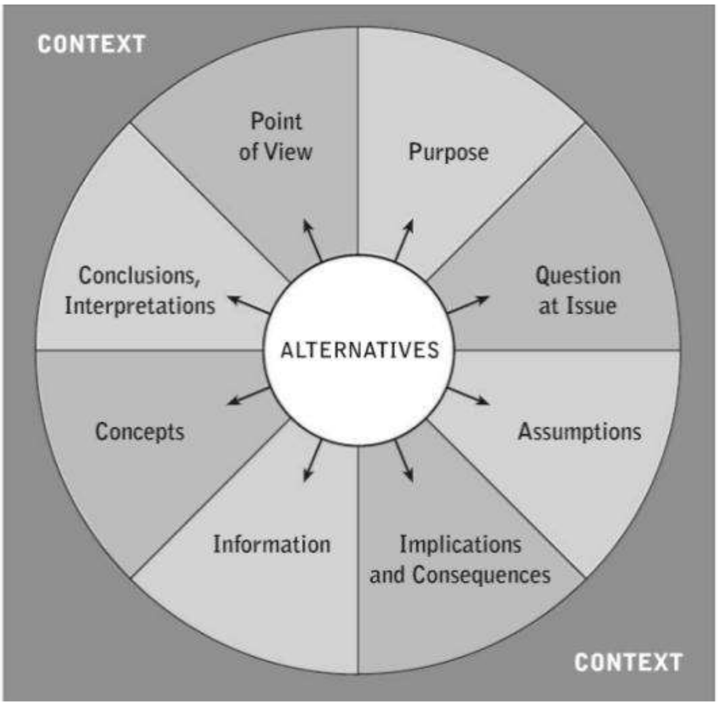


Figure 4: Elements of scientific reasoning (Nosich, 2005).

Table 2: Scientific reasoning dimensions tested by the LCTSR.

Item numbers	Reasoning pattern
1–4	Conservation of mass/volume
5–8	Proportional reasoning
9–14	Identifying and controlling of variables
15–18	Combinational reasoning
19–20	Correlational reasoning
21–24	Hypothetic-deductive reasoning

Note. The numbers indicate the questions of LCTSR which test each skill.

According to (Kind & Osborne, 2017) Answering the three main scientific questions about the material world, is the critical aim of the scientific reasoning theory (National Research Council, 2012b; Osborne, 2011). The questions are: first, the ontic question of what exists. Second, is the causal question of why did that happen, and finally, the epistemic question of how was that known?

How researchers commonly create and build answers to these questions is, Kind and Osborne (2017) insist, best found within the ignored work of cognitive researchers who are studying the history of science. Two main additions in this discipline are study of “shaping the deduction in Greek mathematics” by Reviel Netz’s (Netz, 1999) and Alistair Crombie’s study of “Styles of thinking in the European Tradition” (1994). The cognitive researchers of history of science assumes that human thinking is much superior caught on employing an authentic approach and by taking an outside viewpoint that looks at the items of that thinking; moment, that there are no common, all-inclusive rules of thinking. Instep, the cognitive devices utilized by science have risen verifiably as an unexpected cultural outcome of particular contexts, and given a different historical sequence of events, which may be existed in a different frame, or not exist at all (Kind & Osborne, 2017).

By the usage of this historical methodology in studying SR, Crombie’s exploration of scientific reasoning at the time of the Greeks indicated not only one but six different ‘styles of reasoning’ and it claimed also that the history of science within the European culture is the history of argument, and vision (Crombie, 1994, p. 3). The achievement of science can at that point be credited to the advancement of cognitive instruments, assets and styles of thinking that have been utilized to emphasis a set of ideas—ideas that have at first appeared foolish, such as the thought that day and night are caused by turning of the earth planet and not by Sum movement, or the thought that the continents were one piece of land, or the thought that all species on earth have grew over ages. Particularly, Crombie contends the following six styles of scientific reasoning:

1. *Mathematical Deduction*, that defined as the usage of mathematic as a mean to represent the world and using it in a deductive argument. Of the first people to initiate this style of reasoning were the Greeks. The key to all science is representing the physical phenomena in numerical quantities or in algebraic symbols. Entities from its all kinds can be symbolized in mathematical form, and mathematics is sole main scientific languages used as a means of making deductive estimates not only in engineering , but also in all sciences (Kind & Osborne, 2017).
2. *Experimental Evaluation,* in this style, the prediction of hypothetical models, creation of various patterns, and distinguishing one form of object from the other through

the usage of empirical investigation. Galileo is generally seen as the key figure who started this style of thinking when he dropped two different sizes cannonballs from top of pissa tower testing his theory which states that all masses would drop with the same acceleration. Since that exploration, empirical investigation is commonly used to evaluate scientific and to review claims. Only the ideas that persist such investigations are the ideas been hold by the people to be true (Kind & Osborne, 2017).

3. *Hypothetical Modelling*, that is building hypothetical and analogical models to denote the world. The progression of science happens through building models to explain whatever the scientists observe. Too big things such as the solar systems and or very tiny system like the cell or the atom are represented by analogical models, and these models are used to come up with predictions and are now embedded in computers' software's to simulate the possible behaviour of the real world. So, models are essential tools for scientific reasoning of the real-world behaviour (Kind & Osborne, 2017).

4. *Categorization and Classification*, that is ordering of range of things by taxonomy and comparison. Classification assists in determining what exists. that process is a major standpoint of science, and it has been essential in constructing our understanding of the world. Classification and categorization are achieved via experimental exploration and field work, and unless agreed on the ontic entities which exist, there is no until we agree on the ontic things that exist, there can be no shared language to be used in reasoning about such things (Kind & Osborne, 2017).

5. *Probabilistic Reasoning* is the statistical analysis of patterns in populations and the calculation of these patterns' probability. The determination of patterns is an essential feature of the sciences and the basis, for instance, of the science of epidemiology. For example, the link between a phenomenon and its causal is sometimes related to correlation between the causal effect and the phenomena without really knowing the causal mechanism. Poisson and Gauss are the two scientists whose contributions have made major impact to the establishment of the epistemological standards used to reason about the presence of patterns, in addition to a variety of methods for expressing variation and the probabilities of its existence (Kind & Osborne, 2017).

6. *Historical-Based Evolutionary Reasoning*, which is the construction of historical narratives of the origins of the development of living species, of the elements, of planets, of the solar system, of the universe, and more. A major aspect of reasoning in the

scientific field is the trials to express the origins of the material world with its characteristics. These trials depend on building theories and hypothesis about possible events that might have happened in the past. Examples of such theories are Darwin's ideas developed from his comprehensive observation of the patterns' variations exist in nature and questioning how such variations could have happened. constructing mathematical models has also been practiced presenting some evolutionary accounts. Such models and theories have stood up because they were the best possible available assumptions for what exists, and as noticed not because of any usage of "the scientific method" (Kind & Osborne, 2017).

Each one of these 6 styles of reasoning illustrates the way the sciences answer the epistemic question of how people know. The cultural achievement of any science can't be complete without all the 6 styles of scientific reasoning. And, as these styles of reasoning are the main contribution of science to the modern culture, they provide a rationale and framework for a choice or elimination of any content. Moreover, emphasizing on what we know without understanding the way of how we know it (procedural knowledge), or the cognitive skills that are used to identify the existence of this knowledge (epistemic knowledge), is similar to offering a description of a great building, without any expression of how it was built, or what a creative accomplishment it represents. The procedural and epistemic knowledge are essential to be learned by the students before leaving the formal education, otherwise they will lack a major aspect of the knowledge which is needed to assess scientific arguments (Ryder, 2001).

An important form of reasoning in addition to the discussed styles of deductive reasoning, is the induction, which is actually a persistent and a multidimensional cognitive activity, that is considered a form of reasoning of appropriate interest. Induction may be analysed by asking simple questions that involve cartoon pictures to young children, or it can be analysed by asking adults to make probability decisions after been given several complex verbal arguments. Thus, induction by its nature is uncertain, there are still a set of interesting rich regularities associated with it. Induction also is related and even central to a number of other cognitive skills, such as analogy judgment, classification, probability judgment, and decision making. For instance, a lot of the induction studies were concerned with classification-based induction, and example of classification-based induction is the person assumption that his or her neighbour sleeps, depending on the fact that the class of human being sleeps. even if that neighbour has not been seen sleeping. Also, similarity and induction are so related, and many scientists use similarity in their induction account (Heit & Feeney, 2007). Finally, the value of studying the induction comes from the fact that the induction studies are having the potential to be revealing, as most of the reasoning of the people is inductive reasoning in its nature, and there is a rich set of data associated with it, and because induction is connected to other cognitive activities, its studies reveal a lot about cognition in general and not only reasoning (Heit & Feeney, 2007).

2.2.5 SR and Critical Thinking

Critical thinking contributes to achieving the greatest extent of effectiveness in solving our problems in the best possible way through the use of careful observation and synthesis of a set of correct principles and facts as well as dealing carefully with the procedures of uncertainty (Saiz & Rivas, 2016). These traits exist in this type of thinking, which is known as critical thinking. The most important characteristic of critical thinking is its ability to achieve the best explanation for a problem or dilemma, and that there can be no other better explanation in the context of this problem. In this case, an algorithmic solution is used, which usually follows the approach of critical thinking. Saiz and Rivas (2016) has stressed that there is an urgent need for faculty members in higher educational institutions to harmonize different teaching methods and approaches with policies associated with critical thinking. The failure to keep up with using critical thinking skills and the lack of harmonization and development of teaching methods with the principles of these crucial skills undermine the enormous potential material resources that are currently used to achieve the qualitative leap in higher education. Using critical thinking skills in the higher education institutions, can support achieving learning outcomes and save a great deal of effort and time. As a result, Saiz and Rivas, (2016) aimed through their proposal to provide procedures that allow them to prove that the interpretation of a problem is correct and certain and that a certain science develops because it transforms gradually and depends less on mathematics, statistical probabilities, and some branches of physics are at the top of this hierarchy. That is, critical thinking provides predictions with greater accuracy and certainty, not with a certain degree of probability.

The research conducted by Mustika, and others (2019) also explores the relationship between critical thinking and creative scientific thinking that develops students' skills in sound scientific thinking. To achieve the goal of the study, Mustika and colleagues experimented a test of 42 students studying science in grade 11 at a private school in Bandung. A single package of critical creative thinking was used to solve a sound problem. Mustika and colleagues used the open questions and the scientific thinking skills approach in relation to the multiple-choice skills associated with the sound concept of problem solving. Among the most important results that came out of this study is that the average scores of students in critical thinking and creative scientific thinking, respectively, were 23.67 and 17.36 from the maximum score of 64 and 48. These results show low achievement levels. In addition, the average score for scientific reasoning skills is 36.70 out of a total score of 100, and this also indicates a low level of academic achievement. Thus, the ratio of the relationship between scientific reasoning skills and creative critical thinking skills is 0.57.

The extent of clinical thinking can be indicated by the level of education, specifically the year of education, and the number of training days in hospital health care. Together, the results showed a percentage of 38% (equivalent to p <0.001) And as critical thinking which SR is one of its dimensions is a key skill for clinical reasoning (Bawazir, 2014; Haig, 2008); the researcher expects that the SR also is partially explained with the factors of year education and practical experience in the job environment (Leijser & Spek, 2021).

According to Koenig et al. (2019), sound and effective decision-making process essentially requires the capabilities and skills to identify and then managing to control variables, and thus this is the main starting point for work in the review and examination of Koenig and his colleagues' laboratory. To illustrate this, the correct decision-making requires obtaining complete information, distinguishing and filtering inputs, aggregating large amounts of collected information, weighing inputs based on relevance, considering biases, and engaging in data analysis to improve information because sound final decisions are entirely dependent, based on complete knowledge that has high accuracy, indicated by Myatt (2012), as suggested by Kuhn et al. (2008) in an integrated scientific and logical framework, with the aim of coordinating theoretical evidence. These are higher levels of COV and are important for the shift from concrete thinking to thinking formally.

2.2.6 Scientific Reasoning vs. Common Sense

The difference between "common sense" and "science" was well established by Locke, J. in 1689. In a discussion about human understanding, Locke stated that "knowledge is based in a systematic observation of the physical world, anything else is "faith, or opinion, but not knowledge", He clearly established the position between science and common sense with the "justified belief" concept in between. This establishment by Locke leads us to conclude that this justified belief requires the involvement of mind in a scientific reasoning process. Therefore, developing SR can be one of the means to reach an ideal level of cognitive abilities; the level at which scientific facts and common sense become congruent to each other. And that is possible through training the students to always set their predictions and beliefs and then justify those beliefs. Paul (1993) demonstrated this in his book, when he showed that the mind can be fit if it practices processes of critical thinking, because the mind is like the body in its need for the practice to be fit. Figure3 (Paul& Elder, 2019, p. 25) is showing the scientific mind development. At the top stage SR become a second nature. We may then consider this second nature a common sense. Green (2016) on the systematicity theory, which was first presented by Hoyningen-Huene (2013), has pointed out that this theory is quite appropriate because it has successfully managed to clarify the relationship between scientific knowledge and common sense. However, Green found that in terms of focusing on knowledge products and general categories of daily and scientific activities, he has not clearly indicated many aspects related to the thinking processes involved in the issue of learning and practicing science. Through this theory, Green found that Hoyningen-Huene (2013) has indicated the existence of gaps between scientific knowledge and ordinary knowledge within the framework of specific topics. For example, his emphasis on the issue of hermeneutics is an act incompatible with common sense or saying that common sense has a bit of knowledge to say about the discipline of microbiology. Since this comparison was made only on an abstract and general level, Bellolio 2019 finds that the important role of common-sense concepts, which aims to make these aspects understandable, has been largely neglected. To clarify this issue, Green find that the general comparison did not mention or determine whether leaving common sense axioms is required for the fragmented process by which scientific knowledge is accepted or learned. When all that is related to constructivism theories related to learning is highlighted and the literature quoted in the philosophy of science is that we must analyse both the case of the differences between common sense and scientific knowledge as well as analyse the change itself from pre-scientific thoughts and methods to the analysis of scientific methods and thoughts alike. As a result, reviewing and refining the analysis better highlights the common sense and soundness of scientific thinking that is appropriate to the context. While Hoyningen-Huene believes that achieving and increasing scientific-based methodology sometimes requires the sacrifice of common sense. In fact, empirical studies in science education confirm that common sense is already strongly present in scientific theories, although it should be separated from it. Despite the existence of common sense in the context of scientific theories, these scholars have suggested that our knowledge systems and common-sense concepts should be more flexible than before in order to accept scientific theories. This link between scientific theories and common sense raises a debate about whether scientific knowledge is systematized through a lesser degree of correlation between systematic scientific thinking and common sense. Therefore, this may raise a question about whether increasing systematic scientific knowledge is to be an essential issue or an end in itself. Hoyningen-Huene and others have added that the appropriate adoption of scientific methodology described in the theory of systematicity in analysis can fulfil the dimensions to provide a type of analysis that is more accurate, efficient and reliable. However, these distinctive features in this theory help to increase the gap between innovation and tradition and between innovation and disciplinary standards in various fields and disciplines. Consequently, Bellolio have made a suggestion about the existence of latent potentials related to systematicity theory that can be exploited in analysing texts in the context of their various disciplines by introducing students to the organizational aspects of the sciences of all kinds as well as how students are taking advantage of the growing curriculum and setting its priorities in all these areas. Someone might object to Bellolio’s suggestion that it conflicts with Hoyningen-Huene's aspiration to distinguish between science and everyday knowledge in general. Despite all that has been mentioned, this theory may not be very fruitful to determine the general difference between science and daily knowledge, but the importance of this theory lies in describing and defining how we can support systematic scientific thinking and achieve cooperation externally through activities and institutions based on scientific methodological foundations. John Rawls

(2005) has stated that when educational curricula and scientific conclusions are not controversial, nor do they develop the scientific thinking of students, they are merely thought of as general causes. However, several objections were raised against this statement. In Bellolio's paper, he highlighted only two of these objections. The first of these objections claims that the reasons mentioned here are scientific reasons related to the requirements and needs of the scientific community only and do not include other reasons that are related to society in a comprehensive and general way. The second reason for objecting to this saying is that the condition of difference is when the nature of these reasons resists the scientific assumptions accepted by the majority of the citizens. This research has provided some answers to these two objections. With regard to the first objection, that is the collective objection, the relevant test of a cause to be general is whether the causes have been interpreted according to the rules and the general framework of thought. In the sense that insofar as scientific methods and conclusions correspond to the principles of thought and rules of proof that liberals understand as general, their relational origin is secondary. As for the second objection, the criterion for a scientific argument that is considered non-controversial must be to indicate the degree of its consensus within science itself, because ordinary citizens accept or reject scientific statements that are conditional on their own comprehensive opinions. However, there will be a need for a broad agreement outside the bounds of science arena that includes the cognitive virtues of the scientific method. In the end, Bellolio's scientific paper concludes that there is a good position to think about scientific reasons as general causes to the point where it views scientific thinking as a research method that reflects the basic aspiration of Rawls' liberal political thought, which is summarized about the existence of a general style of thinking that depends on an impersonal point of view, which are to be separated from subjective claims. (Cristóbal Bellolio, 2019).

Bonotti (2019) has responded to the former view of Cristóbal Bellolio (2019), arguing that a lot of scientific conclusions are definite matters rather than being general ones, i.e., like those conclusions found in the situation of Vallier (2011, 2014). He added that usually those scientific conclusions are controversial topics to a greater or lesser degree in the scientific community. This raises the question that does this prevent such conclusions from being thought of these scientific conclusions as general causes? Bellolio has responded to this question by replying yes, and this answer has entailed a major problem because it contradicts with what was stated earlier by Rawls. Rawls (2005) mentioned that the public thinking contains only the non-controversial methods and conclusions of science. However, this scientific discourse indicates that we would stick to only very few general reasons that are based on science, bearing in mind the extent of scholarly disagreement on most scientific issues. Nevertheless, this matter does not have to be the norm. To escape such an undesired conclusion is to think about an aspect of public thinking that Bellolio appears to be neglecting, namely its structure. We find that there are many rational liberals who support notions of “consensus” of public thinking. Those scholars often refer to general causes as “common” causes among citizens of a liberal system; however, there is a different vision to comprehend the public thinking. In other words, it is the notion of "accessibility", literally the idea that for a reason to be public, citizens are only required to share the rationale behind it rather than the reasoning itself (Vallier 2011, 2014). The scientific method for climate sciences provides common foundations and principles of scientific thinking in order to allow scientists of various intellectual spectrums to reach different conclusions and then develop different models for different climate sciences without making any mistakes during the conduct of that process and this is what was demonstrated through the example mentioned above. The fundamental issue in this respect is that given the availability of a concept about accessibility to the public thinking, scientific conclusions can be considered general causes, even though they are a point of scientific disagreement. However, Bellolio seems to reject this concept, and Bonotti (2019) found it implicitly based on the idea of participation of the public thinking, which leads to a narrowing of the scope of scientific conclusions that are considered to be a controversial issue between the circles of the scientific community. Bellolio might avoid this dilemma if he thought about the structure of the public thinking. Another issue that should be touched upon is the normative effect of scientific conclusions. Bonotti found that the normative influence on the evolution of the structure of the public thinking can also help to explain and elucidate the second aspect of the argument on which Bellolio was based. However, Bellolio (2019) rejects the concept of "scientism" and refutes that scientific conclusion is not based on the normative framework and that the process must be accompanied by other general causes. This is a worthwhile claim, as it indicates a distinction that may be overlooked in the areas of addressing the issue of the public reasoning.

In a more precise and specific manner, Vallier (2016) has added that it is not only necessary to separate notions of participation from the accessibility of the structure of the public thinking, but it is also important to include the concept of accessibility to the public thinking and to distinguish between the different types of criteria of thinking that are formed through general causes. Vallier (2016, p.607) further has explained that standards that are "prescriptive and descriptive" are among the different types of criteria of thinking. Descriptive standards include criteria for scientific reasoning for a science, while mandatory standards include both epistemic guidelines for investigation. More importantly, mandatory standards include objective political values such as values of freedoms like religious freedoms and political freedoms exemplified in gender equality, which are widely shared among members of a liberal society that provide its affiliates with normative reasons (Badano & Bonotti, 2020). That is, scientific conclusions can rarely be achieved if they are directed towards achieving specific political goals. At the same time, one can argue in the context of realizing political values, one must use evidence and proofs of a nature based on scientific research. Therefore, it appears that the criterion for descriptive thinking is a relationship that includes both the criteria for critical, creative, innovative, scientific thinking and scientific thinking skills. We find that in this research the product moment correlation is used to find out the relationship between the two variables of creative-critical scientific thinking and scientific thinking, because the data is in the form of an interval or a ratio based on the calculation of the instantaneous correlation of the obtained product with the value of the correlation coefficient rxy = 0.57 with an average level of relationship, which indicates a positive relationship between scientific creativity and critical thinking with scientific reasoning skills. This clearly indicates that the students' scientific reasoning skills variable actually contributes 57% to the critical creative thinking skill variable, while the rest is affected by other factors. It is also possible to integrate different learning strategies into problem-based or project-based learning situations, such as literacy strategy, and various types of non-traditional tasks of writing, conceptual change text, authentic audiences text analogy, production of text and content, and others (Vallier, 2016).

We can conclude that the student's abilities to use critical and creative scientific thinking skills as well as scientific thinking skills in general are still at a low level, according to the results of the case studies that have been conducted by Mustika and others (2019). Among the most important findings of Mustika's study is that there is a direct relationship between creative critical thinking and scientific thinking skills. Therefore, it is imperative that teachers help students to develop and enhance those three important skills presented in this context: SR Skills, scientific creativity Skills and critical thinking skills, especially in the field of teaching physics (Mustika et al., 2019).

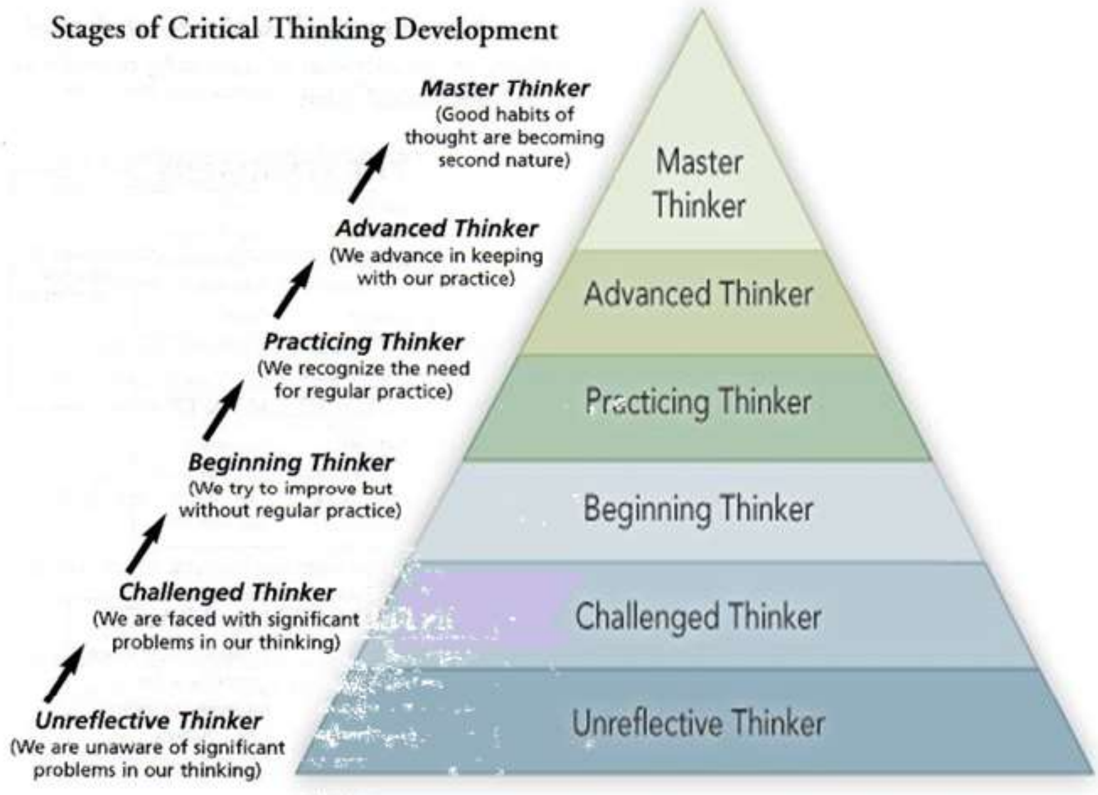


Figure 3: Development of thinking habits) Paul et al., (2006) p.50, Paul & Elder (2006) p. 285).

2.2.7 Science reasoning and teaching approaches

It is previously proved that the teaching approach has certain effect on the students' learning process (Alfaro-Lefevre, R., 2017; Choi et al., 2014; Prosser and Trigwell, 2014), and according to Moore and Rubbo, (2012) to accomplish considerable gains in scientific reasoning a reassessment of the traditional teaching approach to physics for non-STEM populations is required. Therefore, a suitable teaching approach to empower learners with scientific thinking skills, is to be aimed at in order to reach significant enhancement of SR for STEM and non-STEM populations. One of the main elements of a good quality teaching that enhance higher thinking skills is practicing problem-solving (Tanişli & Dur, 2018). As indicated in Clark and Taylor (1993), Many researchers believe that the problem-solving process depends largely on two main variables, the first of which is the extent of experience that we have in a particular area of the profession, and the use of that experience in facing the problems of this field. Thus, we possess the competencies and capabilities necessary to apply them in the face of such problems and thus be able to solve them. The second variable parameter is the provision of a domain-specific set of experiences, knowledge, and skills in solving specific problem categories.

Characteristics of a good problem-solving teaching have been expressed in a study conducted by Taconis and others (2001) which concluded that individual and even group work problem solving, give no considerable positive effects unless accompanied with guidance and immediate feedback from the instructor (Taconis et al., 2001).

The classroom environment also had implications on the problem-solving activities done by students (Kartal,T &Kartal, B, 2019). Classroom environment should be safe for students to ask question and make suggestions and arguments. Students then will be encouraged to solve even the ill-defined problems which are closer to everyday life problems. The teacher in that classroom of problem-based teaching will be a facilitator and students try to solve the problems in small groups (Kartal,T & Kartal, B, 2019).

Teaching implications were also discussed for its relation to SR. For example, Kartal, T and Kartal, B (2019) have criticized pre-service teachers' practices when they often appeal to use the mathematical approach in solving problems. This approach is inappropriate for students thinking because the students do not use proportional logic in solving problems. Secondly, students consider proportional logic to be just a subject whose use is limited to mathematical operations. Thirdly, these students do not see any connection or association between mathematics and physics. In addition, these students still have great confidence in using the old traditional methods such as algorithms. Kartal, T. and Kartal, B. also has reached important conclusions that teachers in the pre-service stage are not aware of the conceptual understanding in stoichiometry, which hinders the performance of problem solving, regardless of the strategies they use in solving problems, and accordingly, specialists and teaching professionals must address these problems by providing awareness sessions to these novice and pre-service teachers on the definitions and the complete concepts behind stoichiometry and explaining how carrying out proportional reasoning on issues of stoichiometry prior to teaching algorithms, as well as encouraging and urging these teachers to practice proportional logic, which does not include the scientific concepts that could help these pre-service teachers understand the ratio and link proportional thinking to stoichiometry. In a similar study by Cramer and Post (1993) assert that the qualitative comparison conducted without numerical quantities serves to enhance the conceptual understanding of proportional reasoning. Thus, it will be a good opportunity to deal with a qualitative comparison as a fruitful way to encourage pre-service teachers to use other strategies, other than the traditional stereotypical approach to algorithms only (Kartal,T &Kartal, B, 2019).

Learners' reflection also, plays a key and vital role in developing their performance in the learning environment. Accordingly, a study by Chen and others (2011) has sought to investigate the possibility of developing learners' reflection levels by adjusting teaching strategies in a way that is compatible and appropriate to the ways of thinking of learners in the learning environment that is based on the Internet. Three instructional strategies are actually designed: constructive, directive and inductive. These three strategies are designed in harmony with legislative, executive, and judicial thinking styles. Subsequently, an online reflection learning system was developed to reflect performance in the learning environment. In order to ascertain the suitability of teaching strategies with the thinking patterns of learners, an experiment was conducted that classified learners into an appropriate or inappropriate group with the aim of analyzing the compatibility and suitability of teaching strategies with students ’thinking styles. 223 male and female students from university levels participated in the experiment. The results of that experiment have shown that the reflection levels of the appropriate group of students outperformed the inappropriate group (Chen et al., 2011).

Stephens & Clement (2010) have conducted their study to suggest a set of applicable combinations and to identify informal thinking processes that take place during students' classroom discussions, and to demonstrate how these processes can be separated from each other and at the same time work on linking them with image indicators. That specifically meant for them that the conceptual distinctions and definitions that Stephens & Clement (2010) have developed allow them to use the transcripts recorded from the classroom video tapes, with the aim of identifying students' use of several categories that are limited to informal thinking processes and that are also used by experts. Specifically, Stephens & Clement studied analogies, extreme cases and Gedanken experiments. It was also possible to define these three processes even when these processes are used in combination with one another. The second issue is that some cases have been created and run collaboratively between several students or between teachers and students. Third, Stephens and Clement found that all the cases of the processes listed in the first and second tables were created automatically by the students, and this occurs during discussions regarding the important issues related to the concepts. On the other hand, for teachers interested in the goals of the scientific process, Stephens and Clement research represents a good opportunity for students to participate in the creative scientific thinking processes in relation to the discussions that take place in the classroom. In addition, evidence can be gathered from video tapes recorded in a particular classroom that indicated that students could use mental images when practicing these three types of scientific thinking. The three types of gestures have also been identified, which are the shape that points, indicates movement, and indicates strength. This distinction is made in the context of identifying evidence for students' use of mental images and animated graphics with kinetic components in a successive manner. In sum, this study is useful in explaining and clarifying the methods used in this study, which provide a set of tools that can be used to analyze informal thinking processes during classroom discussions. The study also suggested that pictures play a role in learning the processes of scientific thinking. This was observed in physics classrooms, which aroused the interest of some researchers to conduct a study on this topic in the future. These indicators have been elaborated to the point of believing that there is a correlation between drawing and building mental images and scientific thinking in most of the students ’classrooms, as these studies have shown that images play a pivotal role in the development and support of scientific thinking processes (Stephens & Clement, 2010). Evidence was shown that imaginary simulations occur during classroom study. However, the matters that represented a real challenge for researchers during Stephens and Clement study have revolved around identifying, criticizing, redefining, and refining the central concepts related to the three types of scientific thinking, namely, analogies, extreme cases and Gedanken experiments and identifying types of gestures as pointers for images. Finally, Stephen and Clement hoped that their research will contribute to the work and conduct of more research to study the nature of students' thinking processes during their attendance and participation in the classroom activities (Stephens & Clement, 2010).

.

Explore- instruct teaching approach, is one of the problem-based inquiry learning approaches (Bhat, 2019; Ramful & Ho, 2015; Tanişli & Dur, 2018). It is one of the approaches that provide knowledge application activity for the students before the explicit instruction (Loehr et al., 2014). And according to Loehr and others (2014) this teaching approach effectively support learning, provided that learners have the opportunity to apply the learned knowledge in an activity after the instruction.

2.2.7.1 Aspects of Scientific Reasoning.

Novia and Riandi (2017) have discussed the facets of scientific thinking and its proportions in students. They have found that the proportional thinking represents the highest, while the conservation of weight facet is the lowest one. Considering the pre and post-tests, the proportional test has received the highest percentage in the pre and post-tests, while the facet of weight conservation was found to be the lowest aspect of scientific thinking also in both pre and post-tests. These findings can be attributed to the stages of inquiry levels that have been trained on, as well as to relative inferences. In fact, the proportional thinking facet represents an interactive and demonstrative sphere, in which teachers create, develop, and ask questions to students in order to give feedback and answers to those questions, while giving explanations to help students arrive at logical conclusions provided by solid and concrete evidence. Students at this stage begin to play a real role based on real evidence that is guided by proper demonstrations and explanations from their teachers (Novia &Riandi, 2017).

Novia and Riandi continued to indicate that the aspects of scientific thinking should be introduced and trained in the early stages of the educational process, which is in contrast to the aspect of weight conservation. In this early stage, learning takes place by means of discovery. The basic idea of learning by discovery is based on the principle of establishing an inductive concept or building knowledgebase that is acquired from students' real-life experiences, which are different experiences. At this stage, students may face difficulties and challenges in determining the relationship between the given phenomena and their experiences, because the phenomena do not match the students' experiences. In order for students to acquire and develop scientific thinking skills, it is necessary for them to practice this type of scientific thinking, perseverance and patience, because scientific thinking in itself represents a wide range of cognitive skills. Accordingly, there is an increasing role on the shoulders of researchers and teachers to provide students with how to learn and develop scientific thinking skills. This can be shown, given that the LKS percentage was found to be slightly higher than 75%, even at the third meeting, and this percentage has not yet reached 80 to 90%. It can also be noticed that the development of students' scientific thinking was not analyzed to the fullest extent at the end of the meeting as the post-test stage (Novia &Riandi, 2017). These data are consistent with what Lawson (1978) has stated, where he indicated that intelligence is not an innate gift, but rather an evolutionary work that takes place gradually over time, requiring a great deal of effort and persistence, and therefore we must know how to develop students' abilities and skills in scientific thinking process.

2.2.7.2 Scientific Reasoning and Teaching Activities

As per many calls all over the world for a STEM teaching that enhance scientific thinking as well as content knowledge, lots of studies were proposed to explore teaching activities which impact scientific reasoning. One of these studies is the research of Novia and Riandi (2017). When the first classroom meeting was held in Novia and Riandi' study, the basic visions and ideas revolving around the concept of motion were presented and developed through the levels of the inquiry model under the framework of the laboratories, with the aim of building scientific thinking system for students that should be completely independent. In the meeting, the scientific reasoning put forward about the conservation weight and the proportional aspects of the score was 8.7 which has been trained through learning by using the means of interactive and demonstrative presentation and discovery. However, this result went down in the second meeting, but that decline was not important because the result increased again in the third meeting. However, various other challenges and problems have arisen due specifically to the issue of lack of sufficient time to examine the developments which often take a long time to take place. Development process is usually repeated several times to determine the extent of development in learning process because the process of development in learning will not occur automatically, but rather has specific requirements and instructions (Novia &

Riandi, 2017). Researchers in Novia and Riandi' study usually meet only three times to search and discuss the issue of data collection, depending on the time agreement provided by the school, and the time was not sufficient to present and discuss the whole visions and ideas of the development of scientific thinking, which normally requires discussion over and over again. In addition to the problem of limitation of time, there is another problem related to the process of describing the development of scientific thinking. This description is not an easy task because each student has come from a different educational background and learned in different teaching methods (Novia &Riandi, 2017). This view is in line with the vision of Jing Han (2013), when he mentioned that the development of the subject of learning and acquiring scientific thinking skills requires more work and effort because it is not an easy process. This is because each student learns in a different way. However, the common ground among the developments of a learning process is that learning is a choice of method based on research evidence (Jing Han, 2013).

There are a number of conclusions reached based on the discussions and results of the data analysis of Novia and Riandi,(2017) . Among those important conclusions, in this context, is that there is an increase in quantitative terms in the issue of scientific thinking, even though its degree is not high. The reason for the increase in the percentage of scientific thinking may be attributed to the students 'knowledge of the learning style based on the characteristic of discovery, and through that method of learning, students began to build their knowledge using the inquiry model. According to these results, the author has made several suggestions to the teachers. Among those important proposals is that teachers should use the Inquiry Levels Model as an alternative to training scientific thinking in order to help students build their knowledge base independently. As for what the author has recommended to other researchers, it is better for them to have the scientific thinking test tool consisting of only four questions from each side in order to ascertain the degree of its consistency. Also, among the recommendations contained in this regard is that the process of developing scientific thinking is not an easy matter and requires more time to acquire and develop it. Thus, the more practice, the more scientific thinking develops (Novia &Riandi, 2017).

A study by van der Graaf et al., (2019)has emphasized that lessons based on the principles of developing inquiries can improve and reinforce scientific thinking in children despite the availability of empirical evidence showing that inquiry-based learning requires some instructions. Generally, this research is based on two basic approaches, although they have not yet been bridged, namely, direct guidance of scientific thinking and training of teachers in verbal support.van der Graaf and colleagues (2019) have studied how to develop and enhance children's scientific thinking through these two types of instructions separately or in combination. This work was done via conducting a comparison of four conditions, namely baseline, direct instruction, verbal support, and finally a common approach. The success and effectiveness of the existing series of lessons were studied along the lines of the inquiry style, which focuses on developing scientific thinking abilities, vocabulary, and knowledge of the field. The test was carried out on a sample around 301 fourth grade school children. The test results have shown that both approaches have enhanced different components of scientific thinking abilities and skills. The study also has illustrated that the instruction set was more effective in developing scientific thinking skills and abilities, vocabulary, and field-specific knowledge. The acquisition of knowledge specific to the field was developed and enhanced when both directions were provided. As a result, it can be concluded that each type of teaching has its own qualitative and special contributions to science learning for children and that these instructions reinforce and complement each other. Thus, (van der Graaf et al., 2019) study has clearly demonstrated that a series of lessons based on the style and principle of inquiry, especially when it is preceded by direct instructions for scientific thinking and supported by verbal support, is most effective. In line with this study is a claim by Healey (2000) that the 21st century students are changing and require instant answers and feedback, by their own investigation skills, not through lecturing.

A study that has been conducted in Malaysia by Tajudin and others(2012), demonstrated that most higher education Malaysian students had concrete-operational level of SR abilities where the general mean was 3.23 out of 12. According to Lawson (2004) acquiring 0 –4 points in LCTSR, means being in the empirical-inductive level (concrete-operational thinkers);acquiring 5 –8 points mean being in the transitional level (transitional thinkers);and acquiring 9 –12 points—hypothetical-deductive level (formal-operational thinkers);indicating that Lawson’s transitional level most probably fits Piaget’s Early Formal thinking and Lawson’s ‘hypothetical-deductive’ reasoning fits Piaget’s Mature Formal thinking. Tajudin and others (2012) study also demonstrated that, the expert and delegator were predominant instructors' teaching styles as indicated by undergraduates' perception. Likewise, there was no relationship between instructors' teaching style and the level of SR aptitudes. Consequently, this examination can't correlate the dominant teachers' instruction style to the level of SR aptitudes of Mathematics, Science and Engineering students in Malaysian public institute of higher education. In a way, this examination gave a few signs that the delegator and expert teaching styles were not of value to the development of undergraduates' thinking aptitudes (Tajudin et al., 2012). In the opposite side and according to Lawson (2004), effective learning mirrors the practice of science, in which students face puzzling observations and then involve in the interpretation production and testing process (Lawson, 2004), and inquiry instruction is a form of this active learning as per (National Research Centre, 2000).When teaching for SR; two things must be known by teachers, the intellectual level of their students (i.e., preoperational, concrete operational, formal, or post-formal), and what prior subject-specific knowledge do they have? (Lawson, 2004). It is necessary then to explicitly instruct for (SR) and plan well the taught topics sequencing. These two steps were found By Bao and others to be two factors affect the ability of the students to reason scientifically (Bao et al., 2009).

Jensen and others (2017) have illustrated that newly admitted students to attend science subject courses appear to be relatively inconsistent in their thinking behavior compared to senior students with regard to the use of systemic thinking process skills in a regular and comprehensive manner, and therefore teachers of mathematics, chemistry, physics, engineering and technology should address this problem and encourage these students to activate their thinking reasoning in order not to drop out and leave the classrooms of these vital sciences. Therefore, it is necessary to address this problem by the teachers of these subjects to develop effective solutions. In this regard, some proposals have been made, including that teaching these subjects should focus on raising thinking process and reasoning skills during the first year of study, and at the same time maintaining students' grades in STEM, which may motivate students and give them a higher chance of success and achieving satisfaction, and thus reduce the possibility of leaving such classrooms. Furthermore, the teachers of these subjects should facilitate the process of scientific investigation for those students, who may benefit the students and make them identify the variables and control them, and then set hypotheses, reach inferences while collecting and analyzing data and drawing conclusions. Given the importance of early development of scientific reasoning skills among students, we find the necessary support and encouragement to reinforce scientific reasoning among students at early stages of education, which is basically stated on the framework of Vision and Change. This trend and call are also encouraged among college students (Jensen et al., 2017).

Marušić & Sliško, (2012) published out a study that examined the impact of three different physics teaching approaches on the gain of scientific reasoning of senior high school students. They found that Experience-discuss teaching approach caused the highest gain in SR among the three approaches which included also Read, Present & Question approach and the traditional lecturing approach. And despite positive trends between the critical thinking of students and Problem-based learning (PBL); outcomes of PBL were not significantly different from traditional lecture in a small study done by Choi, Lindquist & Song (2014), in two South Korean nursing colleges for the first-year students. And larger studies were recommended by them to study the effects of PBL on critical student abilities. In the opposite side, it was suggested by Fyfe, DeCaro, and Rittle-Johnson (2014) that, in spite of arguments in favour of delaying conceptual instruction, sometimes conceptual instruction should come prior to problem solving.

Several scholars and educators who are interested in the need to promote scientific thinking, have indicated that the skills of statistical and causal reasoning are considered more urgent than other skills in dealing with science learning programs (Koenig et al., 2019). From a practical scientific perspective, the acquisition and development of statistical thinking helps to enable students to collect and analyze data, and to achieve quantitative assessments regarding possible relationships between variables, and thus reaching important results and relationships in promoting causal reasoning. In addition, developing the skill of causal thinking is a fundamental issue for testing the hypotheses that are initially set and identified, and then making logical decisions based on concrete evidence and reconciling theory with reality. As a result of the significance of these two skills, at the present time, Koenig and others (2019) embark on studying and examining the learning outcomes, assessment, and the laboratory activities related to these two skills of statistical and causal thinking, at the same time that Koenig and others are working on improving the environment of the laboratories in which they conduct the experiments.

There are also many laboratory curricula available at universities that provide excellent inquiry-based laboratory experiences, but most of their work focuses on developing basic laboratory skills and / or promoting deep conceptual science-based thinking framework. As for Koenig and others' curricula, their courses have been developed from the previous ones in that they have become more determined to support and develop the skills and patterns of scientific reasoning that benefit students to acquire systematic scientific methods and patterns of thinking that are fruitful for them, especially with regard to the sub-skills that have been practically identified, such as SR skills. This type of scientific reasoning is essential and required in the empirical courses and transferable via the STEM system. Koenig and colleagues also hope that a number of other academic institutions will follow the example of their institution in terms of using the advantages and methods of the scientific thinking approach that is illustrated in their study, in addition to learning and developing sub-skills for COV, which are also described in their article. These genuine efforts are all exerted with the aim of helping and demanding each institution to review its curricula in the domain of science subjects, and then attempt to design teaching science programs that contribute effectively and practically to raising the capabilities of their students and enhancing their scientific reasoning skills (Koenig et al., 2019).

2.2.8 Scientific Reasoning Assessment

A major factor that was investigated for its positive impact on SR is the nature of the assessment tool, that will ultimately push the instructor to design several learning activities that enhance these scientific reasoning skills. The thinking ability of children was classified by Piaget in 1964 in which he divided children's thinking abilities into specific four main stages. These stages are named as follows: the sensory-kinaesthetic stage (from 0 to 2 years), the operational stage (from 2 to 7 years), the concrete operational stage (from 7 to 11 years), and finally the formal operational stage (from 11 years and above). Then he has specified that the highest rate of scientific reasoning is formal operational(Piaget, 1964)

2.2.8.1 SR Assessments' construction

Jensen and others (2014) conducted a study in which they have urged and emphasized the utmost importance of the need for a high level of professional evaluation in performance by science subject teachers to develop and strengthen the centralization of the concept of employing scientific reasoning in the context of teaching science subjects. In this regard, these teachers must be urged and encouraged to play their major role precisely in determining the desired learning outcomes first for each subject by using the backward design. For example, biology teachers must make sure that their students are fully aware of the significance of absolute knowledge of the content, and they are at the same time required to develop and strengthened practical scientific process skills when dealing with these types of subjects. It follows that in the backward design, the necessity to articulate clearly and indicate that their students must achieve the learning outcomes and design the necessary assessments according to those outputs. This step is followed by the development and design of appropriate activities that are compatible and aligned with the assessments related to the desired learning outcomes. And through these thoroughly studied procedures and steps, we can offer and help the students of these science subjects more opportunities to translate what they have learned to solve the problems they face in the classroom or outside it, by effectively using logical scientific processes and reasoning skills. However, in many situations, science teachers fail to develop and design learning outcomes and their associated assessments regarding the acquisition and promotion of the necessary scientific processes and reasoning skills from a broad practical perspective, and the assessment is often limited only to testing the student's knowledge of the course content by using the memorization process only, whereas, these tests should have been designed to assess the student's development in the use of skills that sharpen scientific thinking and which are available at the highest levels within the framework of the Bloomfield taxonomy. Accordingly, it must be emphasized that the focus on achieving learning outcomes through the teaching of these subjects must revolve mainly around directing students and urging them to develop scientific reasoning from a practical perspective, i.e. developing their skills inside and outside the academic context and enhancing their intellectual abilities to solve real problems that may encounter them in real life, without focusing solely on the issue of memorization, which serves to discourage the value of employing fruitful and productive scientific reasoning (Jensen et al., 2014).

It has been indicated by Kind and Osborne (2017) that the prominent and growing role that classification, i.e., taxonomy, plays in developing all the different fields of science represents the organizational framework for all the different science curricula. For example, this classification identifies all forms and patterns of life related to the subject matter of biology, classifies stones and rocks in relation to the subject of Earth sciences, that is, what is known as the foundations of geology, as well as determines the basic elements for the development of the periodic table, and distinguishes between heat and temperature as an important and basic way of thinking for evolution. In science, classifying patterns related to epidemiology, with the aim of establishing possible causal relationships for the occurrence of diseases and epidemics. As for school science, classification can be done by studying historical examples, or alternatively those historical examples can be modelled by using exercises that help to explore relationships, for example between eye colour and hair colour, reaction time and its relationship to gender, height, or foot size. Likewise, Kind and Osborne showed that researchers explained that, the emergence of probabilistic thinking in the seventeenth century has become a way of thinking, which is an imperative matter of science and is considered a way of thinking that can be represented by collecting a set of data that produces Gaussian distributions such as the variation in the heights of a group of students of the same age and using Punnett squares for predicting the results of hybridization, whether monohybrid or di-hybrid, as well as in identifying higher levels in statistical and quantitative mechanics(Kind & Osborne, 2017). Even teaching goals in the educational sciences as per Armstrong (2018) and Krathwohl (2002) are good to be classified, that classification was accomplished by Benjamin Bloom Since 1956, Benjamin Bloom's taxonomy has been working effectively in shaping educational plans. The taxonomy is a framework for categorizing educational goals: Taxonomy of Educational Objectives. Familiarly known as Bloom's Taxonomy, this framework has been applied by generations of K-12 teachers and college instructors in their teaching. This taxonomy consists of lower level of thinking: Knowledge, Comprehension, Application, and higher level of thinking: Analysis, Synthesis, and Evaluation. The categories after Knowledge were presented as "skills and abilities," with the insight that knowledge was essential precondition for putting these skills and abilities into practice. So, according to Bloom's Taxonomy, human thinking skills can be broken down into the following six categories: *Knowledge* which is fact recalling, *Comprehension*: which is getting the meaning of informational materials, *Application*: that is applying knowledge to new situations. *Analysis:* is dividing the information into smaller portions to examine it. *Synthesis*: is the application of the learned knowledge to combine parts into a new pattern which was not there before. *Evaluation:* is the judging according to some set of standards, without the existence of wrong or right answers. (Armstrong 2018, Krathwohl, 2002).

Krathwohl (2002) reviewed the previous taxonomy and indicated that the issue of investigation and post-review of classification directed to change the order and naming of cognitive processes in the original version. The revised Bloom's levels became remembering, understanding, creating, analysing, applying, and evaluating. The issue of reorganizing the taxonomy focuses on developing the skill of synthesis in the level at the top of the hierarchy, rather than focusing on the evaluation. Most important of all, Krathwohl and others found that the review of the taxonomy is very useful because it represents a real addition to all the six cognitive processes. In addition, this review identifies the four types of cognitive processes that can be addressed by Learning activity (Appendix IX). As mentioned earlier, these four types of knowledge are represented by realistic knowledge that is related to terms and separate facts, conceptual knowledge that includes categories, theories, principles and models, procedural knowledge which is a technical knowledge or also known as practical or methodological knowledge, and finally what is so-called metacognitive. This type of knowledge includes the abilities related to knowledge of various learning skills and techniques as well as self-assessment (Krathwohl, 2002).

Interestingly, Kind and Osborne stated that, the whole attention of science is centred on providing a convincing and satisfactory answer to the causal question, "why does the world behave in the manner required by the hypothetical structure within the framework of analogue models and in the manner of the pattern of thinking prevalent in science". For instance, students are required to learn about the physical models and structures of the human body and then they are required to create physical models for cells. They are also required to use particle models for the material in order to explain some phenomena such as condensation, evaporation and refraction phenomena. The atom is also represented by presenting the Bohr model as one of the types of miniature solar system that is made by using a wavelength model of light. Other types of molecular models of chemical structures are also used to explain ionic and covalent bonding, the nature of chemical reactions and the properties of substances. Despite the importance of presenting, explaining and experimenting with all these physical and chemical models as tools for studying scientific projects and the need to use scientific reasoning as an effective and powerful tool to solve the problems surrounding these scientific projects, researchers noticed that this proposition is not openly discussed in the context of science education (Kind & Osborn, 2017).

Sternberg (2020) has emphasized that there are many countries in the world in which standardized test scores are required, adopted and used in order to accept students who wish to study in science-based programs. It is preferable that these tests measure the basic skills that should be met by students who intend to study science subjects. In fact, this is done in a non-essential way, compared to measuring the skills of scientific creativity and the recognition of scientific influence. In three major scientific studies, Sternberg (2020) conducted an investigation on the role of scientific creativity and the perception of the scientific impact on scientific reasoning. To conduct these three scientific studies, a university in the north-eastern United States of America was intentionally selected, and 219 students were being tested. These student respondents received assessments related specifically to scientific creativity and recognition of scientific influence. In addition, a number of various questions have been placed, among the evaluations that were used previously, with the aim of measuring the issue of scientific inference, where alternative hypotheses were developed, experiments generated, and conclusions drawn. The flexibility aspect of general intelligence that includes letter combinations and number series was also evaluated. These students were also required to sit and present scores on either or both the SAT and ACT tests in order to be admitted to the university or college, taking into account some demographic information. The main objective in Sternberg (2020) study was to identify and articulate clearly the relationship between the new tests that were conducted regarding the scientific influence and creativity and the extent of their relevance to each of the scientific reasoning tests and fluid intelligence, by studying them separately or both together. Although the result of the global structure of the test on the issue of determining the scientific influence was less clear compared to the test of scientific creativity, Sternberg (2020) found that the new measures have greatly benefited from the aspects of scientific thinking, which they have previously studied. They also found that those who participated in the test rated the studies as having high-impact and that they were more rigorous from a scientific perspective and were of very great benefit in practice, but they were considered to be as low-impact studies. In general, these studies were criticized as being less innovative and creative, and this may be due to the fact that the titles and academic abstracts that were produced had low levels of creativity and innovation by the participants in terms of their performance. The results repeated through these studies also included the existence of a direct relationship between the groups of letters and the number series, on all levels and measures of fluid intelligence, as well as the existence of this direct relationship between scientific reasoning and scientific creativity. (Sternberg et al., 2020).

According to Erlina (2018), the elements of scientific reasoning consist of a set of key components. These components can be identified in several ways: First, the variable control (CV), which specifies the types of independent variables and their dependent variables; Second, relative thinking (PPT), which is the determination of the relationship between variables by using mathematical equations, numbers, tables, and graphs; Third, it is known as Probabilistic Thinking (PBT), which is what predicts the resultant opportunity that has been gained when replicating. The fourth type of the components of scientific thinking is known as hypothetical deductive reasoning (HDR), and it synthesizes a hypothesis based on a general concept related to a specific concept. As for the fifth and final type of the components of the elements of scientific thinking, it is known as associative or correlational thinking (CT). This type of thinking is done by establishing reciprocal relationships, which may be interdependent or unrelated, within the framework of variables. Referring to Bloom's revised classifications, scientific thinking and its maturation process according to education researchers, must be viewed through pretesting, based on mastery of important capabilities related to understanding, analysis, remembering, application, evaluation, and creativity (Erlina et al., 2018). In fact, Kind and Osborne found that in the previous period, specifically in the last six decades, a great qualitative development occurred in human perception regarding the types of real cognitive levels that people need in their various fields of life. For example, researchers mentioned by Kind & Osborne have classified four classes of cognitive types, not just one. They classified these types of cognitive thoughts as follows: conceptual, factual, and procedural cognitive, and the last one known as metacognitive. In spite of that classification, no clear vision was developed and presented in order to identify the advantages and merits of the three cognitive types related to each of the content area, the procedural, and the cognitive framework. These areas are fundamental and vital in building, developing and learning correct scientific reasoning (Kind & Osborne, 2017).

Effendy and colleagues (2018) showed in their study that is about the ability of scientific reasoning and mastery of physics concept, that when students reach the cognitive level of C1, which is the highest cognitive level at 87.01%, its students are included in the studied class (cognitive levels and its symbols are demonstrated in table 3). At this level, the students are able to exercise good scientific reasoning in dealing with problems and thus become capable to solve problems correctly. However, at this level of cognitive mastery, these students face the problem of memorization. On the other hand, Effendy and colleagues found that the lowest level of intellectual and cognitive perception takes place at the cognitive level C4, which is 41.99%. This cognitive level is included in the being category. At this level, most students face difficulties and problems in the use of scientific analysis in order to reach logical and correct conclusions and only a small percentage of them are able to solve the problems correctly. It all depends according to Effendy and colleagues, on students 'ability and skills to use good scientific reasoning. This clearly proves the importance and ability of mature scientific reasoning in solving problems through employing analytical thinking to reach appropriate and correct conclusions, especially in the study of physics because it relates to a number of interlocking concepts with each other. Effendy and colleagues (2018) have stated that, the total ability to master a student's concept is reached at number 60.38% and at this level it is included in the medium category. In fact, there are different levels of cognitive knowledge. Students can accurately answer questions within the framework of the cognitive level C1 that are remembered, C2 which is the nature of understanding and C3 which is its application. While Effendy and others (2018) find in the C4 cognitive level questions that are analysed, C5 that assesses and C6 creates, students attempt or strive to get into trouble but then they encounter a real difficulty in dealing with and answering the problems. This occurs because in the learning process students seldom encounter something contextually related to the subject being actually studied. Teachers are usually directed towards enabling students to acquire mathematical abilities only, but when students encounter problems that require honing analytical skills, in order to build and evaluate them, they find difficulty to solve such problems (Effendy et al., 2018).

Table 3: Mastery of Concept Indicator retrieved from (Effendy et al., 2018).

Cognitive Level	Aspect of Mastery	Concepts Indicator
C1	Remember	a. Recognize b. Recalling
C2	Understand	a. Interpret b. Exemplify c. Classify d. Summarize e. Conclude f. Comparing g. Explain
C3	Apply	a. Execute b. Implementing
C4	Analyse	a. Distinguish b. Organize c. Attribute
C5	Evaluate	a. Check b. Criticize
C6	Create	a. Formulate b. Plan a. Produce

2.2.8.2 Examples of SR Assessment instruments

Kim and others (2014) discussed that, when different types of skills are taught, as in the case of physics or any scientific or experimental course, various measures should be utilized to assess all the learning dimensions. So, both traditional tests, and non-traditional tests (i.e., Performance Based Assessments) should be prepared by instructors to assess the students' learning.

In 2013 Jing Han studied in a PhD thesis scientific reasoning development and assessment. Han conducted in his research a series of studies, one of those studies studied Lawson Classroom Scientific Reasoning Test (LCTSR) and its utilization in SR assessment. Han indicated that LCTSR is the most widely used test in SR assessment. But the test's validity was not thoroughly investigated. Han showed few issues with the design of the current version of LCTSR such as, the choices of question 8 as per Han, which seems to include two correct answers, the thing that resulted in a low scoring of many students in on this question, in addition to similar concerns in some other items of the test. These issues caused according to Han, a low ceiling even for good- developed students. In spite of the concerns about LCTSR, the wide utilization of it provided a huge pool of data that provide a rich source for researchers to compare their studied students' skills with the test outcomes. Han results showed that, studying individual dimensions of SR are better in reflecting in depth picture of SR skills of the studied sample. For example, Han found that, on the Lawson's test, the Chinese and American students are comparable in their total SR scores, but they differ in five skill dimensions out of the six. The actual reasons for these differences were under investigation; but initial indicators suggest according to Han, that educational and cultural settings of the two countries contribute a lot to these differences. Han's analysis also provided numerical system of measurement for the levels of difficulty of the LCTSR, and the participant's skill dimensions. Han's results showed that LCTSR is optimum for high school students' assessment.

In a study by Yuksel (2019), a comparison happened to explain why LCTSR is to be utilized to assess SR and not Scientific Reasoning Skills Test (SRST) which was developed by Yuksel (2015). The study showed that Lawson's transitional level is more likely to relate to Piaget's previous official level of 'speculative deductive' than to Piaget's mature formal level. If a comparison of the Scientific Reasoning Skills Test (SRST) developed by Yuksel (2015) is made with the LCTSR test which has been utilized as the main instrument in the current study, then the Cronbach Alpha reliability factor for the Scientific Reasoning Test (SRST) that Yuksel (2015) developed, found to be 0.76. In fact, the Scientific Reasoning Skills Test (SRST) consists of a total of seven sub-dimensions, and the memorization laws and the six dimensions expected to be included in an individual must be understood during the formal operational stage. The questions are presented in the SRST sub-dimensions as shown in table 4.

Table 4: SRST Sub-Dimensions Questions

SR sub-dimension	Question number
Conservation Laws	1, 2, 3
Proportional Thinking	4, 5, 6, 7
Identifying and Controlling Variables	8, 9, 10, 11
Combinational Thinking	12, 13, 14, 15
Correlational Thinking	16, 17, 18, 19
Probabilistic Thinking	20, 21, 22, 23
Hypothetical Thinking	24, 25, 26

Looking at things all together, the results by Schiefer and others (2019) have shown that, through written tests, it is possible to assess primary school children's understanding of SIC by developing and implementing the SIC test in an objective manner. More importantly, the results obtained from Schiefer, and others' study have also helped in providing and proving the preliminary evidence on the correctness of the construction of the test scores obtained by these students. It is also possible to expand the package of tools that are used to measure scientific reasoning for this age group of students. These results have been supported by other recent studies, for example the two studies conducted by Mayer and Koerber in 2014 and 2015, which have both provided more empirical evidence that primary school children have the abilities and skills necessary to use scientific reasoning to solve problems. In addition, the SIC test can be conducted with the aim of evaluating the essence of scientific reasoning and then applying it in future research and studies in order to evaluate the competencies of scientific reasoning among primary school children or measure educational progress in the field of science learning, for example, but not limited to, within the evaluation of scientific interventions for students with high capacities or taking school curriculum. (Schiefer et al., 2019)

In our current time, Wati and others (2019) found that the issue of science education does not focus only on quantity but goes beyond that to sharpen students' motive and urge them to develop scientific thinking in order to solve the problems they face in case studies. As a result, the main objective behind Wati and colleagues' study is to identify the abilities and skills of students within the context of using scientific thinking in solving problems related to light science courses. Thus, the method employed in that study is the descriptive approach. Survey tools were also used, and the instrument used in the description is a light material consisting of 8 elements. The test was conducted on a random sample of 201 elementary school students, who are the eighth-grade students in Banjarmasin. To analyse the data and process it into recorded numbers, the RASCH model was used, and the win step program was also employed to provide information on fit, dress and unilateralism. This study by Wati and colleagues has achieved its intended goal through the quality of the elements that are represented in measuring the item and the person, the item bias, the reliability of the item and the person, and a variable map. The results of Wati and others' study have shown that the abilities of the students (studied from the random sample) of scientific reasoning is low. Therefore, those interested in this matter, including officials, professors, and researchers, should strive to raise and develop the scientific reasoning capabilities of students in the future (Wati et al., 2019).

CHAPTER 3

METHODOLOGY

This research study is a field experiment that has been conducted in a college campus with real undergraduates, using a mixed method approach. Johnson & Christensen (2014) stated in their book that the factors that determine the appropriate mixture are the research questions and the practical issues and situations that face the researcher(P.33).So, the researcher as an educator wanted to quantitatively answer the first main question in this study, and then reflect on some of the students' perspectives that she often hears from students after sessions or read it written by students in the course evaluation.

Quantitatively, the study tried to find a common pattern of the impact of a change in the teaching approach (Independent variable) in the scientific reasoning ability (dependent variable) of sample of students studying introductory physics course. The advantage of the field experiment is that it shows if certain manipulation can work in real life setting, but extraneous factors should be controlled (Johnson & Christensen, 2014). The necessity for explicit instruction of (SR) and the taught topics sequencing are two factors found to affect the ability of the students to reason scientifically(Bao et al., 2009), and these factors were considered in the setting of this study. The introductory physics class was selected because the subject is logically sophisticated and is emphasized commonly in science education (Lawson et al., 2000). The resulting pattern can then be generalized as an outcome of a quasi-study, with purposeful selection and randomly assigned sampling.

According to Johnson and Christensen (2014), purposeful sampling and random assignment can be used to design a quasi-experimental research and this design is weaker than the experimental design which is the strongest possible design (random sampling, random assignment).But two-groups post-test-pre-test quasi-experimental research design which is followed in this study is stronger than the week experimental research designs, such as the one-group post-test-only design or one-group post-test-pre-test design or the post-test-only with non-equivalent groups design. In the last-mentioned week design, the non-equivalent word indicates that the two groups were not equal in all variables (demographical, cognitive ability…etc.), most probably because of non-random assignment (Johnson & Christensen, 2014). So, the effect of the independent variable on the dependent variable can then be confirmed by controlling any confounding variable resulted from the non-random assigning of students in the study groups as they were already enrolled in sections by choosing the section's time that suits them, and generalization of the results will be limited to population that may has similar characteristics of the participating population.

(Ivankova et al., 2006) declared that selecting certain research method may be decided based on the purpose of the research and the specific design of each stage. In the current mixed method research, a sequential exploratory design was used to test the impact of delaying the lecture session to come after the laboratory and problem-solving session, by revealing the perspectives of the students and analyse it with the descriptive qualitative results. The priority here will be given to the quantitative data, as the aim is to depend as much as possible on the statistical difference between the change in scientific reasoning of the treated group and the one of the control group, and that can be best analysed by quantitative comparing procedures. Participating students' views and perceptions are to be reflected upon to know "should the order of teaching sequence change; how can each session (Laboratory/Problem solving& Lecture)be conducted to achieve the maximum possible positive impact", qualitative procedures such as in-depth interviews are the best to study this part(Yin, 2003).

To clarify why the planned design of the current study is mixed methods design; some points are stated as follow; First, the uniqueness of the topic-context combination, especially in a country where the local and Arab expats learners are learning in a second language (English). So, by using the mixed methods, the qualitative analysis will reveal unique in-depth information regarding students' perceptions on the proposed teaching technique, on the other hand; the statistical analysis of the numerical data will give valuable ideas for science educators in general and physics educators in particular about which sequence of teaching better be adopted if they are targeting the scientific reasoning of the students, or whether it does not make a difference. Moreover, the quantitative method produces generalizable results which will be relevant and of value to science education departments at the UAE and the Arab countries (Creswell, 2003).

Eventually, using both quantitative and qualitative methods will make the analysis of the results understandable. Due to Johnson and Onwuegbuzie (2004), the main limitation of quantitative data is that the researchers do not comprehensibly know what information can be figured out from the statistical tests, the thing that make it difficult for the findings' consumers such as educational policy makers to depend on numerical un-explained data in their policies.

As per Sandelowski (2000) each method will be supported by one philosophical paradigm different from the one supporting the accompanying method which take a place in the same research. Choosing a paradigm at the beginning of the research, clarify the expectations, motivation, and objectives of the research. It will also make it easier to select the proper methodology and the design of the research (Mackenzie & Knipe, 2006). In this study, a post-positivism philosophy will underpin the quantitative part, to assert what is the reason behind the achieved outcomes. That is obtainable via the scientific reasoning and common-sense reasoning which will be acquired from the statistical numerical results (Sharp et al., 2011).

It was found in a quasi-study by Loehr, Fyfe and Rittle (2014) that the absence of the phase of knowledge application after the instruction phase during the lecturing session may remove potential benefits of the explore-instruct teaching approach. This finding by Loehr, Fyfe and Rittle (2014) was also proved by She and Liao (2009) in a quantitative study, for grade eight students in Taiwan. She and Liao conducted an experimental design study to test the impact of the instruction approach and the effect of the level of the scientific reasoning of the students, on their achievement and scores in three tests which were provided by the researchers to measure their scientific reasoning, concept construction and conceptual change. And they found that the gained knowledge from the delayed lecture should be implemented directly in the same session for the SR and conceptual change to be enhanced. So, such findings by such quantitative studies are expected to be generalized and it was considered in the context of the current study. To reveal such conditions in the current study, qualitative part proposed by the researcher to take place after the quantitative, aiming to know more about the optimum conditions for the new explore-instruct approach to explore as many implicit factors as possible that affect the targeted impact of the proposed technique; qualitative paradigm, in the form of semi-structured interviews had been used beside the quantitative one.

The current research considered the study by Ruiz-Primo and others (2011) at which they tried to develop scientific research of instructional innovations by emphasizing critical issues, and recommended the following: first, descriptive statistics (sample sizes, means, standard deviations) need to be included by all studies for all control and intervention groups on all statistical analysis cases. Second, researchers should try to randomly assign participants to control and treatment groups whenever possible. if this is not feasible, researcher should try to demonstrate that the study groups are equivalent before the intervention with respect to variables (e.g., demographics, prior academic achievement). Finally, researchers should be alert to the quality of their results; if results are not reliable and valid, subsequent interpretations may become misleading, and specific deficiencies may weaken the studies. (Ruiz-Primo and colleagues view that, ensuring the previous recommendations are being considered is a joint responsibility of; researchers, reviewers, journal editors and funding agencies, and they emphasise that experts in experimental methodology and research in education and educational assessment's experts can contribute enormously to improve instructional innovations in science's research (Ruiz-Primo et al.,2011).

Validity may be evidenced by assessing the test content by an expert; in this regard, Lawson made, assessed and updated Lawson Classroom Scientific Reasoning Test (Lawson 2003).

To conclude this methodology part; the proposed design of this study, and due to the previously mentioned purposes, is to be a sequential exploratory mixed design. In the first phase, a quantitative quasi-experiment approach will be utilized to find the difference in SR. ability of the students' pre- intervention and post- intervention, and then compare that difference with the same difference of a control group who will experience a traditional instruct-solve (Loehr et al., 2014) (Loehr, Fyfe &Rittle-Johnson, 2014) teaching approach.

3.1 Context of the study

The study took place in scientific colleges of a private university at the United Arab Emirates, students who were registered in Introductory Physics for scientists and engineers' course in several undergraduate levels of college were enrolled in four sections, the four sections were assigned into two groups. according to the population of students in the sections accessible for the researcher, the studied sample size was defined (should be in average 103 students out of a population of 140 students) according to a sample size identification table: 10.5 in Johnson & Christensen (2014, p. 266-267).

The total number of participating students was 204 (57 males, 147 females). Age of all participating students ranged between 17 and 26 years old. As per (Drachsler & Kirschner, 2011) learner characteristics are important to be known before tailoring any teaching pedagogies.

Demographic characteristics of the participating learners were as follow, 78 female and 22 males, with 40 UAE nationals participants and 60 expatriates. These students' curricula of high school were a mixture of National UAE curricula, American curriculum, National English Curriculum and Indian curriculum. This curriculum mixture is found in almost all UAE universities (Kurian & Al-Assaf, 2020). The participants' level of college year extended between foundation year to 5th year of university.

Intellectual characteristics of language and SRA of participants had also been analysed via studying the participants' IELTs scores and SRA score in Lawson test of SR. Participants' English language level extended between 5.5 and 7.5, where minimum admission requirement in UAE universities is 5.5 in IELTs. SRA of participating students showed 64 concrete reasoner, 35 transitional and 1 formal.

After studying the performance of students prior to the course and post the course, participants with 3 or more points decreasing SRA were excluded from the sample as per (Cone et al., 2016) where it has shown that the health sciences reasoning test manual indicated "students who score three or more points lower than their previous test score were likely less engaged in the test, as an individual's critical thinking ability does not decrease significantly over time, except in case of disease or accidental injury"*p*.719 . Also, students who did not finish more than 60% of the questions or who took less than 15 minutes to complete the test and those who had the test either pre- or post- the treatment and not both were all excluded. Finally, 100 (27 males, 84 females) students were left, out of which 20% were UAE Nationals, and the rest were from other different nationalities. Fifty-one students were assigned to the control group and forty-nine to the experimental group. Both groups received 45 hours of physics sessions in a period of 15 weeks (3 Hours/Week).

The course covered the topics of; Units, Vectors and Scalars, vectors product, motion in one and two dimensions, Newton's laws of Motion, Circular motion, Work and Energy, Conservation of Energy and Oscillatory Motion.

The control group did not experience Explore-instruct approach, they were taught using the common physics lectures teaching approach that starts with explaining the concepts and the steps of solving the related problems, followed by the instructor solving an example with the students and draw their attention to the pitfalls which students usually do when solving such examples and finally assign to the students' extra problems to be solved as homework. The laboratory experiments related to the topics are either conducted later during the same term or in a later term as the laboratory is considered a separate module not registered by some students in the same term with the Physics module (not a co-requisite to the Physics module), but the Physics module is a pre-requisite to the Physics-laboratory module.

The experimental group in the other side contained students who had registered the laboratory and the physics module in the same term but were conducting the experiments in the laboratory prior to having lectures in the physics module. This group also was taught using the Explore-instruct teaching approach. At the beginning of each theory session the students are exposed to problems to be solved prior to explaining the concept and trials of students were then discussed and good trials students were rewarded incentive marks, that was then followed by explaining the concept and accordingly asking the students to solve the previously discussed problems with the guidance of the instructor and reminding the students of the reached results of related experiments previously conducted in the laboratory.

3.2 Instrumentation

3.2.1 Quantitative instrument- LCTSR

In the quantitative part of the research and to answer the first question, Lawson Classroom Test of SR (LCTSR) have been utilized (Appendix II). LCTSR is a multiple-choice two-tier tool that was developed by Lawson in 1978 (Lawson, 1978) updated in 2000, to measure the students' scientific reasoning (Lawson, 2003).It evaluates students' abilities in six dimensions containing; conservation of matter and volume, proportional thinking, Correlational thinking, correlative thinking, control and identification of variables, and hypothetic–deductive reasoning ability(Boudreaux et al. 2008; Lawson 2000; Chen & Klahr 1999). LCTSR has 12 items (questions), with each item containing two tiers, the first tier is for selecting the answer, and the second tier is for selecting the reason of choosing that answer. The second tier demonstrates the ability of the student to use one of the previously mentioned thinking dimensions. One point for each item will be scored only if the two tiers of the item been correctly solved by the student. So, the maximum score is 12 (100%). Based on the score, everyone can be categorized according to his/her formal reasoning level. Below 33% are the concrete operational reasoners, 41% - 66% are the transitional reasoners and students who score more than 66% are the formal operational reasoners(Marušić & Sliško, 2012).

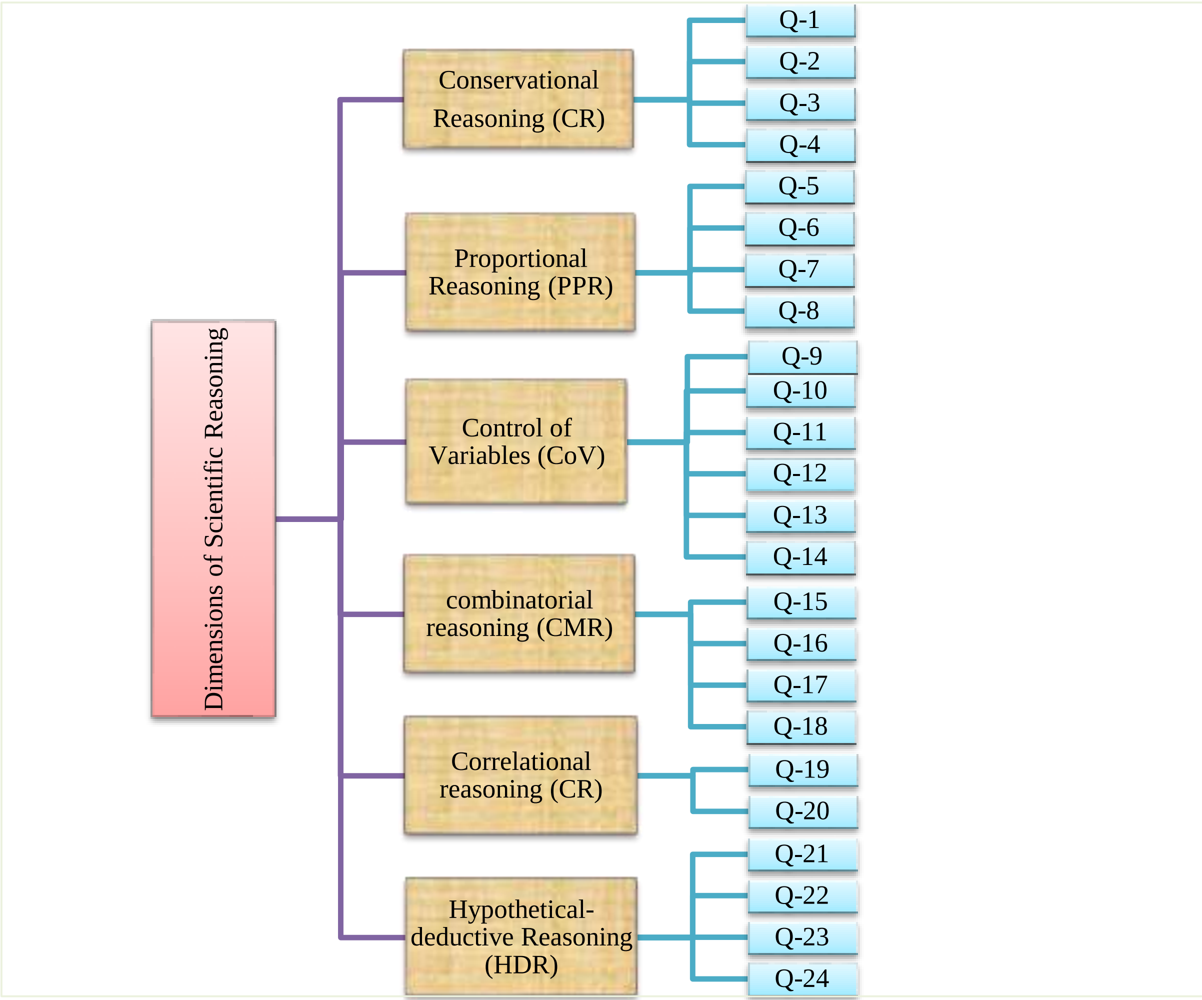


Figure 5. Dimensions of SR in LCTSR.

3.2.1.1 Procedures undertaken to support validity:

Accuracy and validity of the inferences made by the researcher depending on LCRSR test scores can be evidenced based on the *internal structure of the test*. Lawson (2003) showed that reliability of scores achieved in this test to the actual Scientific Reasoning Ability (SRA) of people achieving these scores is acceptable, providing Cronbach's alpha of 0.88 ($\alpha \geq 0.70$) which is according to Johnson and Christensen (2014) evidence of reliability and validity. Validity can also be evidenced by assessing the test content by an expert; in this regard, Lawson made, assessed and updated Lawson Classroom Scientific Reasoning Test (Lawson 2003) But other experts showed that even though LCTSR was popular and widely used (Lei Bao et al., 2018), the results of Lei Bao and others (2018) suggest that the test is a practical instrument to assess a unidimensional scale of SR. The tool has a good overall reliability for the whole test. But examinations of individual tier of question pairs raise a range of validity concerns. So the researcher reviewed literature on results validity of LCTSR. Procedures have also been taken to ensure the validity of LCTSR results in the current study.

3.2.1.1.1 LCTSR results' validity in literature :

Moore and Rubbo, (2012) indicated that investigating differences in populations of students is important when we compare the normalized gains on SR or on concept inventories, and according to Coletta & Steinert (2020), different student populations may show significantly different reactions to the same pedagogy, so we need to consider measuring the abilities of the participating students before we interpret any normalized gain(Coletta & Steinert, 2020).

Luo and others (2020) found that LCTSR is suitable for different grades' students as their scores were mostly between 9 and 25; Mean=16.41, SD=4.492, and all students were able to finish the test during the assigned time. Student saw that LCTSR test was for assessing their general "intelligence" or "thinking", not for science knowledge. Students in Luo and others' study found that some of the questions in LCTSR were difficult to understand, so the questions were verbally explained or rephrased by the researcher for the students to make it understandable (Luo et al., 2020). The influence of English language proficiency and SRA skills on the performance of Hispanic English language learners and native English speakers' students in grade 10 on a standardized science test have been studied by Torres and Zeidler (2002). The results presented by Torres and Zeidler (2002) implies that both factors of higher order of English language competence and SR skills were proven to forecast success in learning concepts of science. (N Torres & L. Zeidler, 2002)

3.2.1.1.2 Validity in the current study:

In the present study, Reliability and Internal validity have been ensured throughout the research. Following is a brief about quantitative results validity, and more details are shown in the results' section.

Reliability refers to the consistency of a set of scores of certain tests. One of the methods to compute that reliability is internal consistency of the test. Internal consistency assesses the consistency with which a test measures a single concept. But in case of a heterogeneous test, where more than one dimension or construct are being measured; reliability of each dimension needs to be assessed. And Coefficient alpha "Cronbach's alpha α" is the formula that gives an idea about the reliability. value of alpha also depends on context, but generally, if α equal to or greater than 0.7 it reflects adequate reliability (Johnson and Christensen, 2014). To ensure reliability and consistency of the used test's items in the current study, reliability of the test in the recent context of this research was checked for each subscale of the six subscales of LCTSR at the beginning of the term before experiencing the physics course (Pre-test). The data collected via Pilot study from 18 students which were not included later in the main sample, and was analysed using SPSS 21 to reveal adequate reliability of each subscale and of total test (more details are in **table 6** in the results' section). The aspects or dimensions of scientific reasoning and the related number of questions in LCTSR are as follow (Lawson, 1995).

Dimension	**Question No.**
Conservation of weight	**1**
Conservation of volume	**2**
Proportional reasoning	**3**
Proportional reasoning	**4**
Identification and control of variables	**5**

Identification and control of variables	**6**
Identification and control of variables	**7**
probabilistic reasoning	**8**
probabilistic reasoning	**9**
probabilistic reasoning	**10**
Combinational reasoning	**11**
Correlational reasoning	**12**

To ensure internal validity (causal validity) which means according to Johnson and Christensen (2014) "establishing a trustworthy evidence of cause and effect" (p. 281), both factors of student's understanding of the test's questions and the causal effect of the independent variable (teaching strategy) have been ensured.

LCTSR in English language, was used at the beginning of the semester to gauge the SR ability of the students in the two groups before the intervention. One group which received the traditional instruct-solve approach (Control) and the other one that received explore-instruct (Experimental). The same test then had been provided for all students at end of the semester. Participants in both groups were not native English speakers. But The researcher and the instructor had the same home language of 86% of the participating students, and as per Quinn and others (2012), native language support for English Language Learners (ELLs) can be used to learn academic processes and content in English. That has been shown by Quinn and others in academic classrooms including the science classroom.

In Science classroom, communication and reinforcement of key scientific concepts and vocabularies in the students' native language impact learning as per (Hudicourt-Barnes, 2003). Students are also to be allowed to communicate with peers using mixtures of English and their first language, described as "translanguaging." by Garcia (2009) (Quinn and others ,2012)

Sireci and Faulkner-Bond (2015) discussed and illustrated several validation studies, using many statistical methods that can evaluate the consistency of test's properties across both populations of ELL and non-ELL (e.g., confirmatory factor analysis), assess item functioning such as Differential item functioning (DIF) analysis, and assess the existence of factors irrelevant to the test construction that may hinder ELL's performance (e.g., multiple regression analysis) (Sireci & Faulkner-Bond, 2015).

A study by (Turkan & Liu, 2012) conducted Differential item functioning (DIF) analysis with regard to ELL status on an inquiry-based science assessment and results showed that, overall, ELLs performance was significantly lower than non-ELLs. Three of the four items which showed DIF have favoured non-ELLs in Turkan and Liu (2012) study, while one item favoured ELLs. The item which favoured ELLs was the one that provided a graphical representation of a science concept inside a family context. The reason may be behind the constructed response represented in the graphical representation. Because as per Turkan and Lio some evidence is there supporting the fact that constructed-response items can support ELLs communicate SR using their own expressions. So, teachers need to pay attention to the probable interaction between science content and linguistic challenges when teaching or when designing assessment to ELLs. Sireci and Faulkner-Bond (2015) have learned a lot about different accommodations' types that may help ELLs when assigned academic assessments. They benefited from Abedi and Ewers (2013) meta-analyses, and classifications in understanding the kinds of accommodations which are most likely to be efficient in eliminating linguistic barriers on a test while staying genuine to the measured content (appendix IIX) (Sireci & Faulkner-Bond, 2015)

In the United Arab Emirates (UAE), undergraduates are diverse in their nationalities as well as in their followed high school curriculum. Paper of (Kurian & Al-Assaf, 2020) presents a pilot study conducted in a context close the current study context, at Rochester Institute of Technology (RIT), Dubai to identify the impact different high school curricula have on undergraduates' performance in the university. Academic as well as non-cognitive measurements of records of 213 undergraduates were studied. And as per Kurian and Al-Assaf (2020) Indian curriculum undergraduates were found to perform better than other students in university. UAE's national curriculum was found to be a revised version of American curriculum. Readiness for college of UAE's national curriculum undergraduates was noticed to be the lowest in the studied population. Revised American curriculum undergraduates showed a noticeable gap between their actual performance (Grade Point Average) and the self-perception of their capabilities (Kurian & Al-Assaf, 2020).

Significant difference tests were performed in the multi-group design followed in this study, to check differences in scores and categories of SR collected prior to the physics course in the two groups. The reason for those tests in general is because researcher had to use groups which are already formed (undergraduates already enrolled in sections) giving up the random selection and accordingly the participants in the two groups may differ which is the case in this study.

So, scores of SR dimensions are to be analysed against gender, age, nationality, IELTs scores, and the starting SRA prior to having the intervention. If p-value of the t-test (between two groups) or ANOVA test (between more than two groups) are less than 0.05 ($p \leq 0.05$), that shows less than 5% acceptance of the null hypothesis that states "There is no difference in scores and categories of SRA between the groups" and more than 95% rejection of the null hypothesis (Christensen et al., 2014). Once any confounding factor related to the previously mentioned demographics of participants is discarded; the causal validity of the teaching pedagogy will be more trusted.

3.2.2 Qualitative instrument- Email's interview

In the 1980s and 1990s the semi-structured interviews were developed by the psychologists Scheele and Groeben to study subjective theories in fields as schools and other professional work areas (Uwe Flick, 2014, p.217). Accordingly, in the qualitative part here, semi-structured email interviews had been carried out. In taking the decision of choosing semi-structured interviews as a mean of data collection in this part of the research, the researcher depended on the following points: First, the second research question represent a mean for exploring the students' views, this will guide us in deciding how each teaching session – Laboratory/Tutorial for problem solving or lecture – should differ when switching from traditional style to the style under investigation to achieve the optimum impact. Second, semi-structured interviews are featured according to Uwe Flick, (2014) with prepared open-ended, hypothesis-directed, and conferential questions which leave a space for the interviewee's perspectives which are under consideration here to be revealed. Thirdly, the semi-structured interviews are not difficult to be applied by the researcher who is an educator with a reasonable experience in communicating with students and getting their feedback. Lastly, this type of interviews is believed by the researcher to suite the type of interviewees in the current study; as the college students are expected -to some extent- to value the educational research and to be willing to give their own perspectives.

Flick also demonstrated in his book the different types of questions to be asked and their sequencing in the interview protocol. Open-ended questions to be at the beginning of the interview and to be answered based on the knowledge which the interviewee has at hand. Following the open questions are the hypothesis-directed questions in which the interviewer gives his assumptions as offers for the interviewees, either they take it up or reject it. Together with the previous two types of questions come the confrontational questions at the end to validate the participants' perspectives (p. 218). The set questions for the participants in this study, aimed at exploring their opinion on the tried teaching strategy which they experienced in the physics course, exploring their view about their lab experience, and in which case do they think they had been able to understand the concept or the principle of the experiment more; If they conduct the experiment (perform it by following the steps provided in the sheet or the manual) then receive the lecture from the instructor? Or, if they fully had the lecture by the instructor and then have the experiment? (hypothesis-guided open-ended question).The last question for the interviewees was the confrontation 2-choises question, about their opinion; if they feel comfortable with the Explore-instruct strategy? and if they agree with the view that "Explore-instruct approach ultimately clarifies the concept more than the instruct-solve (conventional) strategy, but the student doesn't feel comfortable in conducting the experiment before being instructed"? and if there was any point they would like to mention about explore-instruct teaching strategy?

E-mail was used to interview the students. The reason for selecting the e-mail as an interviewing tool is that the participants and the researcher happened to be at two different cities at the time of the interviews, and as per(James, 2016); email interviews give interviewees space and time to reflect on their academic experiences. And the asynchronous communication gives them time preference. Two from the middle and two from the low Normalized gain of SR replied to the email interviews' questions. The replies were then coded and analysed using thematic coding by the researcher to induce conclusions (subjective theories of students as per the grounded theory by Glaser and Strauss (1976)).

Some disadvantages of content analysis are details' loss and the loss of "sense of individual participants" (Wilkinson, 2008, p. 201) in addition to some issues in coding that appear when trying to integrate many quotes into related meanings. However, if the qualitative data had been collected from big number of participants, coding problems would be greater. In the current study coding problems are less likely to happen as a small sample is less likely to be compelled into category groups; small number of responses are much easier to interpret than bigger number as per Hseih, and Shannon (2005). But there may be a loss of some perspectives as not all participants were included in the interviews.

3.3 Ethical consideration

The ethical clearance was obtained prior to starting any processes of investigation or data collection, because human resources of the considered institute to be used in the research (Appendix I). The proposal was submitted to the Institutional Board of Research (IBR) and a form of ethical clearance was filled and attached to it. The students' consent was also obtained prior to participating in any of the study parts.

CHAPTER 4

DATA ANALYSIS AND RESULTS

This section presents the Quantitative and Qualitative data collected and the statistical analysis to answer the first and second research questions. The researcher collected quantitative data via Quasi-experimental method with two groups design. Students enrolled in four sections of University Introductory Physics were tested to level their Scientific Reasoning Ability (SRA) at the beginning of a three-months term. The total sample size was 110 students at the beginning. The sample was then split into experimental and control groups and that was to eliminate threats to internal validity that may result from any extraneous variable other than the independent variable (Johnson & Christensen, 2014). After collecting data and cleaning it from data of incomplete tests' scores and scores that showed decrease in the total SRA which is decided to be inaccurate (Cone et al., 2016), V.P. Coletta showed with Steinert (2020) that Pre-test and post-test statistical results on conceptual tests are significantly disturbed by guessing. The disturbance decreases if the scores increase. But the normalized gain is almost unaffected by the guessing and may be securely used even if pre-test and post-test scores need significant correction. eventually the number of participants became 100 (51 control, 49 experimental). Randomizing the students in sections was not authorized for the researcher in this study, because the registration process of students in the different sections took place earlier by the students themselves and the registration department. Average number of students in each section was 25. The control group experienced the traditional lecturing strategy of presenting the topic by the instructor and solve example problems and finally engaging the students in solving problems. Participating students' SRA was measured again at the end of that term. The experimental group in the other hand; taught via the Explore-Instruct teaching strategy under study where they explore solving problems first followed by discussion of solutions and the instructor presentation of the topic at last. After documenting their scores pre/post receiving the physics course; the participants' data was then analysed using 'IBM SPSS 21' to answer the first research question "Is there a difference in college students' scientific reasoning ability's (SRA) development if taught via explore-instruct teaching style compared to the instruct-solve style (traditional approach) in physics?". In answering this question, the following hypotheses and their alternatives were set:

$\boldsymbol{H_1}$: There is significantly higher scores of SR Post-test than Pre-test for all participants in both the control and experimental groups.

$\boldsymbol{H_{01}}$: There is no significant difference between SRA Post-test and Pre-test for all participants in both the control and experimental groups.

$\boldsymbol{H_2}$: There would be a significantly higher Normalized Gain in all dimensions of SR in the experimental group than in the control.

$\boldsymbol{H_{02}}$: There is no significant difference in all dimensions of SR between the experimental and the control group.

$\boldsymbol{H_3}$: There is a significantly higher tendency of improving the category of SRA in the experimental group than the control group.

$\boldsymbol{H_{03}}$: There is no significant difference in tendency to improve category of SRA between the experimental and the control group.

Description of participants' *demographics* was the preliminary analysis presenting the characteristics of the studied samples. Following the sample description is the *descriptive* analysis to study the characteristics of the collected data of variables under investigation, starting with data distribution, and then presentation of numerical characteristics of collected data, such as the mean, standard deviations, minimum, maximum and percentages distribution in categories. Lastly the *inferential* tests are to be performed to study the relation between variables and be able to predict the effect on a variable if another variable experienced a change (Johnson & Christensen, 2014).

Email interviews were sent to six participants to answer the second question. 2 students with the highest SRA Growth, 2 with the middle SRA Growth and 2 with the lowest SRA in the experimental group. Four of the six replied, the purpose was to explore students' perceptions related to Explore-Instruct teaching approach in Physics courses and laboratories.

4.1 Results of Question 1- Lawson Classroom Test

The results of this part have been carried out to answer the first research question; "Is there a difference in college students' scientific reasoning ability's (SRA) growth if taught via explore-instruct teaching strategy compared to the instruct-solve strategy (traditional approach) in physics?" The software suggested by Muijs, (2011) "the Statistical Packages for Social Studies (SPSS)" was used, and the section is divided into four parts. It begins with a description of the participants' *demographics* in the first part. The second part presents the *descriptive* analysis of quantitively collected data. This second part intended to analyse the dependent variables and relations between them; including Pre-intervention variables (LCTSR Scores prior to the intervention, Category of SR of the student prior to the intervention "Categ.SR-PRE") and Post-intervention variables (LCTSR Scores post the intervention "Post-test", Category of SR of the student prior to the intervention "Categ.SR-POST." And SRA Growth) in both groups. The third part of this section deals with the *reliability and validity* of the SR test to this study.

In the last part of this section, *inferential* statistical analysis carried out to study the hypothesis set above. It described the inferences and predictions that may be made about the studied population depending on data from the sample of this population (Johnson & Christensen, 2014). Differences between the experimental and control groups in SRA Growth is studied in this inferential statistics section. Distribution of students among the different stages of SR and their mean score in each category after experiencing the two teaching styles was also analysed, followed by correlation tests between dependent and independent variables which as per (Johnson & Christensen, 2014) provides information about the relationship between two variables. Finally, *regression analysis* took place to study the relation between *Post-CoV* and *Pre-CoV*. and, between the *teaching strategy* and *N-Gain in CoV.*

4.1.1 Descriptive statistics -Demographics

Demographic analysis of participants had been carried out for $(N = 100)$ participating students, the control group contained $(n = 51)$ students enrolled in two sections of General Physics-I, and the other $(n = 49)$ students in the experimental group enrolled also in two sections (***Figure 6***) . These participants were in various academic levels (Foundation year to 5^{th} year) of a private university in the United Arab Emirates and General Physics-I course here was a college requirement course for scientists and engineers.

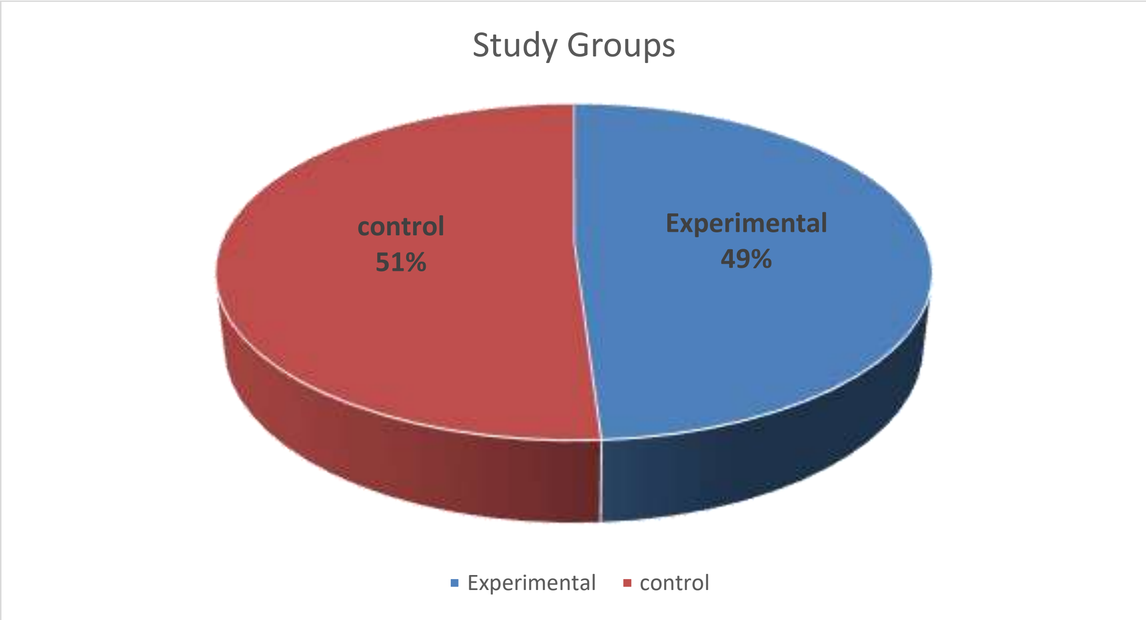


Figure 6. Percentage of students in the control and experimental groups

Age of all participating students ranged between 17 and 26 years old. In the experimental group 98% of the undergraduates were below 21 (19-20) years old and only 2% were more than 21 years old. The control group included 53% below 21 years old (17 – 20) and 47% of the total number of participants were at /elder than 21 years old (21-26) years old (***Figure 7***).

Gender distribution of the sample was as follow; The experimental group was $100\% \ (n = 49)$ females, while the control group included 57% (n=29) females and 43% (n=22) males.

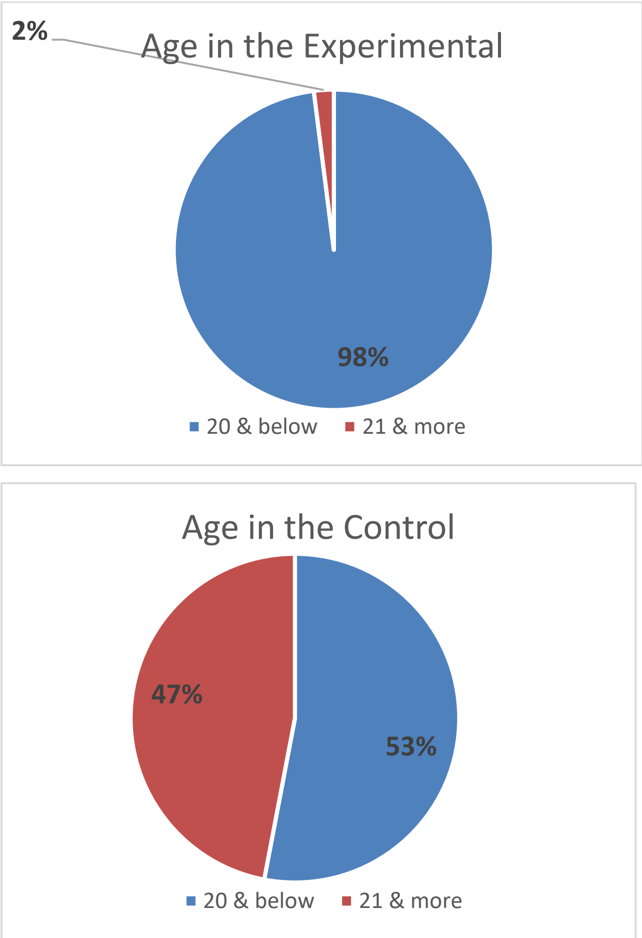


Figure 7. Age distribution of the participating students

Nationalities of the participates had also been studied, and among the students in the experimental group, 69%, (*n*=34) were UAE national students while 29% (*n*=14) were Arabs non-UAE nationals and 2% (*n*=1) Pakistani student, The control group included 12% UAE nationals, while the rest were expatriates as shown in **Figure *8*** .

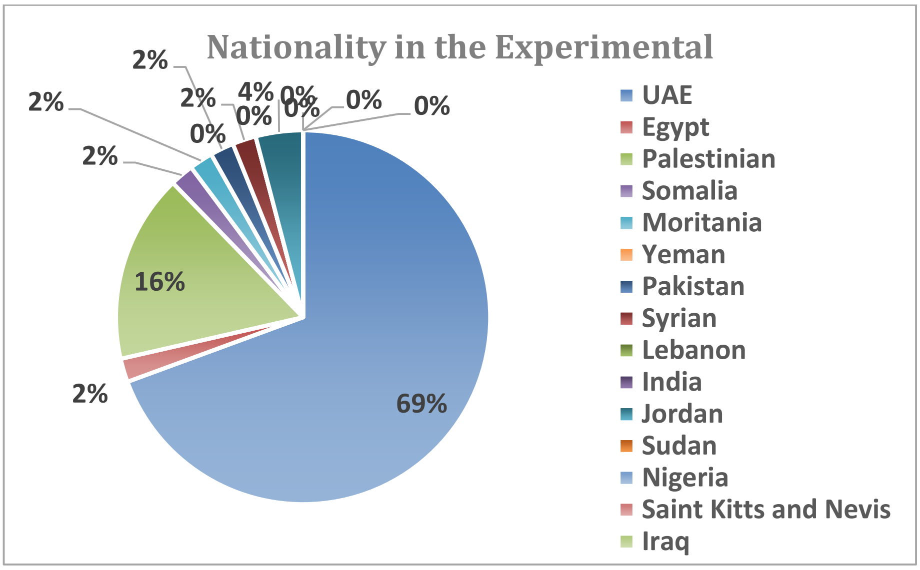


Figure 8:Nationality distribution of the participants in the experimental group.

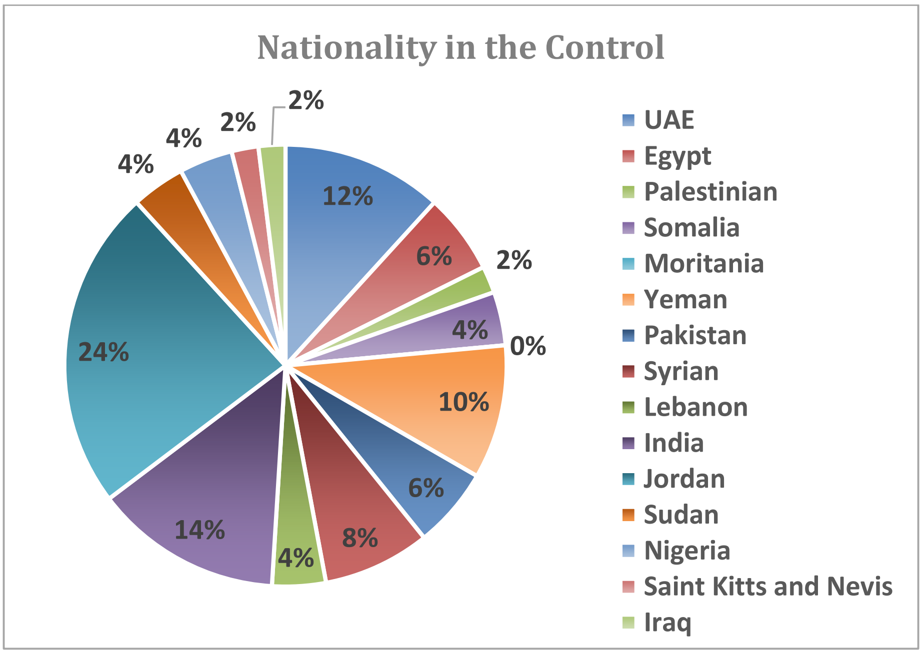


Figure 9. Nationality distribution of the participants in the control group.

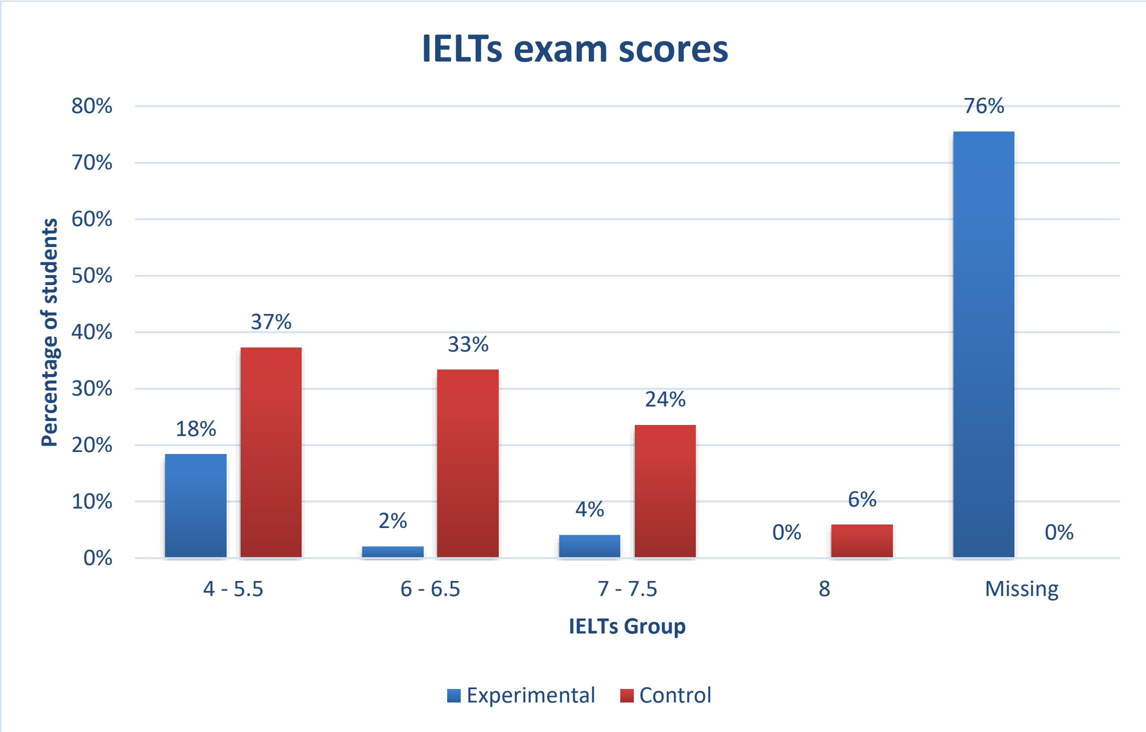


Figure 10. English language level-IELTs

The *language* of LCTSR was English, while the participant's native language was not English (mostly Arabic). So, students' English language level was indicated by their IELTs scores and presented in **Figure *10***. English language level-IELTs The IELT's scores were categorized for this study context and as per a statistician view into four groups: (1) 4-5.5 , (2) 6-6.5, (3) 7-7.5, (4) 8 and more. 76% of the experimental group data of IELTs was missing (*Figure 10*).

The last demographic to be presented here; the SR level of participating students prior to the intervention (Pre-test% and *Categ.SR-Pre*), it gives an idea about the participant's SRA before being affected by any factor during the time of this study. The percentage of the total SR of each participant identify the category of SR level that participant is in. As per Lawson (1995); Concrete operational score 33% and below, while *transitional* reasoners score between 41% and 66% and formal reasoners score between

67% and 100%. Percentages of student's categories in both study groups were presented in **Figure *11***

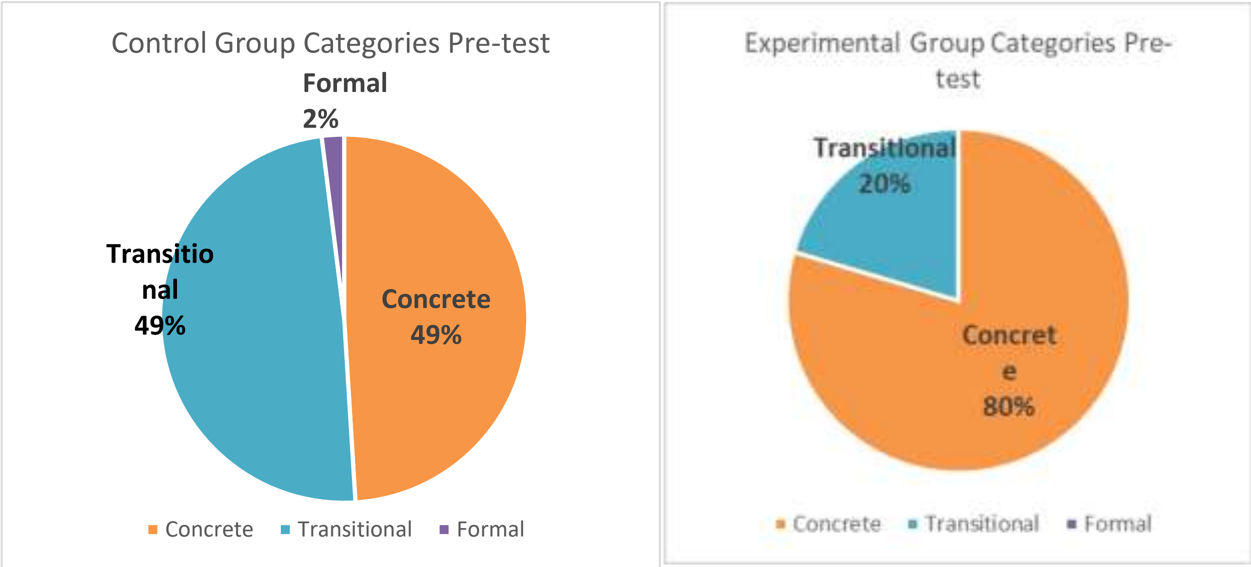


Figure 11. Categories of SR pre-instruction

At the beginning of the term, out of the 49 students *in the experimental* group 80% (*n*=39) were in the *concrete* level of SRA, 57 % (*n*=10) *transitional* reasoners and none was a formal reasoner. In the *control group,* which contain 51 students, 49% (n=25) were categorized as concrete reasoners, 49% (n=25) transitional and 2% (n=1) formal reasoners.

Starting score of each dimension of SR, can be described here as a demographical variable of participants in the two study groups. SR have according to (Anton E. Lawson, 2004) six dimensions. Table 5 presents descriptive statistics of the six SR dimensions' percentage prior to the instruction for participants of this study ($N = 100$) in both the experimental ($n = 51$) and control ($n = 49$) groups. Percentage of any dimension either prior or post the course was calculated utilizing the following formula:

$$Pre\ test\ \% = \frac{frquency\ of correct\ answers\ or\ frequency\ of\ 1's\ for\ the\ dimension's\ questions}{Maximum\ possible\ frequency\ of\ 1's\ for\ the\ dimension's\ questions} \times 100\ \%$$

Table 5: *Pre-test % Descriptive Statistics*

SR dimension	*Group*	*N*	*Minimum*	*Maximum*	*Mean*	*SD*
CR Pre-test	Control	51	0	100	63.73	28.423
	Experimental	49	0	100	50	30.619

PPR Pre-test	Control	51	0	100	26.47	39.183
	Experimental	49	0	100	13.27	26.567
CoV Pre-test	Control	51	0	100	26.73	28.346
	Experimental	49	0	100	20.33	27.055
CMR Pre-test	Control	51	0	100	66.67	40.825
	Experimental	49	0	100	23.47	34.007
CRR Pre-test.	Control	51	0	100	29.41	46.018
	Experimental	49	0	100	22.45	42.157
HDR Pre-test	Control	51	0	50	6.86	17.377
	Experimental	49	0	50	24.49	25.254

(CR) Conservational Reasoning, (PPR) Proportional Reasoning, (CoV) Control of Variables, (CMR) Combinatorial reasoning, (CRR) Correlational reasoning, (HDR) Hypothetical-deductive Reasoning.

4.1.2 Reliability – LCTSR test

Before accomplishing the empirical tests, Reliability, and validity of the used LCTSR Lawson Classroom Test of Scientific Reasoning was checked in the context of this study. As per Johnson and Christensen (2014) "when judging the performance of a test and your interpretation based on that test, remember that reliability and validity are both important properties. You need both." (p. 166).

Reliability refers to the consistency of a set of scores of certain tests. One of the methods to compute that reliability is internal consistency of the test. Internal consistency assesses the consistency with which a test measures a single concept. But in case of a heterogeneous test, where more than one dimension or construct are being measured; reliability of each dimension needs to be assessed. And Coefficient alpha "Cronbach's alpha α" is the formula that gives an idea about the reliability. value of alpha also depends on context, but generally, if α equal to or greater than 0.7 it reflects adequate reliability (Johnson and Christensen, 2014).

Lawson Classroom Test of Scientific Reasoning (LCTSR)evaluates students' abilities in six dimensions containing; conservation of matter and volume (CR) covered in *questions 1 - 4*, proportional Reasoning (PR) covered in *questions 5 - 8*, control and identification of variables (CoV) covered in *questions 9 - 14*, Combinational Reasoning (CMR) covered in *questions 15 - 18*, Correlative Reasoning covered in *questions 19 – 20* and Hypothetic–Deductive Reasoning (HDR) covered in *questions 21 - 24* (Boudreaux et al. 2008; Lawson 2000; Lawson, 2004; Chen & Klahr, 1999). LCTSR questions' distribution among the aspects or dimensions of SR are demonstrated in **Error! Reference source not found.**.

Accuracy or validity of the inferences made by the researcher, after data collection and recording of the test scores, may be evidenced based on the internal structure of the test that was discussed earlier in the reliability section. Validity may also be evidenced based on assessing the test content by an expert; Lawson created, assessed and updated Lawson Classroom Scientific Reasoning Test (Lawson 2003)and showed that reliability of this test to the study of Scientific Reasoning Ability is acceptable with Cronbach's alpha of 0.88 ($\alpha \geq 0.70$)which is according to Johnson and Christensen (2014) an evidence of reliability and validity. When (Luo et al., 2020) analysed scores of 552 students, the correlation between the students' SRA test scores and the scores of LCTSR was .527, Such association indicated that, in line with the SRA test which had Cronbach's α of .809 indicating good reliability, LCTSR also had good and statistically significant practical validity and would help determine students' SRA.

Though the popularity and wide usage of LCTSR; (Lei Bao et al., 2018) The results of Lei Bao and others (2018) suggest that the LCTSR is a practical instrument to assess a unidimensional scale of SR. The tool has a good overall reliability for the whole test. But examinations of individual tier of question pairs raise a range of validity concerns. So, reliability of the test in the recent context of this research was checked for each subscale of the six subscales of LCTSR at the beginning of the term before experiencing the physics course (Pre-test). The data collected via Pilot study from 18 students which were not included later in the main sample, was analysed using SPSS 21 and revealed adequate reliability of each subscale (Table 6). It is clear in Table 6 that all sub-scales are trusted to be reliable, even the lowest α value of the Control of Variables sub-scale ($\alpha = 0.665$) may be rounded to 0.70. After that, all factors together with their 12 sub-items were used to compute the total reliability of the test and it is found to be good and equal to 0.703. one may therefore conclude that the consistency of data collected from the current population is assured.

Table 6*:* Reliability test

Factors	24 Sub-Items	12 Sub-items	Cronbach's Alpha	N of Items	Number of cases
Conservational reasoning	Item 1 Item 2	Q 1	.717	2	18
	Item 3 Item 4	Q 2			
Proportional reasoning	Item 5 Item 6	Q 3	.750	2	
	Item 7 Item 8	Q 4			
Control and identification of variables	Item 9 Item 10	Q 5	.667	3	
	Item 11 Item 12	Q 6			
	Item 13 Item 14	Q 7			
Combinational reasoning	Item 15 Item 16	Q 8	.776	2	
	Item 17 Item 18	Q 9			
Correlational reasoning	Item 19 Item 20	Q 10	1	1	
Hypothetical-Deductive reasoning	Item 21 Item 22	Q 11	.750	2	
	Item 23 Item 24	Q 12			

After checking the reliability, and to ensure internal validity and the causal effect as per Jonson and Christensen (2014), demographics of students need to be proved not to have a significant effect on participants' performance. That non-significant effect is essential to assure strong relation between the noticed difference -between the experimental and control group - and the studied independent variable (Explore-Instruct teaching strategy) and minimize threats to internal validity.

4.1.3 Internal validity - Testing for SR Differences related to Demographic differences.

To ensure internal validity (causal validity) which means according to Johnson and Christensen (2014) "establishing a trustworthy evidence of cause and effect" (p. 281) significant difference tests were performed in the multigroup design followed in this study, to check differences in scores and categories of SR collected prior to the physics course in the two groups. The reason for those tests in general is because researcher may have to use groups which are already formed (undergraduates already enrolled in sections) giving up the random selection and accordingly the participants in the two groups may differ which is the case in this study. Moore and Rubbo, (2012) indicated that investigating differences in populations of students is important when we compare the normalized gains on SR or on concept inventories, and according to Coletta & Steinert (2020), different student populations may show significantly different reactions to the same pedagogy, so we need to consider measuring the abilities of the participating students before we interpret any normalized gain (Coletta & Steinert, 2020).

Therefore, the scores of SR need to be analysed against gender, age, nationality, IELTs scores, and the starting SRA prior to having the intervention. If *p*-value of the t-test (between two groups) or ANOVA test (between more than two groups) are less than 0.05, that shows less than 5% acceptance of the null hypothesis that states "There is no difference in scores and categories of SRA between the groups" and more than 95% rejection of the null hypothesis (Christensen et al., 2014). So, there will be a significant difference between the variables only if $p \leq 0.05$.

The data distribution of variables should be tested to ensure a normal distribution. That normality test is essential to decide on the types of statistical tests needed further; whether it should be parametric in case of normal distribution or non-parametric tests if the distribution is not normal.

According to Ahmad et al. (2018) One Sample Kolmogorov-Smirnov Z-test compare the distribution of the data under study against a standard theoretical distribution curve and normality of data would be proved if Kolmogorov-Smirnova *p* values were less than 0.05 (Huizingh 2007; Ahmad et al., 2018).

Therefore, One Sample Kolmogorov-Smirnov *Z*-test of normality was performed to decide on normality of the *Pre-test% of SR-dimensions*, *Post-test% of SR-dimensions* and *N-Gain in each dimension* for the total sample size and results are demonstrated in **Table** *7*. The abbreviation of each variable is used in the table with the full-form provided in the notes under the table.

P- values of Kolmogorov-Smirnova for pre-test scores , post-test scores and SRA Growth of all dimensions were less than 0.05 ($p< 0.05$) revealing a non-normal distribution. Accordingly, significant testing among demographics should be performed via non-parametric tests of Mann Whitney instead of t-test and Kruskall-walis instead of ANOVA.

Table 7: *One Sample Kolmogorov-Smirnov Z- test*

		Variable	N	Mean	SD	Kolmogorov-Smirnov Z	p
One-Sample Kolmogorov-Smirnov Test	Pre-instruction	CR	100	23.6	27.8	2.92	0.00
		PPR	100	45.5	43.3	2.73	0.00
		CoV	100	26.0	44.1	4.62	0.00
		CMR	100	15.5	23.2	4.38	0.00
		CRR	100	67.5	31.3	2.82	0.00
		HDR	100	30.0	40.2	3.72	0.00
	Post-instruction	CR	100	41.6	30.6	2.41	0.00
		PPR	100	51.5	44.6	2.71	0.00
		CoV	100	49.0	50.2	3.45	0.00
		CMR	100	24.5	29.7	3.55	0.00
		CRR	100	23.6	27.8	2.92	0.00
		HDR	100	45.5	43.3	2.73	0.00
	N-Gain	CR	100	.195	.512	3.58	0.000
		PPR	100	.120	.477	3.69	0.000
		CoV	100	.178	.470	2.02	.001
		CMR	100	.110	.544	3.10	0.000
		CRR	100	.300	.461	4.43	0.000
		HDR	100	.035	.489	3.41	0.000

Category of SR	Pre-Instruction	100	1.37	.506	4.08	0.000
	Post-Instruction	100	1.66	.590	3.18	0.000

Note. Pre-test = Scores of Lawson Classroom Scientific Reasoning Test Prior to the course**; Post-test = S**cores of Lawson Classroom Scientific Reasoning Test Post the course**; N-Gain**= Normalized Gain; **CR**= Conservational reasoning; **PPR** = Proportional reasoning; **CoV** = Control and identification of variables; **CMR**= Combinational reasoning; **CRR** = Correlational reasoning; **HDR** = Hypothetical-Deductive reasoning.

Starting by *gender*- Mann Whitney results of students' six dimensions of Scientific Reasoning prior to the intervention, among the two *gender* groups was shown in **Error! Reference source not found.**. The gender has no effect on the students' s cientific reasoning dimensions at the beginning of the course except for the Combinational reasoning U ($N_{Male} = 22, N_{Female} = 78$) = 612.000, Z= -2.19= $p = 0.029$, and the Hypothetical-deductive reasoning U ($N_{Male} = 22, N_{Female} = 78$) = 667.000, Z= -1.984= $p = 0.047$

Table 8: *Mann-Whitney U test of SR dimensions- among Gender*

SR Dimension Pre-test	Gender	N	Mean Rank	Mann-Whitney U	*P*
CR Pre-test	M	22	57.82	697.000	
	F	78	48.44	Z= -1.55	.120
	Total	100			
PPR Pre-test	M	22	47.43	790.500	
	F	78	51.37	Z= -0.705	.481
	Total	100			
CoV Pre-test	M	22	46.23	764.000	
	F	78	51.71	Z= -0.855	.393
	Total	100			
CMR Pre-test	M	22	61.68	612.000	
	F	78	47.35	Z= -2.19	.029*
	Total	100			
CRR Pre-test	M	22	53.41	794.000	
	F	78	49.68	Z= -0.701	.483
	Total	100			
HDR Pre-test	M	22	41.82	667.000	.047*
	F	78	52.95	Z= -1.984	
	Total	100			

Note. **M**=Male , **F**=Female
$^{*}P < 0.05$

Studying the *Age,* Mann Whitney Results of students' *Scientific Reasoning Pre-test* in the six dimensions among the two *age* groups is presented in ***Table 9***. Results show that there is no significant difference in the students' SR dimensions Pre-test among the different *age* groups except for the Combinational U($N_{Age\leq 20} = 75, N_{Age>20} = 25$) = 620.000, Z= 1580 , $p = .007$andfor the Hypothetical-DeductiveReasoningU($N_{Age\leq 20} = 75, N_{Age>20} = 25$) = 750.000, Z= 1075, $p = .028$. These results make the participant's age an extraneous factor if the combinational reasoning or Hypothetical-Deductive reasoning of the participants were compared between the control and the experimental groups. So, age should be controlled when looking for an impact of the followed teaching strategy on these two dimensions of SR.

Table 9: *Mann-Whitney U test SR Dimensions among Age groups*

Age Group		N	Mean Rank	Mann-Whitney U	p
CR Pre-test	Equal or less then 20	75	48.05	753.500 Z= 1446.5	.089
	More than 20	25	57.86		
	Total	100			
PPR Pre-test	Equal or less then 20	75	49.81	886.000 Z= 1314	.607
	More than 20	25	52.56		
	Total	100			
CoV Pre-test	Equal or less then 20	75	49.49	861.500 Z= 1338.5	.509
	More than 20	25	53.54		
	Total	100			
CMR Pre-test	Equal or less then 20	75	46.27	620.000 Z= 1580	.007*
	More than 20	25	63.20		
	Total	100			
CRR Pre-test	Equal or less then 20	75	51.50	862.500 Z= 1187.5	.432
	More than 20	25	47.50		
	Total	100			
HDR Pre-test	Equal or less then 20	75	53.00	750.000 Z= 1075	.028*
	More than 20	25	43.00		
	Total	100			

*$P < 0.05$

Here pops up a question; Is the difference in starting CMR and HDR is due to Age difference or Gender difference, or because gender and age percentages are different between the control and experimental groups? So, a confounding variable is the study group which include different starting CMR% and HDR%.

Nationality- of participants' effect on SR was also investigated. *Kruskal-Wallis* test results of students' Scientific Reasoning Pre-test scores in the six dimensions among the different *nationalities* was shown in Table 10. The Chi-square values were presented, and it's proved that the participants' (N=100) nationalities made no significant difference on the students' scientific reasoning Pre-test scores with (all p values > 0.05) of the six dimensions.

Table 10: Kruskal-Wallis test for SR among Nationalities

SR dimension		Nationality														
		UAE	Egypt	Palestinian	Somalia	Mauritania	Yemen	Pakistan	Syrian	Lebanon	India	Jordan	Sudan	Nigeria	Saint Kitts and Nevis	Iraq
	N	40	4	9	3	1	5	4	5	2	7	14	2	2	1	1
CR Pre-test	Mean	45.65	65.5	54.06	58.17	6.5	44.9	43.5	37.5	65.5	44.5	62.36	65.5	65.5	87.5	43.5
	X^2	16.408														
	p	0.289														
PPR Pre-test	Mean	51.55	47.13	40.94	36	80.5	47.8	47.13	77.4	58.25	59.21	40.21	36	65.5	95	36
	X^2	21.211														
	p	0.096														
CoV Pre-test	Mean	51.56	59.75	39	46.83	90.5	48.1	46	46.5	78.75	49	55.36	25	46	67	25
	X^2	10.797														
	p	0.702														
CMR Pre-test	Mean	39.14	52.75	43.33	42.33	84	65.7	61.13	59	69.5	66.79	58.18	84	55	84	21.5
	X^2	22.54														
	p	0.068														
CRR Pre-test	Mean	55	62.5	37.5	37.5	37.5	47.5	37.5	37.5	62.5	51.79	55.36	62.5	37.5	37.5	37.5
	X^2	13.878														
	p	0.459														

HD Pre-test	Mean	51	66	63.2	41	41	41	41	51	41	48.14	51.71	41	41	41	41
	X^2								11.28							
	p								0.664							

English language- Kruskal Wallis test was also performed to study the difference in Pre-test % scores of SR among the different levels of English language of participants. That test was necessary because LCTSR language was English while participant's native language was not English (mostly Arabic). Students' English language level was indicated by their IELTs scores, and result of Kruskal Wallis test for the whole sample was presented in **Table *11***. It was clear that participants' English language level makes no difference on their scientific reasoning ($Sig. > 0.05$).

Table 11: *Kruskal-Wallis test - English language level.*

SR Dimension	IELTs Group	N	Mean Rank	Chi-Square	*p*
CR Pre-test	4 - 5.5	28	35.61		
	6 - 6.5	18	26.56	4.325	.228
	7 - 7.5	14	33.29		
	8	3	25.00		
	Total	63			
PPR Pre-test	4 - 5.5	28	29.91	3.277	.351
	6 - 6.5	18	35.83		
	7 - 7.5	14	33.61		
	8	3	21.00		
	Total	63			
CoV Pre-test	4 - 5.5	28	33.54	1.147	.766
	6 - 6.5	18	30.78		
	7 - 7.5	14	32.39		
	8	3	23.17		
	Total	63			
CMR Pre-test	4 - 5.5	28	30.05		
	6 - 6.5	18	32.97	1.269	.737
	7 - 7.5	14	32.75		
	8	3	40.83		
	Total	63			
CRR Pre-test	4 - 5.5	28	31.88		
	6 - 6.5	18	34.50	1.711	.635

	7 - 7.5	14	30.75		
	8	3	24.00		
	Total	63			
HDR Pre-test	4 - 5.5	28	32.50	.564	.905
	6 - 6.5	18	31.50		
	7 - 7.5	14	32.50		
	8	3	28.00		
	Total	63			

a. Kruskal Wallis Test

b. Grouping Variable: IELTs Group

Starting categories of SRA(*Categ.Of.SR-PRE)*were investigated for differences in each of the six dimensions. Kruskal Wallis test results are shown in ***Table 12***. The only dimension that is not showing a significant difference between students of the different categories is the Hypothetical-Deductive dimension $\chi^2(2, N = 100) = 3.398, p > .05$.All the other dimensions' scores significantly differ among the categories ($p < 0.001$) with the lowest mean rank on the concrete level followed by a higher rank at the transitional and the highest mean rank of SR dimension at the formal category.That means the higher the category of the student's SR , the higher are the accompanying dimensions of SR except for the HDR dimension.

Table 12: *Kruskal Wallis test – Categ.Of.SR-PRE*

	Categ.Of.SR-PRE	N	Mean Rank	χ^2	P
CR Pre-test	Concrete	64	43.22	16.035	.000**
	Transitional	35	62.76		
	Formal	1	87.50		
	Total	100			
PPR Pre-test	Concrete	64	42.02	25.610	.000**
	Transitional	35	64.74		
	Formal	1	95.00		
	Total	100			
CoV Pre-test	Concrete	64	38.20	39.121	.000**
	Transitional	35	71.61		
	Formal	1	98.50		
	Total	100			
CMR Pre-test	Concrete	64	40.16	26.125	.000**
	Transitional	35	68.44		
	Formal	1	84.00		
	Total	100			

CRR Pre-test.	Concrete	64	43.75	17.995	.000**
	Transitional	35	61.79		
	Formal	1	87.50		
	Total	100			
HDR Pre-test	Concrete	64	48.28	3.398	.183
	Transitional	35	53.57		
	Formal	1	85.00		
	Total	100			

$^{**}P < 0.001$

4.1.4 Descriptive Statistics -Study Variables

Descriptive statistics aim at describing or summarizing a set of collected data after teaching performance in the two groups. The *teaching strategy* followed with each group was the *independent variable* in this study and the change in students' scientific reasoning *(N-gain)* was the *dependent variable* under focus. Variables such as students' score of SR at the beginning of the term *Pre-test%*, score at end of the term *Post-test%*, the Growth or the Gain in SRA *Normalized Gain*, *Categories* of the students' SR pre- and post-the treatment in the two groups (experimental and control) were statistically described. But as SR is assured in Lawson framework to be multi-dimensional cognitive skill which contains six aspects, and LCTSR measures these six aspects as shown in **Error! Reference source not found.** (Novia &Riandi, 2017); all variables which have been studied and described here were split into and related to these six dimensions.

Participants of this study $(N = 100)$were assigned into experimental and control groups. The *experimental* group consisted of $(n = 49)$ participants and had a starting average SR % of $M = 25.5$, $SD = 14.3$ and ended up with SR % of $M = 41.2$; $SD = 19$. The c*ontrol* group in the other side consisted of $(n = 51)$participant which got an average starting SR % of $M = 36.5$; $SD = 19$ and ended up with a SR % of $M = 45.6$; $SD = 20.4$ (**Figure *12***).

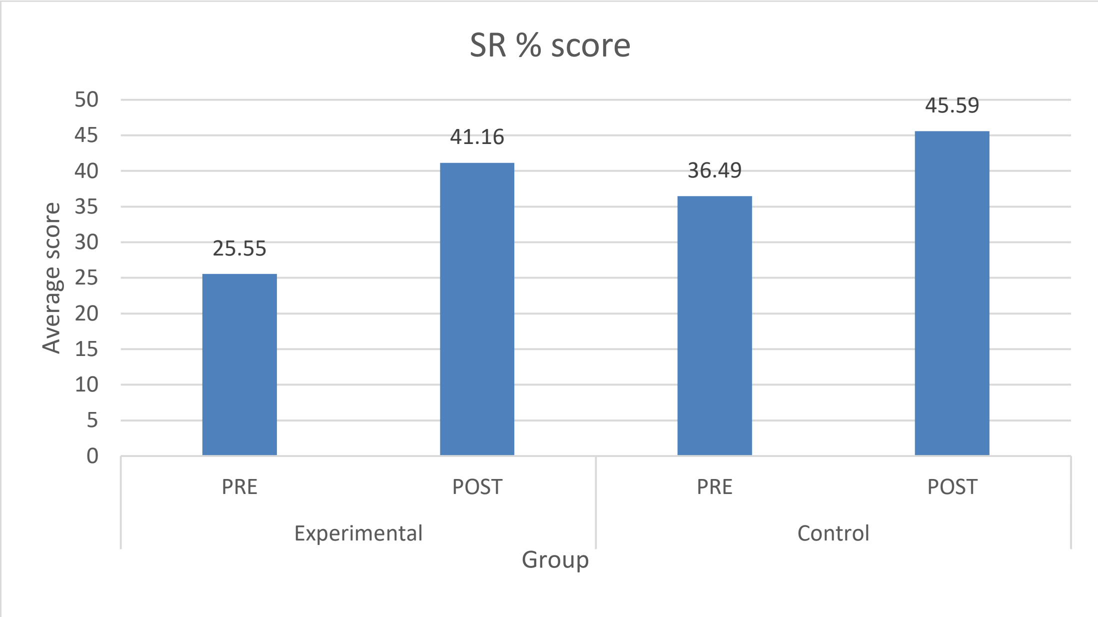


Figure 12. Mean SR % Pre- and Post-instruction

The students' Pre-test% mean scores for each of the six dimensions of SR for the experimental and control groups were presented in Figure 13. It is clear in the chart the existence of difference between the two groups in the pre-test% of each dimension, accordingly we need to perform a test to check if this difference is significant or not to assure validity of conclusions.

The students' post-test% mean scores for each of the six dimensions of SR for the experimental and control groups had been presented in Figure 14.The chart shows difference between the two groups in post-test% of each dimension, accordingly we need to perform further tests to check if there are significant differences in the change or in the growth of each dimension pre-post or not.

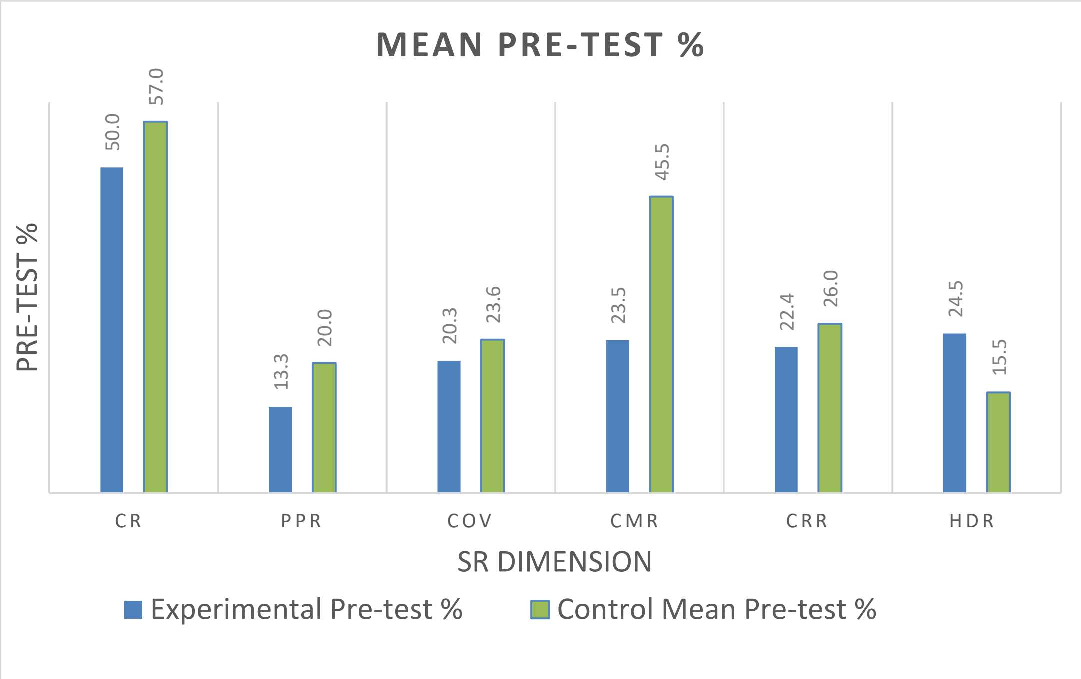


Figure 13. Pre-test Percentage of each SR dimension in the Control and Experimental groups

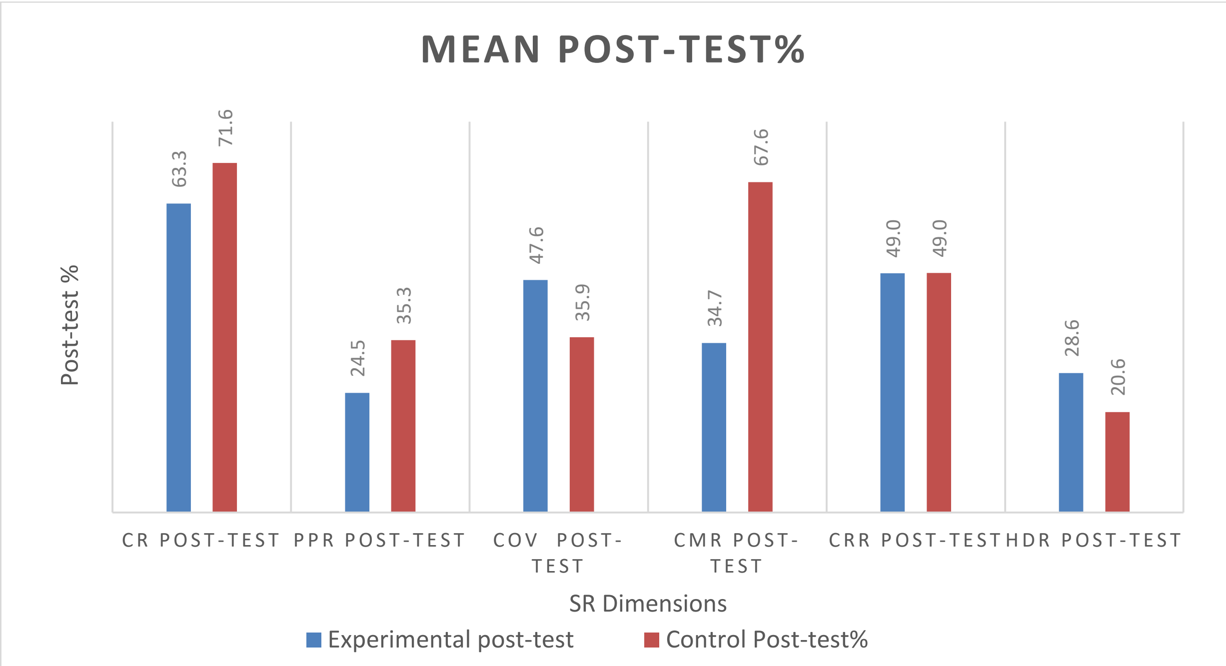


Figure 14. Post-test Percentage of each SR dimension in the Control and Experimental groups

The Post-test scores were compared then to the Pre-test for both the experimental and control groups, in a trial to look for any major impact of teaching the Physics subject regardless of the utilized pedagogy. So, Wilcoxon Signed Ranks test was used, and the results are in Table 13.

Table 13: *Wilcoxon Signed Ranks test for SR Pre/Post comparison*

Group	SR %	M (*SD*)	*N*	*Z*	*p*
Experimental	Pre	25.55 (14.30)	49	-4.984	0.000
	Post	41.16 (18.92)			
Control	Pre	36.49 (18.94)	51	-4.638	0.000
	Post	45.59 (20.35)			

Wilcoxon statistical test indicated that post-test SR scores were significantly higher than pre-test scores in both study groups (Z_{Cont}= -4.638; $p < 0.001$, Z_{Exp}= -4.984; p< 0.001)

As a multidimensional variable, SR dimension post-test were also compered to the dimensions' pre-test while controlling for the starting category. Error! Reference source not found.showed the results of Wilcoxon T est comparing SR dimensions pre-test and post-test.

Table 14: *Wilcoxon Signed Ranks Test – Pre / Post*

SR Dimension	Group	SR %	M (*SD*)	*N*	*Z*	*p*
CV	Experimental	Pre	50.00 (30.6)	49	-2.208 [a]	.027
		Post	63.27 (30.2)			
	Control	Pre	63.73 (28.4)	51	-2.138 [a]	.033
		post	71.57 (32.02)	49		
PPR	Experimental	Pre	13.27 (26.6)	49	-1.807[a]	.071
		Post	24.49 (38.4)			
	Control	Pre	26.5 (39.2)	51	-1.572 [a]	.116
		Post	35.3 (41.6)			
COV	Experimental	Pre	20.33 (27.055)	49	-4.013 [a]	.000
		Post	47.57 (30.56)			
	Control	Pre	26.73 (28.346)	51	-1.999 [a]	.046
		Post	35.9 (29.80)			
CMR	Experimental	Pre	23.47 (34.01)		-1.54 [a]	.123
		Post	34.69 41.1	49		
	Control	Pre	66.67 (40.8)	51	-.164 [a]	.870

		Post	67.65 (42.218)			
CRR	Experimental	Pre	48.98 (50.51)	49	-2.837 [a]	.005
		Post	22.45 (42.16)			
	Control	Pre	29.41 (46.018)	51	-2.500 [a]	.012
		Post	49.02 (50.49)			

a. Based on negative ranks.

Wilcoxon Signed Ranks Test indicated that *post-test%* of the students in the experimental and control groups did show significant statistical difference from *pre-test%* for *CV* (Z_{Exp}= -2.208; *p*<.05, Z_{Cont}= -2.138; $p < 0.05$), *COV* (Z_{Exp}= -4.013; *p*<.001, Z_{Cont}=-1.999; $p = 0.046 < 0.05$) and *CRR* (Z_{Exp}= -2.837; *p* =0.005<0.05, Z_{Cont}=-2.500; $p < 0.05$) dimensions. While for the *PPR and CMR* dimensions the *post-test%* did not differ significantly from *pre-test %*.

Effect size of Wilcoxon test was calculated ($r = \frac{z}{\sqrt{2 \times N}}$) to examine the practical significance of Wilcoxon results and it was ($r_{Experimental}$ = .223, $r_{Control}$ = .211) for *CV*, ($r_{Experimental}$ = 0.41, $r_{Control}$ = 0.20) *COV* and ($r_{Experimental}$=0.29, $r_{control}$= 0.25) for CRR. Accordingly, the answer for $\boldsymbol{q_1}$: "Is there a significant difference between SR dimensions prior to the intervention and the same dimension post the intervention for the two teaching strategies?" is that; both teaching strategies had no significant effect on the *PPR* of the participants, but they had a significant effect on the *CV, COV* and *CRR.* That urge us to focus the following statistical tests on *CV* , *COV* and *CRR* only.

The Post-test scores were compared then to the Pre-test for both the experimental and control groups, in a trial to look for any major impact of teaching the Physics subject regardless of the utilized pedagogy. So, Wilcoxon Signed Ranks test was used, and the results are in Table 13.

Table 15: *Wilcoxon Signed Ranks test for SR Pre/Post comparison*

Group	SR %	M (*SD*)	*N*	*Z*	*p*
Experimental	Pre	25.55 (14.30)	49	-4.984	0.000
	Post	41.16 (18.92)			
Control	Pre	36.49 (18.94)	51	-4.638	0.000
	Post	45.59 (20.35)			

Wilcoxon statistical test indicated that post-test SR scores were significantly higher than pre-test scores in both study groups (Z_{Cont}= -4.638; $p < 0.001$, Z_{Exp}= -4.984; p< 0.001)

As a multidimensional variable, SR dimension post-test were also compered to the dimensions' pre-test while controlling for the starting category. **Error! Reference source not found.**showed the results of Wilcoxon T est comparing SR dimensions pre-test and post-test.

The *change in the six dimensions of SR* after the control and experimental groups experienced the course, was also revealed by calculating the normalized gain (N-Gain). In 2002, Hake applied the following formula to calculate *N-Gain* to individual participant scores (V.P. Coletta & Steinert, 2020):

$$N\_Gain = \frac{postscore\% - prescore\%}{100\% - prescore\%}$$

Figure 15 presents the experienced *N-Gain* in SR dimensions as per the study group. based on the criteria by Hake, (2002): if N-gain ≥ 0.70 it's considered (high); if 0.30 < N-gain < 0.70 it's considered (moderate); and if N-gain ≤ 0.30 it's considered (low) (Erlina et al., 2018).

The N-gain aims to calculate the gain in SR correcting for guessing cases by students. It's clear that highest change among the dimensions in the experimental group happened in the *Control of Variables* dimension ($N_{Experimental}$= 49, M = .342, *SD*=0.36) which was a moderate gain as per criteria by Hake (2002), while the gain for all other aspects in the same group and also in the other group was low M ≤ *.30*.

The highest difference in N-gain was between the experimental and control group in favour of the experimental at the *COV* dimension. It was also clear from the figure that the least difference in the normalized gain was between the two study groups in the proportional reasoning. This comparison needs to be quantitatively analysed deeper in the inferential statistics section.

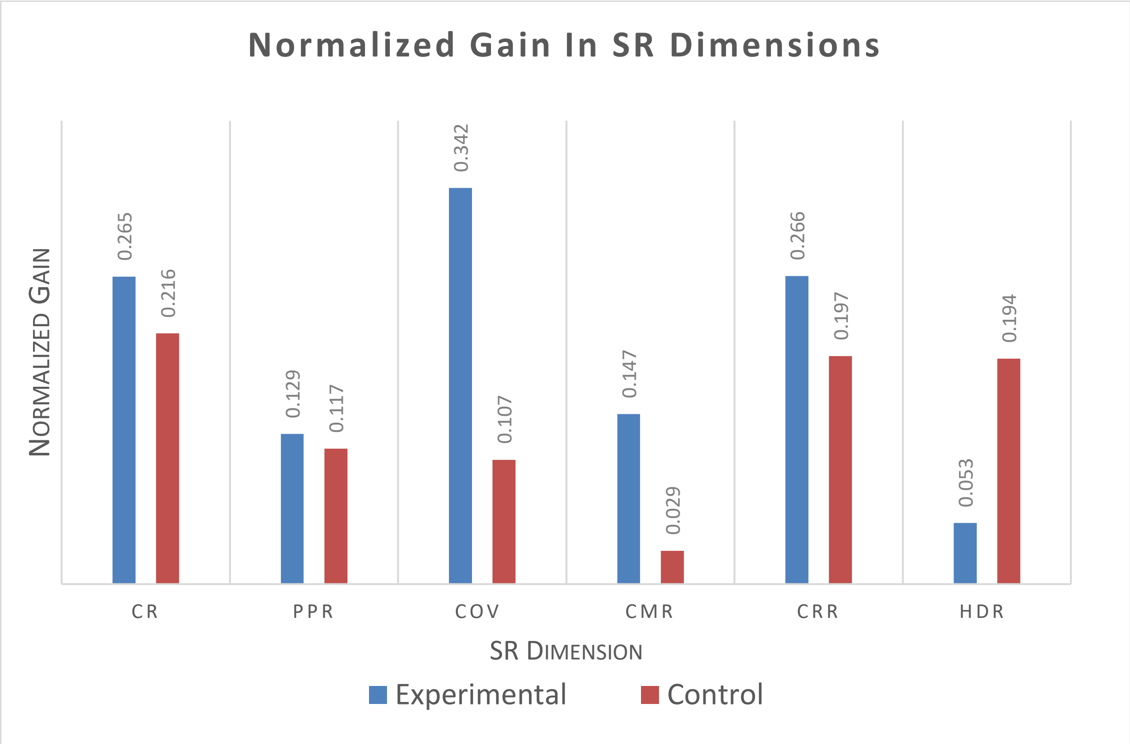


Figure 15. N-Gain of SR dimensions

The *category* of Scientific Reasoning Ability (SRA), where each student fitted in at the end of the semester after the instruction was also studied to check for any impact of the intervention on percentages of categories among the groups, and results were as shown in Figure 16. The figure shows that the control group ended up with 21% (n=11) concrete reasoners, 61% (*n*=31) transitional and 18% (n=9) formal reasoners. While the experimental group had 18% (*n*=9) concrete, 76% (*n*=37) transitional and 6% (*n*=3) formal reasoners.

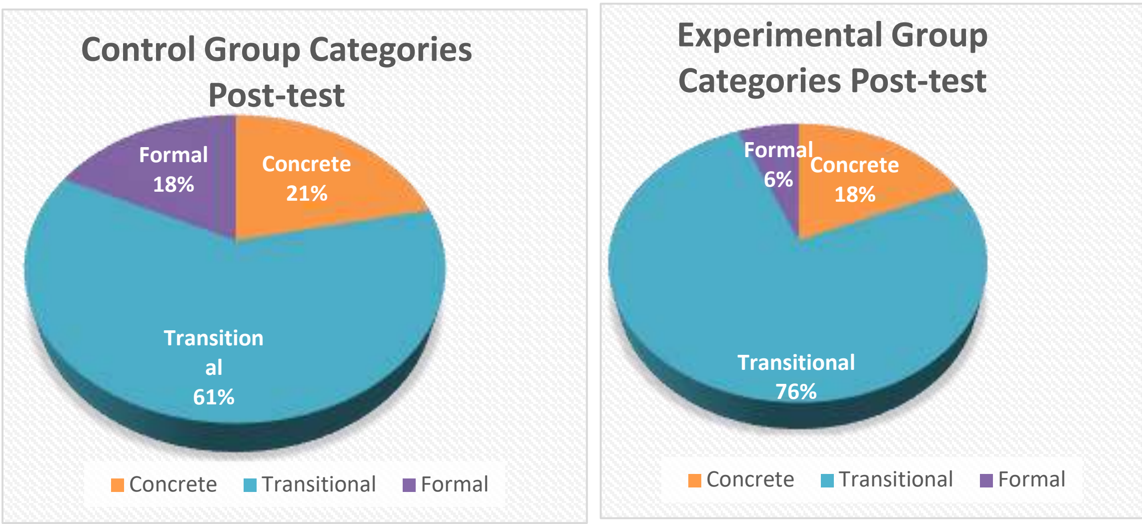


Figure 16. Categories' Percentages Post-test of both groups.

Cross tabulation of frequencies and percentages of students in each level of SR (concrete, transitional, or formal) before (Categ. -PRE) and after (Categ. -POST) the intervention was presented in Table 15. The null hypothesis in this test, H_0: "There is no significant relationship between the starting SR category and the post-instruction category" It indicated that; in the experimental group; 36 % of the concrete SR had moved out to the transitional category and 5% to the formal category. In contrast, in the control group, 36% moved out of the concrete category to the transitional and no one had moved out of this category to the formal. The formal reasoner participant stayed as a formal reasoner. Transitional reasoners mostly stayed in the transitional category in both groups, but in the control group, 12% of those transitional reasoners shifted to the formal category. and formal reasoners are in the ceiling already, so cannot be investigated for category raising.

Table 16. *Contingency table- categories of SR Pre- and Post- intervention.*

Group				Categ.Of.SR-POST				Gamma
				Concrete	Transitional	Formal	Total	
Experimental	Categ.-Pre	Concret	Count	23	14	2	39	.840
			% Within Categ.Of.SR-PRE	59.0%	35.9%	5.1%	100.0%	
		Transitional	Count	0	10	0	10	
			% Within Categ. -PRE	0.0%	100.0%	0.0%	100.0%	*P* = .000
	Total		Count	23	24	2	49	
			% Within Categ. -PRE	46.9%	49.0%	4.1%	100.0%	
Control	Categ.-Pre	Concrete	Count	16	9	0	25	
			% Within Categ.Of.SR-PRE	64.0%	36.0%	0.0%	100.0%	
		Transitional	Count	1	21	3	25	.961
			% Within Categ. -PRE	4.0%	84.0%	12.0%	100.0%	*p* = .000
		Formal	Count	0	0	1	1	
			% Within Categ. -PRE	0.0%	0.0%	100.0%	100.0%	
	Total		Count	17	30	4	51	
			% Within Categ. -PRE	33.3%	58.8%	7.8%	100.0%	

The null hypothesis H_0 may now be rejected (p<.001) with an effect size shown in Table 16 indicated practically significant results in both the experimental (Gamma = .840) and the control groups (Gamma = 0.961). But the percentages in the table does not show a significant relation between the *category shifting* and the teaching approach. In other words, there is no noticeable difference between the Explore-Instruct and Instruct-solve approaches in lifting of categories.

4.1.5 Inferential statistical analysis:

Inferential statistical analysis had been carried out to study the inferences that may be made about the population targeted in this study, depending on data collected from a sample of that population. The hypotheses set were:

$\boldsymbol{H_1}$: There is significantly higher scores of SR Post-test than Pre-test for all participants in both the control and experimental groups".

$\boldsymbol{H_2}$: There would be a significantly higher Normalized Gain in all dimensions of SR in the experimental group than in the control".

$\boldsymbol{H_3}$: There is a significantly higher tendency of improving the category of SRA in the experimental group than the control group".

Studying the previously set hypothesis would help in answering the first main question in this study "Is there a difference in college students' scientific reasoning ability (SRA) development if taught by explore-instruct teaching approach compared to the instruct-solve approach (traditional approach) in physics?" so this section will be outlined as follow.

First, a decision on types of tests (*parametric or non-parametric*)to accomplish the objectives of this inferential analysis section is to be made; data distribution results of variables considered in this inferential section was revisited in **Table *7***. It was clear in **Table *7*** that the data distribution was not normal for each of the following variables to be studied in this section; *COV Reasoning Pre-test*, *COV Reasoning Post-test* and *N-Gain in COV Reasoning* ($p \leq 0.05$). Accordingly, non-parametric tests would fulfil the numerical data analysis in this inferential statistical analysis section.

Second, the two study groups (Experimental & Control) are to be tested for a similar start. Because starting from the same point assure the *inferential statistics' conclusions validity*, which means the noticed effect in the dependent is due to the independent factor and not to a different starting condition (Johnson & Christensen, 2014).

Third, the inferential statistics which answer sub-questions and validate or reject hypotheses that led to answer the first main question of this study were performed. First sub-question in this context was q_1:"Is there a significant difference between SR dimensions prior to the intervention and the same dimension post the intervention?". *Wilcoxon test* was performed in the descriptive statistics section, to examine the SRA at its six different dimensions in each group post the intervention, there found a significant difference exist Pre-Post the intervention in *CV*, *COV* and *CRR*; the researcher is to investigate in this inferential statistics section if that difference is similar in case of following the Explore-Instruct teaching strategy or when teaching via the Instruct-Solve strategy.

So, the researcher will study the difference in the *Normalized Gain* of the *CV*, *COV* and *CRR* dimensions of SR in the experimental and control groups using *Mann-Whitney U test* to answer sub-question *2* (q_2): "If the SR-dimension experienced a significant change post the intervention; is that change significantly different between the control and experimental groups?"

Regression test was performed as a last inferential analysis here to estimate percentage-contribution of normalized gain of the individual SR-dimension (the ones which experienced a change due to the Explore-instruct strategy) to the associated normalized gain in total SRA resulted due to Explore-instruct approach under investigation (Johnson & Christensen, 2014). Finally, inferential statistics results are to be summarized concluding this results section.

Starting with *Validity of inferential conclusions* which is to be assured at the beginning of any inferential statistics section. Because as per Johnson & Christensen (2014), purposive assignment or non-random assignment of students who were already enrolled in classes, threaten the validity of inferential conclusions. So, the Pre-test data of SR is to be compared between the control and experimental groups, and as agreed on the non-parametric tests; Mann Whitney U test is to be used for this purpose.

In **Table *17***, Pre-test% of SRA in the experimental group ($M = 25.6\,, SD = 14.3$) was lower than it in the control group ($M = 36.5\,, SD = 17.6, N = 51$). Mann-Whitney U test results indicated that the difference was statistically significant U($N_{Experimental} = 49, N_{Control} = 51$) = 816.500, Z=-3.021 , p = 0.003.that difference forbids the researcher from relating any noticed impact on the dependent (N-Gain in SRA) to the independent variable (Group = teaching strategy).So, both groups did not start at the same level of SRA.

Table 17: *Mann-Whitney U test of Pre-test SR per groups*

Variable	Group	N	M	*S.D*	Mann-Whitney U	*p*
Pre-test%	Experimental	49	25.55	14.299	816.500	
	Control	51	36.49	18.943	Z= -3.021	.003
	Total	100	31.13	17.623		

But as *SRA was a multi-dimensional* variable, and LCTSR was a heterogeneous test; each dimension of SR, need to be compared first before the treatment, between the experimental and the control groups. That step will ensure validity of statistical conclusions related to the specified dimension. Accordingly, Mann-Whitney U test was performed this time to study the beginning of each single dimension of SR between the experimental and control group. Deciding then which dimension can be trusted for inferential conclusions.

Table18: *Mann Whitney - Pre-test SR dimensions between groups*

SR Dimension	Group	N	Mean Rank	Mann-Whitney U	*p*
Conservational Reasoning (CV) Pre-test	Experimental	49	44.79	969.500 *	
	Control	51	55.99	Z= -2.241	.025
	Experimental	49	46.58	1057.500	.097

Proportional Reasoning (PPR) Pre-test	Control	51	54.26	Z= -1.661	
Control of Variables Reasoning (COV)Pre-test	Experimental	49	47.06	1081.000	.204
	Control	51	53.80	Z= -1.269	
Combinational Reasoning (CMR) Pre-test	Experimental	49	36.77	576.500 *	.000
	Control	51	63.70	Z= -4.962	
Correlational Reasoning (CRR) Pre-test	Experimental	49	48.72	1162.500	.430
	Control	51	52.21	Z= -.790	
Hypothetical-deductive Reasoning (HDR) Pre-test	Experimental	49	57.33	915.000 *	.000
	Control	51	43.94	Z= -3.394	

Note. Grouping Variable: The study group

Table18 showed that dimensions of Control of Variables reasoning and Correlational reasoning prior to the instruction were not significantly different between the control and experimental groups (p > 0.05), while the other dimensions including Conservational reasoning were significantly different for the two study groups (p < 0.05). But it was found in the difference among demographic section that gender, age and starting SR category influences the starting SRA of the participants. As per Barnett and colleagues (2015), for research with control and experimental groups, ANCOVA can be used to control regression of the mean, to eliminate then any confounding effect while comparing between the control and experimental groups MANCOVA test was run to study the impact of each independent variable on the dependent variables of the six SR dimension while controlling for the other independent variables. The results are in Table 19

Table 19: *MANCOVA- Tests of Between-Subjects Effects*

Source	Dependent Variable	df	Mean Square	F	*p*	η^2
Group	CV Pre-test	1	44.268	.063	.803	.001
	PPR Pre-test	1	8.321	.009	.926	.000
	CoV Pre-test	1	1191.802	2.937	.092	.050
	CMR Pre-test	1	2690.104	1.955	.168	.034
	CR Pre-test.	1	2709.816	2.121	.151	.036

	HDR Pre-test	1	3936.517	10.955	.002	.164
Age	CV Pre-test	1	389.395	.550	.461	.010
	PPR Pre-test	1	2.141	.002	.963	.000
	CoV Pre-test	1	39.654	.098	.756	.002
	CMR Pre-test	1	377.691	.275	.602	.005
	CR Pre-test.	1	3311.292	2.591	.113	.044
	HDR Pre-test	1	12.705	.035	.852	.001
Gender	CV Pre-test	1	849.162	1.199	.278	.021
	PPR Pre-test	1	2280.733	2.370	.129	.041
	CoV Pre-test	1	884.617	2.180	.145	.037
	CMR Pre-test	1	4.229	.003	.956	.000
	CR Pre-test.	1	1679.621	1.314	.256	.023
	HDR Pre-test	1	2.596	.007	.933	.000
Categ.Of.SR PRE	CV Pre-test	1	8509.642	12.017	.001	.177
	PPR Pre-test	1	29610.191	30.774	.000	.355
	CoV Pre-test	1	16453.842	40.541	.000	.420
	CMR Pre-test	1	22066.344	16.038	.000	.223
	CR Pre-test.	1	34365.013	26.892	.000	.324
	HDR Pre-test	1	1933.082	5.379	.024	.088

The multivariate test in Table 19 revealed the significant difference between the two study groups in the starting HDR F(1, 3936.52) = 10.955, P<.05) , Wilks' =.781, η^2 = .164. That means all dimensions except *HDR* can be studied with trusted inferential conclusions. But we are interested in studying *CV* , *CoV* and *CRR*.

MANCOVA test also revealed that, the starting dimensions of SR are significantly different among the different starting categories of SR $P < .05\, , Wilks' = .188$. Therefore, the following tests will be performed for *the five dimensions other than HDR* while *controlling* the *starting category*.

Normalized Gain *N-Gain:*

After descriptive statistics section and at this stage, we need to study the difference in "Normalized Gain in CV, COV and CRR":

1. *Between the experimental and control groups*

Normalized Gain in SR (N-Gain) is a variable resulted from the difference between Post-test% and Pre-test% divided by hundred minus the pre-test% and it represents the SRA Growth of the student with correction for the guessing problem in answering the MCQs of LCTSR (Hake, 2002). This variable will be analysed further to study the SRA Growth more. But as *N-Gain* is not normally distributed (***Table 7***); Mann-Whitney U test will replace the t-test and Kruskal-Wallis test will replace ANOVA in analysing differences between groups in the SRA normalized gain. And in studying correlation between this dependent variable and other variables in the study; Spearman correlation will replace Pearson correlation (Huizingh 2007; Ahmad et al., 2018).

To answer the second sub-question in this section ***q₂***: "In case the SR-dimension experienced a significant change post the intervention; would that change significantly differ between the control and experimental groups?" the null hypothesis here H_0 will state "students population taught via Explore-instruct and the ones taught via Instruct-Solve have the same N-Gain in CV, CoV and CRR dimensions of SR". and MANCOVA test controlling the different demographic factors and the starting SRA revealed the following.

Table 20: *MANCOVA test*

Source	Dependent Variable	df	Mean Square	F	p	η^2
	N_Gain CV	1	.142	.531	.468	.006
Group	N_GainCOV[a]	1	2.80	14.5	.000	.132
	N_Gain CRR	1	.158	.773	.381	.008
	N_Gain CV	1	.005	.017	.895	.000
Categ.Of.SRPRE	N_GainCoV[a]	1	.073	.377	.541	.004
	N_Gain CRR	1	.008	.041	.840	.000

a. R^2 = .161, (Adjusted R^2 = .126)

According to Table 20, the null hypothesisH_0 is to be rejected for the N-Gain of *COV*F(1, 2.80) = 14.5, p < .001) between the control and experimental groups . But the null hypothesis $H_0$is to be accepted for CV and CRR, proving that while controlling for the starting category of SR and most demographic factors, the only aspect of SR which its N-Gain significantly differ between the experimental and control groups, is the *COV*. The value of partial eta square ($\eta^2 = .132$) shows that the difference in teaching approach explains 13.2% of the difference in *N-gain* between the control and experimental groups. So, the answer to sub-questions q_2 is "Yes, there is a significant difference between the two groups in N-Gain of COV".

After noticing a significant difference between the two study groups in the *N-Gain of COV*, the strength of the relationship of that *N-Gain of COV* with the followed teaching approach needs to be investigated, keeping in mind, controlling for the starting category of SR which has been considered a confounding variable according to MANCOVA ***Table 19***. That can be studied by performing Partial correlation.

Partial correlation test has been performed between variables of the teaching strategy and all aspects of SR. *Spearman's correlation* coefficients are presented in **Table 21** below.

Table 21: *ANCOVA - N-Gain in COV& the teaching strategy*

Control Variables			N_Gain in COV
Categ. Of.SR-PRE & Post group & Gender	Teaching strategy	r_s	-.364
		p	.000
		df	95

Table 21 indicated that *COV* dimension found to negatively correlate in its N-Gain to the followed teaching approach ($r_s(95) = -.364,\ p < .001$), Negativity comes from the labelling numbers of the teaching approaches (1: Explore-instruct, 2: Instruct-solve).

The effect size indicator of the correlation coefficient square (Johnson & Christensen, 2014) "r_s^2 = $0.364^2 = 0.132$" was calculated. Which means 13.2% shared variance or overlap between the *N-Gain of COV*, and the followed teaching strategy was practically proved. Assuring by that the strength of the correlation between the normalized gain in COV pattern of SR and the Explore-Instruct teaching approach while controlling the category,

Categories of SR prior to the test (Categ.SR-Pre) and Categories of SR post the test (Categ.SR-Post) are ordinal values which are also not normally distributed($Sig. < 0.05$) as presented in ***Table 7***So, Mann-Whitney U test will be performed replacing independent t-test to study the difference in starting SRA category, between the experimental and control groups to ensure inferential conclusion validity of results related to categories.

Results in ***table22***; proved that there is a significant difference in the beginning categories of SRU($N = 100$) = 862.500, Z= -3.201, p = 0.001 between the experimental and control groups with the mean rank of the control (58.09) higher than that of the experimental (42.60), with effect size of (r $=\frac{z}{\sqrt{N}} = \frac{3.201}{\sqrt{100}}$) 0.32.

Table22: *Mann-Whitney Test comparing beginning Category.*

	Group	N	Mean Rank	Sum of Ranks	Mann Whitney U[a]	p
Categ.Of.SR-PRE	Experimental	49	42.60	2087.50	862.500	
	Control	51	58.09	2962.50	Z = -3.201	.001
	Total	100				

a. Grouping Variable: Group

But it has been revealed in ***Table 20*** that ,N-Gain on COV showed no significant difference between the different pre-instruction Categories.

The aim at this point of the study is to investigate the effect of the teaching strategy on the aspect of *Control of Variables* as well as on SR categories distribution. As COV found to be the only among the other SR' aspects to show a significant growth related to the treatment. *N-Gain in COV.* was noticed when compared between the different *beginning categories of SR while controlling for the group in **Table 20** and if it was similar for all the categories*, one can expect a significant impact of the treatment on any student regardless of his/er category at the beginning of the course. Results in ***Table 20*** assured that in any of the teaching approaches; there is no significant difference ($p >$ 0.05) in the *normalized gain of COV* between the different pre instruction categories. That result of similar *N-Gain of COV* among the categories (controlled variable), encourage the researcher to investigate the effect of the treatment (Group) on a larger frame, which is the effect on categories distribution that accompany the *N-Gain of COV*

.

Linear R*egression Analysis (LRA)* was the last inferential test *which* was performed to study the relation between *Post-COV* and *Pre-COV* in each study group separately.

Table 23: *Linear Regression- Post COV& Pre COV*

Group	Variable[a]	B	95% CI	Beta	t	*p*
Experimental[b]	(Constant)	42.335	[31.48, 53.18]		7.850	.000
	COV Reasoning Pre-test	.258	[-0.065, 0.580]	.228	1.606	.115
Control[C]	(Constant)	23.382	[12.91, 33.85]		4.487	.000
	COV Reasoning Pre-test	.468	[.198, -.739]	.446	3.485	.001

a. Dependent Variable: COV Reasoning Post-test
b. Adjusted R^2 = .032
c. Adjusted R^2 = .182

Results in Table 23 revealed that *pre-test COV* was a significantly high predictor of *post-test COV* (β = .468, p = .001) in the control group F(1,49) = 12.14, explaining around 18.2% of the post-test COV in that group. In contrast, pre-test COV in the experimental group; was not a significant predictor (β = .258, *p* = .115) of the post-test COV F(1,47) = 2.58, accounting for only 3.2% of the variance in post-test COV.

Linear regression analysis was also performed to study the relation between *teaching strategy* and *N-Gain in COV.*

Table24*: Linear Regression- N-Gain and Strategy*

Variable[a]	B	95% CI	Beta	t	*p*
(Constant)	.567	.144		3.924	.000
Teaching strategy[b]	-.257	.091	-.275	-2.833	.006

a. Dependent Variable: N_GainCOV
b. Adjusted R^2 = .066

Test results in ***Table24*** revealed that *the teaching strategy* was a high significant predictor of *N-Gain in COV* (F(1,99) = 8.024*,* β = -.257, *p* = .006), accounting for around 6.6% of the variance in *N-Gain in COV* between the two groups.

4.1.6 Summary of results of the quantitative part- question 1.

Descriptively

- Consistency of results was evident in the studied population with an α value equal to .73 for the total test. And $\alpha > .67$ for each of the six SR dimensions.

- Causal effect of the independent variable (the teaching approach) was evident for the dependent variable (SR). That was done by Investigating the beginning SR of the two groups to guarantee the causal effect.
- Gender, Age, Nationality, English language level and starting category or level of SR were investigated for interactive impact on the pre-test% using MANCOVA. MANCOVA indicated that, the starting category of the participants was the only factor that showed impact on all dimensions of SR except HDR , so it was the confounding variable that needs to be controlled when performing the inferential statistics.
- The starting HDR was the only dimension of SR that is significantly different between the control and experimental groups.
- *The students' Pre-test% and Post-test%* showed significant difference from each other in both study groups, indicating occurring of change in both groups.

- *Normalized gain in (PPR, COV, CMR and CRR)* was studied and compared between the two studied group and the two groups had a significant difference in N-Gain in COV dimension only. N-Gain in COV found to be significantly higher in the experimental group than in the control group. While similar *N-Gain of COV* found in the three categories when controlling the group.
- Correlation test showed significant association between the N-Gain in COV and the group.
- Distribution of students among the different stages of SR and their mean score in each category after experiencing the two teaching styles have also been analysed, and the concrete category experienced the same amount of shifting toward higher categories in both groups, but the transitional and formal categories did not experience such shifting. Transitional and formal students mostly stayed within the same category after instruction. But there was no distinct impact on category distribution by the Explore-Instruct approach compared to the traditional Instruct-Solve approach.

Inferentially

Regression analysis took place to study the relation between *Post-COV* and *Pre-COV*. and, between the *teaching strategy* and *N-Gain in COV*. The researcher proposed three hypotheses at the beginning of this study to answer the first study question.

The alternative hypotheses set were $\boldsymbol{H_1}$: "There would be a significantly higher scoring in all dimensions of SR post-treatment than pre-treatment in both the EI and the traditional IS group". Results validated the acceptance of this hypothesis. $\boldsymbol{H_2}$:"There would be a significantly higher Normalized Gain in all dimensions of SR in the EI group than in the traditional IS group". That hypothesis could be validated only for *COV*. For the rest dimensions of SR there was no significant difference between the two stud groups. The third hypothesis $\boldsymbol{H_3}$**: "**There is a significantly higher tendency of moving out of any category to a higher category of SR in the EI group than the IS control group." That tendency was found only for the Concrete category but was not higher in the EI than in the IS group, both groups experienced similar concrete shifting tendency.

4.2 Results of Question 2- Email interviews with students.

It was found in a quasi-study by Loehr, Fyfe and Rittle (2014) that the absence of the phase of knowledge application after the instruction phase during the lecturing session may remove potential benefits of the explore-instruct teaching approach. This finding by Loehr, Fyfe and Rittle (2014) was also proved by She and Liao (2009) in a quantitative study, for grade eight students in Taiwan. She and Liao conducted an experimental design study to test the impact of the instruction approach and the effect of the level of the scientific reasoning of the students, on their achievement and scores in three tests which were provided by the researchers to measure their scientific reasoning, concept construction and conceptual change. And they found that the gained knowledge from the delayed lecture should be implemented directly in the same session for the SR and conceptual change to be enhanced. So, such findings by such quantitative studies are expected to be generalized and it was considered in the context of the current study. To reveal such conditions in the current study, qualitative part proposed by the researcher to take place after the quantitative part, aiming to reveal more data about the optimum conditions for the new explore-instruct approach. In addition to exploring as many implicit factors as possible that affect the targeted impact of the proposed technique; qualitative paradigm, in the form of semi-structured interviews had been used beside the quantitative one.

This section analyses students' responses via email interviews to answer the second research question "What are the perspectives of college students regarding the explore-instruct approach after experiencing it?" .After collecting the quantitative data of this study, four students in the experimental group; two with low normalized gain in SR, and two with medium normalized gain in SR were interviewed. Qualitative data was collected via emails' interviews to support the formerly collected quantitative data and to help revealing perceptions of students' group that received the intervention. The researcher then used Thematic coding was used for this analysis as this type of coding is the appropriate to the purpose of "analysing the perspectives of the group which received the intervention" as per Flick (2014). Hsieh and Shannon (2005) presented three forms of content analysis: summative, directed, and conventional. The three forms differ in their sources. Directed content analysis has a deductive nature; it may validate a theory or extend it. The conventional content analysis starts with no pre-defined theory and the aim is to define a new theory. The summative content analysis starts with counting the occurrence of a word and then tries to understand the meaning of the word through analysing the hidden content.

Questions of the interview were set by the researcher after reviewing the available literature, and after the quantitative analysis of SRA pre-post the intervention. The invited interviewees were six participants, two students from the high SR normalized gain, two from the medium and two from the low normalized gain in the experimental group, but only 4 students replied to the emails. The researcher was in a different city and the only contact mean with the students was the email.

the researcher analysed each interview in brief and discussed the proposed themes: Teaching approach and engagement domain, Problem-solving activities domain, Student's stress, and motivation domain. The process of validation of these themes included reviewing qualitative parts of related literatures (Erlina et al., 2018; Koenig et al., 2019; Wong & Kowitlawakul, 2020) in addition to the researcher expertise, who used to verbally ask the students in the experimental group at the end of each session about their perceptions and opinions on the new EI teaching strategy, in a form of - informal conversational interview - as classified by (Johnson & Christensen, 2014). It has been observed at the beginning of the term, that the students do not accept the idea of being asked to solve a problem or answer short questions before having the lecture. Few sessions later in the term, students adapted to the pedagogy but there appeared a concern about the grades as it impacts the students' GPA.

The interviews conducted by the researcher attempted to follow the feature for semi-structured interviews shown by Flick (2014), it starts by open-ended questions to be at the beginning of the interview and to be answered based on the knowledge which the interviewee has at hand. Following the open-ended questions are the hypothesis-directed questions in which the interviewer gives his assumptions as offers for the interviewees, either they take it up or reject it. Together with the previous two types of questions come the confrontational questions at the end to validate the participants' perspectives.

Creswell (2013) stated that "the qualitative research process is emergent which means that the initial research plan put forth by the researcher cannot be tightly prescribed, and that all phases of the process may change or shift after the researchers enter the field and begin to collect data" (p. 47). The data collection methodologies may change, questions may change, and the interviewees may be modified throughout the study (Creswell 2013). Supported by this statement, the researcher needed to replace the in-depth semi-structure interviews with email interviews, and to further interpret and validate the responses of the interviewees, the responses to a one-page survey adopted from (Örnek et al., 2008) administered to a group of 25 students from the same population during the piloting phase, became data to reflect upon during the qualitative data analysis process. However, these 25 students did not participate in the main study. In the study by Örnek and others (2008) it appeared that students' perceptions of reasons of difficulty of introductory physics are different for the instructors' perceptions, but they are similar to the

Teaching Assistants' (TA s) perceptions. The cause of difficulty agreed upon by students, TA s, and instructors as per (Örnek et al., 2008) is that students don't study more. In the current study, in an exit questionnaire about the most factor contributing to the difficulty of Physics given to a sample of 24 students who were not part of the real study but are from the same population of the study, 46% of students rate the student-controlled factor "Do not practice many problems and work only assigned problems" as.

4.2.1 The interviews

The students from the experimental group had been contacted a year later after passing the course, they had been interviewed to investigate the features of Physics course they still remember and think it had impact on their learning. The structure of the interview questions followed the semi-structured interviews protocol which according to Uwe Flick, (2014) begins with open-ended questions, followed by hypothesis-directed questions, and conferential questions which leave a space for the interviewee's perspectives to be revealed. Questions set to reveal their opinion on the tried teaching strategy which they experienced in the physics course, their experience in the Lab, and the confrontation 2-choises question was to reveal in which case do they think they had been able to understand the concept or the principle of the experiment more, If they conduct the experiment (perform it by following the steps provided in the sheet or the manual) then receive the lecture from the instructor? Or, if they fully had the lecture by the instructor and then have the experiment? The last question for the interviewees was, about their opinion(hypothesis-guided open-ended question).;it reveals if they feel comfortable with the Explore-instruct strategy? and if they agree with the view that states: "Explore-instruct approach ultimately clarifies the concept more than the instruct-solve (conventional) strategy, but the student doesn't feel comfortable in conducting the experiment before being instructed"? and finally the interviewees were asked if there was any point they would like to mention about explore-instruct teaching strategy?

4.2.2 Identified Themes across Interviews

Given that the sample size was only four students who responded to the emails, the researcher did engage in a manual coding following the way of other researchers (Bottge et al., 2009; Braun & Clarke, 2006) and based on that, established three themes. using an Excel spreadsheet, an analysis matrix was formed for each question (**Table *25***). First two open ended questions were placed at the left column of the matrix, and respondents' levels of N-Gain were indicated on the second column of the matrix. Basic statements have been extrapolated and inserted into the designated groups on the matrix. Then the patterns, frequency of statements, and emerging themes, were denoted by the researcher.

The data collected have been categorized as per the SR development level of the student. Responses to *the first interview question* of "What are the features of the Physics course you remember that had impact on your learning?" were as follow:

Students with low N-Gain in SR replied by highlighting the main resources the student think were helpful in understanding the subject such as the work sheets been solved in and outside the class, and resources that attracted and motivated the students to think and interact in the class such as KAHOOT website which was used instead of the clickers to allow students to answer questions using their mobile phones or laptops at the very beginning of the lecture. The questions in KAHOOT were all related to the upcoming non-studied yet topic. While *Students with medium N-Gain in SR* responded that they learned lots of new knowledge related to many concepts in Physics.

Responses to *the second interview question* of "What do you think of the teaching style you have experienced in the physics course? Explain the pros and cons." were as follow:

Two students in the *low N-gain level*, responded differently to this question. One thought that it helped them to experience and approach the learning outcomes earlier before being instructed and guided by the instructor, so it was helpful in making the lesson easier to be understood. The other student believed that the IS approach was better than the EI in carrying up the experiments, but in the lecture, exploring the problems first and then being instructed was preferred, but with the condition of having enough number of problems available for exploration before and after the lecture.

Students with medium N-Gain in SR replied by emphasising that most students alienate from physics, and there is a responsibility on the instructor to deliver the subject in more simple approach than the EI, to be understood. They also urged exposing the students to more problems and more activities. These students preferred having the lecture first before the laboratory, because they found it hard to write the lab report with the EI approach.

Responses were as follow to *the third confrontation 2-choices questio*n "In the Lab, when do you think you understood the concept or the principle of the experiment more? a) If you conduct the experiment (do it from the steps in the sheet or the manual) then receive the lecture from the instructor? Or b) If you had the lecture fully by the instructor and then have the experiment.":

The low N-Gain students (2 interviewees) had chosen taking the experiment first before the lecture which go over the concept explored in that experiment (Explore-Instruct), while *the two medium N-Gain students* (2 interviewee) had chosen to take the lecture before the experiment (Instruct-Solve).

Responses *to the last (hypothesis-guided open-ended) question* "write your opinion: Did you feel comfortable with the Explore-instruct style? What do you say about the statement "it clarifies the concept better (than the instruct-explore style) at the end, but you don't feel comfortable while you try to conduct the experiment before being instructed"? What else can you say about this style of teaching?" were as follow:

Students with low N-Gain preferred to have the Experiment before Lect. because they think they will practically experience what is happening, and then the lecture will stress on what did the results proven, one of these students stated "Explore then instruct is more beneficial and efficient to the students, because student's thinking will be wider and deeper. Furthermore, the first deep information will stay on our minds for long time. And linking between information in Explore- instruct was strongest than the other styles."

Students with medium N-Gain stated "I believe that having the experiments after the lecture keep you on the understanding path. Experiments helps in more understanding of the points given in the lecture. I know explore instruct style is more exciting for those who love physics, but I will talk about the vast majority students who don't like to do things without understanding why they are doing. Knowing what you are doing and what you are measuring and the ways you do so helps most of the students to be fully aware of all information given in theory classes." ,"From my point of view that method of teaching which depends first on the lecturer then the attempt of the experiment is more useful because the lecture first engages the mind in thinking process and improve the scientific imagery.

Table 25: *Responses' brief of students to questions in the interview.*

Interview Questions	Level of N-Gain of the interviewee.	Responses of the interviews
1"What are the features of the Physics course you remember that had impact on your learning?"	Low N-Gain	• **Work sheets** were very helpful. • Course was **very easy**, because the **instructor made it too simple** to us. • "I remember that **we memorize the definitions and main points** by **using KAHOOT application which make all of us attract** to the lesson".
	Medium N-Gain	**"learned many** from the information in the physics course **and new definitions** on basics in magnetism and laws of Newton **and many new concepts"**
2 "What do you think of the teaching style you have experienced in the physics course? "Explain the pros and cons.	Low N-Gain in	• Doing the experiment before the lesson helps us **to know the result and to confirm that before having it taught** will make the lecture easier. But taking experiment after the lesson will cost **a lot of time.** • "Through scientific experiments I benefited a lot of the new concepts and in my opinion the explanation first then the attempt is good for establishment of the information".

		• "Attempts to solve questions then the lecture, is good for applying what has been learned and in my viewpoint **enough questions** should be available related to each concept in the lectures".
	Medium N-Gain .	• "I do love physics and I loved the way of teaching. However, **most of student aren't fans of this subject** so the **class need to have more simple way of explanation** for the material given. Moreover, class need to have **extra activates** and teachers need to **motivate the student** to learn more. **Solving more equations helps in understanding the concept** of the formulas and it helps in understanding the meaning of some terms. On the other hand, if a student **understands a lecture, it helps a lot in lab sections** but if **not, there is a chance in lab sections to understand the lectures in a practical way".** • **"It was new style of teaching for me**, it made me **think more while doing the experiments** to achieve full understanding of the concept of the experiment. **The negative** thing about it I think, I found **it hard to write the Lab. Report, it took me long time and sometimes I find myself missing some data after leaving the lab**.".

3In the Lab, when do you think you understood the concept or the principle of the experiment more? a) If you conduct the experiment (do it from the steps in the sheet or the manual) then receive the lecture from the instructor? Or b) If you had the lecture fully by the instructor and then have the experiment." :	Low N-Gain.	• Both responders selected (a- Explore-Instruct approach)
	Medium N-Gain.	• Both selected (b- Instruct-solve approach)
4Write your opinion: Did you feel comfortable with the Explore-instruct style? What do you say about the say "it clarifies the concept better (than the instruct-explore style) at the end, but you don't feel comfortable while you try to conduct the experiment before being instructed"? What else can you say about this style of teaching?	Low N-Gain.	• "In my opinion, Experiment before Lect. Is better because I will practically experience what is happening and then the lecture will stress on what did the results proven". • "Explore-instruct is more beneficial and to the students, because **student's thinking will be wider and deeper"**.
	Medium N-Gain.	• "I believe that having the experiments after the lecture keep you on the understanding path". • "First the lecturer then the attempt of the experiment is more useful because the lecture first engages the mind in thinking process and improve the scientific imaginary."

Based on the collected responses of the interviewed students, the researcher proposed the following themes: *Teaching approach and engagement* theme, *Problem-solving activities'* theme, and *Student's anxiety and motivation* theme.

These qualitative data which was collected to study the perspectives of the students; revealed that the students were aware of the impact of the teaching approach on their learning. The researcher infers the preference of Explore-instruct teaching approach by some students only if there is no work to be submitted for assessment. But students in general preferred the traditional approach when lab reports or any post-section work to be completed and submitted for formal assessment as they feel less stress and spend less time in completing and turning in the assessed work. So, the students do not trust their skills in completing assessed work with the new teaching approach withing the available time.

CHAPTER 5

DISCUSSION, LIMITATIONS, RECOMMENDATIONS AND CONCLUSION

This study was proposed to investigate the impact of *Explore-Instruct* teaching approach on college students' scientific reasoning. Population was undergraduates of scientific collages in a private university at the United Arab Emirates. Two research questions were set up,

RQ1."Is there a difference in college students' scientific reasoning ability (SRA) development if taught via Explore-instruct teaching approach compared to the Instruct-solve approach (traditional approach) in physics?"

RQ2. "What are the perspectives of the college students and instructors regarding the explore-instruct approach after experiencing it?"

Theories and literature in this topic, encouraged the researcher to expect higher development in SR when teaching via Explore-Instruct teaching strategy, and set the following three hypotheses and their alternatives***:***

H_{01} : There is no significant difference between SRA Post-test and Pre-test for all participants in both the control and experimental groups.

H_1 : There is significantly higher scores of SR Post-test than Pre-test for all participants in both the control and experimental groups.

H_{02} : There is no significant difference in all dimensions of SR between the experimental and the control group.

H_2 : There would be a significantly higher Normalized Gain in all dimensions of SR in the experimental group than in the control.

H_{03} : There is no significant difference in tendency to improve category of SRA between the experimental EI and the control group.

H_3 : There is a significantly higher tendency of improving category of SRA in the experimental group than the control group.

4.3 Discussion

The methodology utilized for this study was the mixed method. The design was quasi-experimental comparison-group design for the quantitative part, followed by semi-structured interviews for the qualitative part. Quantitative findings were based on statistical analysis of pre-instruction and post-instruction scores of Lawson Classroom Test of Scientific Reasoning, and qualitative findings were based on theme coding of semi-structured email interviews. In other studies, such as (Wuriyudani et al., 2018). The design of the research was One-Group Pre-test-Post-test. This design of one experimental group only, ignores the effect of factors other than the heuristic method of problem solving embedded in the discovery learning model. The heuristic method was embedded in a student practicum sheet in the topic of Pressure in physics. So may be the observed effect is due to the discovery learning model itself and not the heuristic method. So, the researcher preferred to use a control group in this study.

Detailed discussion is subsequently following and is presented into sections of *interpretations* of *key findings* and comparisons with findings of other studies, *Limitations* of the study and what parts could not be revealed on it, and finally *recommendations* and future practical actions that may be taken.

4.3.1 Discussion and interpretations of Quantitative part- question 1

A main purpose of this study was to quantitatively compare the impact of the new Explore-instruct teaching approach with the traditional Instruct-solve teaching approach. Comparison was done by comparing the gain in SR of the participants experiencing each teaching approach. Ultimately, the researcher aimed at answering the main question 1 which was: “Is there a difference in college students’ (SRA) development if taught by explore-instruct teaching approach compared to the instruct-solve approach (traditional approach) in physics?” first step that has been performed is ensuring reliability and consistency of LCTSR results. That was evident by a pilot study to the chosen population performed prior to the real study. The sample in the pilot study included nineteen students which were not included in the real study. Cronbach’s alpha of the total score of the test found to be 0.73,but as SR is multidimensions, consistency of each dimension of its six patterns was also investigated. All dimensions of LCTSR showed reliable results ($\alpha \geq 0.667$).

In Lawson (2003) study $the\ value\ of\ \alpha\ was$ 0.88, so the researcher started to collect the data for the real study trusting its reliability.

The sample included 100 students, 51 in the control group and 49 in the experimental. The independent variable was "the teaching approach" and the dependent variable was the "SRA" of the students.

4.3.1.1 Extraneous Confounding factors

To ensure Causal effect of the teaching approach at any observed change in SRA of the students, the researcher performed the following:

1. Comparing of the starting SRA between the two study groups, where MANCOVA test revealed that; all dimensions except *HDR* can be studied with trusted inferential conclusions in our sample.
2. Eliminating the impact of any confounding factor, resulting from the demographical differences which includes gender, age, English language level, nationality, and the starting SRA-Category. The demographical groups though, showed no significant difference when investigated for interactive impact on the pre-test%. But when MANCOVA test was performed for starting SRA-categories, it was clear that, the *participants' starting category* of SR was the confounding variable needs to be controlled when performing the inferential statistics. There may be a difference in any dimension of SR between the different age groups but may be the reason is that certain category of SRA was by accident accumulating more at certain age category, due to the non-random assignment of the students. So, the difference in that dimension would be related to the difference in SRA and not to the age factor. Another possible un-real impact may be due to accumulation of certain category at certain gender , nationality, English language level or study group by accident.

This section will start with discussing and interpreting the difference in SR dimensions among demographics including g*ender, Age, nationality, English language* and *starting SRA* of the participants; followed by a discussion of the descriptive statistics and inferential statistics.

4.3.1.1.1 Significant testing among demographics

Differences in the participants' populations found to be important when comparing normalized gains on concept inventories (Moore and Rubbo, 2012). It is then essential to study the participants' demographics when comparing normalized gain in SRA.

Starting with *gender*, it was found that gender has no impact on the SRA of the student. that was in line with other studies (i.e. Alzoubi et al, 2009; Piraksa, Srisawasdi & Koul 2014;Wuriyudani et al., 2018; Yaman&Karamustafaoglu,2006 &Luo et al., 2020), also in line with a study by Yuksel (2019) in Science History course and Nature of Science course, where Scientific Reasoning Skills Test (SRST) was developed by Yuksel (2015), and after using learning approach based on research-inquiry, and in terms of students' total scores of SR, there was no significant difference among gender groups (Yuksel, 2019). In contrast to a study by Yuksel and Tarakci (2018) on middle school students, post-test corrected scores for sub-dimensions of COV and HDR was significantly different between female and male students. Researchers need to consider other factors that may cause this difference if found between genders, such as the level of SRA of the participant before the test(Yuksel, 2019).

Regarding the participants' *age, and* in line with the results of this study; literature (Sokolowski, 2013) indicated that there are no age-related differences between college students and adolescent in the process or the development of SR. But the difference was there between 6th grade students and university students.

Studying the impact of the participants' *nationality*, even when controlling for other factors of *gender, age, English language level, starting category of SRA* and group; nationalities' sample sizes does not give trusted results on its impact on SR because of the small number of participants in some nationalities (less than 10 participants) which is one of the limitations of this study.

English Language level had no impact on the students' SRA as shown by the non-significant difference in SR skills between the different English language levels. That's in line to the study by (Kempert & Hardy, 2015) which analysed the data of 57 students in the elementary school and results revealed that there exist particular differences between bilingual and monolingual groups on the measurements of executive functions. On the measure of self-consciousness, the bilingual group performed superior to the monolingual, but no differences were found on the attentional control measure. Furthermore, regression analyses in that study of (Kempert & Hardy, 2015) revealed that the measures of attentional control and cognitive ability, explained students' performance on the reasoning task. And as there was no difference because of the language differences in the attentional control, accordingly no differences in the reasoning as was found here in our study.

Categories of SRA.

The starting categories for the control (C) group and the experimental (E) group were $\mathrm{C} = 49\%, \mathrm{E} = 80\%$ concrete level, $\mathrm{C} = 49\%, \mathrm{E} = 20\%$ Transitional level, and $\mathrm{C} = 2\%,\ \mathrm{E} = 0\%$ Formal level.The results of Khoirina and others (2018) indicated that for 51.14% of the sample size, the average score of the SRA for 11^{th} grade Indonesian population of students was at the category of concrete reasoning, at that concrete reasoning level, students were able to; answer the questions on conservation of mass and volume dimension, and proportional thinking very well. Of the studied sample 47.72% were at the transitional reasoning level, at which students were capable of answering all conservation of mass and volume questions, proportional thinking, and a few questions on dimensions of control of variables, correlational thinking indicators and Hypothetical-deductive reasoning. The lowest among the SRA categories in the study of Khoirina and colleagues was the formal reasoning category (1.14%), at that category of formal reasoning, students were able to properly answer all SR questions of LCTSR. Khoirina and colleagues attributed the causes of low ability of operational reasoning of high school students to the learning process which had not been classified properly to motivate and enhance students' SRA. the percentages of students in the different levels in Khoirina and others (2018) study was similar to the control group in this study. But The categories percentages were significantly different between the two study groups in this study, and it was indicated by the researcher, that the starting category is the confounding variable which needs to be controlled while studying the causal effect of the teaching approach on the growth or the gain in SR.

4.3.1.1.2 The students' Pre-test% and Post-test%.

The first hypothesis proposed to answer the first question was H_1 : "There is significantly higher scores of SR Post-test than Pre-test for both the Explore-Instruct and Instruct-Solve groups".

The students' average total SR % pre-test was in the concrete level (26%) and Post-test was in the transitional level (41.2 %). For the control group pre-test average score was in the concrete36.5 and ended up with transitional level SR % of 45.6.

Comparing the total SR scores of the two study groups before and after the intervention, Wilcoxon Signed Ranks Test, (Table 13)indicated that post-test total SR scores were significantly higher than pre-test scores in both study groups. With regard to the dimensions of SR, they were tested via Wilcoxon statistical test (Table 14),which indicated that *post-test%* of the students in both the experimental and control groups did show significant statistical difference from *pre-test%* for *CV*, *COV* and *CRR* dimensions. In the control group HDR had a significant difference between pre-test and post-test %, but not in the experimental group. While for the *PPR and CMR* dimensions the *post-test%* did not differ significantly from *pre-test %*. Those findings were in line with the findings of (Yulianti et al., 2020) which investigated the SRA of 32 students in a junior high school. The model was applied in three consecutive sessions, and the students' answers of 10 multiple-choice two-tier questions, which comprised of 7 SR dimensions, have been analysed. a few key findings were in line with the findings of this research as follows: (1) enhanced scores of students' tests were noticed for each SR dimension and (2) each syntax of the guided inquiry produced an uplift to the students' SRA. An indicator of the level of typical reasoning skills which was considered in(Yulianti et al., 2020) and also in study by (Lagubeau et al., 2020), was the percentages of correct answers for each dimension .

In interpreting that significant increment, it is good to recall the followed strategies in the two study groups. The experimental group experienced the Explore-Instruct (EI) teaching strategy while the control group was taught using the traditional way of Instruct-Solve (IS). Instruct-solve strategy often starts with concept explanation by the instructor, followed by examples solved also by the instructor, ending up by students solving problems independently (classroom observations were done by the researcher). *EI* strategy is the other way around, it starts with students exploring the problems independently, the instructor then provides guidance and encourage discussion between students, and the session ends up with concept presentation and explanation by the instructor. As the physics laboratory is a co-requisite to the physics course in the academic plan of the university where this study took place; laboratory experiments related to taught topics followed the related lectures in the traditional teaching strategy, but the experiments advance (within a week) the lectures in the EI strategy. All participants in the two study groups registered Physics in common. Otherwise, participating students study different other subjects, as the course or the section enrolment is individual and not similar for the hall section, it included students from different baches study other courses depending on their university specializations and levels.

Even though it is seen by students as an abstract complex subject, and it is logically sophisticated (Bao et. al., 2009), and due to the nature of the contemporary knowledge community which based upon science (Rusli, 2012); the basic physics lectures may lend themselves usefully for developing the scientific thinking of the students. And as an experimental science, Physics surrendered itself to the natural enhancement of the hypothetico-deductive reasoning; this reasoning is widely used in science practice and it involves utilizing a hypothesis to explain observed facts, designing an experiment and predict its outcomes using the proposed hypothesis, carrying the experiment out and then compare the outcomes with the predictions and accordingly accept or reject the hypothesis (Etkina et al., 2007). The Physics though, and any scientific subject may be taught via several approaches, but the Inquiry-based teaching approaches found to make a noticeable change to student's SRA in just one semester(L. Bao et al., 2009). That has been considered by Bao and his colleagues an encouraging outcome. However, to make it a solid finding, extra research is needed to understand the SRA change underlying mechanisms . In line with the study of Bao and others comes the results of Wilcoxon Signed Ranks test in **Table 14** . It has indicated that *post-test%* of the students in the experimental and control groups did show significant statistical difference from *pre-test%* for *CV*, *COV* and *CRR* dimensions. HDR in the control group had a significant difference between pre-test and post-test, but not in the experimental group. While for the *PPR and CMR* dimensions the *post-test%* did not differ significantly from *pre-test %* . So, both teaching strategies had no significant effect on the *PPR or CMR* of the participants, but they had a significant effect on the *CV, COV* and *CRR.* That urged us to focus the statistical tests that came after that on the difference between the EI and IS approaches in *CV,COV* and *CRR* only. In contrast to that, Yulianti and others (2020) observed the increment of all SR dimensions' scores after applying the guided inquiry model for 32 junior high school students, in three consecutive meetings on the topic of pressure, they concluded that each syntax of the guided inquiry model elevates the students' SRA.

The statistics in the result section showed that the COV experienced statistically and practically significant increment after the EI teaching design. When Koenig and other(2019) redesigned the traditional Physics Lab section and that resulted in development of aspects and moving the students to higher levels of reasoning of COV of algebra-based physics students (n = 296)and calculus-based physics students (n = 503), who used similar curriculums in their particular lab courses. It was indicated in (Koenig et al., 2019)that at the intermediate and low levels of COV, students made significant improvements. whereas at the highest COV level, little progress occurs.(Koenig et al., 2019) followed a random-group, repeated cycles, multi post-test methodology to assess more accurately by increasing the measurement frequency of both the occurring students' changes and the timing of those changes, to understand better the way by which certain pre-lab and in-lab activities affected the student progress of the COV skills. With more than 500 undergraduates in all sessions of the lab course every term, Koenig and her colleagues were able to form five groups randomly of around 100 undergraduate each. Therefore, statistically significant differences could be figured out with signals bigger than 5% with an error of measure on the order of2%. All 5 subgroups have been taking the SR pre-test at the term's first week, while the post-test has been randomly assigned to one group of the five at end of the course on weeks 4, 7, 9, 11, or 12 ; the first four weeks matched up with the due dates of the lab reports, indicating that an investigation cycle was completed, and the next investigation cycle will start. Because of the random assignment, the subgroups had been considered as equal samples of the same population. Accordingly, multiple measurements of the assigned weeks had been combined, Koenig and others were able to investigate the covariations between the teaching activities which occurred throughout the term period and the students' COV skills, with a resolution of time of about 2 weeks (one learning cycle). The results enabled Koenig and colleagues to discover that most of the effect on the COV dimension progress was a result of the first four weeks of the laboratory course, and mostly in the low and intermediate levels range. These results encouraged later in the course to make amendments to the activities. The amendments involved adding extra exercises prior to the lab and prompting questions for in-class activities in addition to lab reports that offer students with extra intentional practice of the lacking SR skills. Subsequent investigation showed that these amendments further developed students' abilities; the mean total pre–post shift increased from 10.8% in the year of 2016to 14.8% in the academic year 2017. Even though (Koenig et al., 2019)could demonstrate that specific SR skills may be taught inside existing structures of the course, students reached the maximum point early and did not progress to higher levels in the course. For intermediate and basic levels of COV, most of the progress happened during the first 4 weeks in the course. These COV skills are essential because they lead to the higher reasoning skill segments essential in hypothesis testing and validation, according to Kuhn's framework of theory-evidence coordination reasoning (Kuhn, et al. 2008). Koenig and other accordingly, suggests that extra course work is needed to develop students' COV reasoning further and target these complex skills better, particularly on the higher level COV.

About the importance of CoV aspect of reasoning, it has been revealed in a cross-sectional research with 1283 students in grades 5–13, Schwichow and others (2020) investigated whether there is a relation between students' mastery of Control of Variable Strategy (CVS) and their scientific content knowledge in the subject of physics. It has been indicated that CVS is significantly associated with science content knowledge of the students, even with controlling for the common impacts of the general reasoning skills. In lower-secondary school grades, Significant differences exist between students' CVS abilities and their science content knowledge when they receive physics education. Analysis reveals that the most difficult CVS aspect for students is understanding the effect of confounding. This sub-skill appears in the upper- secondary grade and it necessitates that students master CVS' procedural sub-skills more. Schwichow and others(2020) concluded that CoV Strategy and science content knowledge are strongly related in the contexts of secondary school science. The identified sub-skills different patterns among students as per Schwichow and others' suggestion can inform science educators and researchers about the CVS sub-skills that students lack to be trained and focused at during interventions. (Schwichow et al., 2020)

I. Inferential

The inferential statistics studied the impact occurred to all the SR dimensions because of the instruction design in both the control and experimental groups. The starting point in both groups needs to be similar. Because as per Johnson & Christensen (2014), purposive assignment or non-random assignment of students who were already enrolled in classes, threaten the validity of inferential conclusions. So, the Pre-test data of SR test had been compared between the control and experimental groups. MANCOVA test was performed, and it was revealed in Table 19 that; HDR was the only dimension that showed significant difference between the two groups prior to the treatment. But starting categories of SRA significantly differed in all SR dimensions, which is expected, as concrete reasoners will own by definition lower levels of SR dimensions than formal reasoners. The researcher is expecting that students who previously studied courses together or graduated from the same high school and planning to join the same major, enrolled themselves in the same sections leading to kind of purposive assignment of students in the control and experimental groups. Confounding factor then found to be the starting SRA of the students, that was in line with (Coletta & Steinert, 2020) which indicated that different populations of students can significantly have different reactions to the same pedagogy, so before interpreting any normalized gain, it is essential to consider measuring the starting abilities of the students. Because Coletta & Steinert (2020) wanted to avoid the confounding impact of the starting students' ability with the impact of the specific pedagogy. so, in interpreting normalized gains, average scores on either the SAT or LCTSR should be considered, since these scores are strongly associated with normalized gain(Coletta & Steinert, 2020).

After gauging the starting level of both study groups, and as people according to Lawson (2005) seem to process information in highly abstract sequences of hypothetico-deductive reasoning. It is then natural that, science teaching should offer students chances to produce and examine highly abstract and complicated theories and hypotheses in a hypothetico-deductive way. Students in this way, are expected to practice and become more and more aware of their underlying hypothetico-deductive thinking processes, progressively more competent in their HD application, and consequently more and more scientifically literate. So, depending on this conceptual framework, the researcher proposed the second hypothesis, $\boldsymbol{H_2}$: "There would be a significantly higher Normalized gain in all dimensions of SR in the EI group than in the IS group".The dependent variable here is the Normalized gain and the independent variable is the teaching strategy followed in each group.

Normalized gain for the total SR as a whole, and for each of the six SR dimensions individually was calculated using Hack 2002 rule. The difference in "Normalized Gain" in the total *SRA and* the difference in "Normalized Gain" of dimensions of PPR, COV, CMR and CRR was then investigated comparing the experimental and control groups to validate or not validate the hypothesis $\boldsymbol{H_2}$. The results showed that when expected extraneous factors are controlled, the two studied groups significantly differed in the N-Gain in COV dimension only. N-Gain in COV found to be significantly higher in the experimental group than the other group.

N-Gain in COV was the only dimension among the other SR dimensions that is significantly higher in the EI group than the control group (table 19). Nissen et al. argued recently for using *Cohen's d*, instead of the most widely used *N-gain*, in analysing the preintervention and postintervention scores on concept accounts utilised to assess the efficiency of instruction. They advocated such a change because they think N-gains are "prescore biased." Coletta& Steinert(2020) provided five examples, which show that if, data is carefully analysed , no pre score bias found, showing that their problem was the analysis' omitted variable bias. Coletta & Steinert showed that Cohen's d is not as informative as the N-gain when utilized as a singular parameter measuring of instruction effectiveness, although, Cohen's d is more commonly utilised in other fields. But Coletta & Steinert (2020)believed that researchers of physics education should continue using N-gain to measure efficacy of pedagogy(Coletta & Steinert, 2020).

These findings are in line with the study by (Erlina et al., 2018) where they analysed the Evidence-Based Reasoning (EBR) which they adopted as a framework for inquiry-based teaching to develop SR. Their research aimed to examine the EBR effectiveness in inquiry-based teaching of Physics in a trial to improve students' SR. Their study involved 139 secondary school students having similar prior knowledge. sample determination had been calculated by applying solving formula. The design of this study was one group pre-test post-test with repetition. Paired Sample T-test was used to study the effectiveness of teaching on developing SR and ANOVA test was used to examine the consistency of the effectiveness throughout in test group. The results indicated that EBR efficiently improved students' SR in Physics based upon two primary grounds. First, the category of the N-gain in SR component showed significant difference between pre and post-test, it proved that (COV)dimension reached high category, while proportional reasoning and probabilistic reasoning reached moderate category; correlational reasoning attained low category and hypothetical-deductive reasoning also attained low category. second, the level of SR has reflected that the experience was portrayed by somewhat imperfect answers. So Erlina and colleagues (2018) drew several conclusions, First, a highly significant improvement (2-tailed asymptotic significance of p<.0001)on SR of students resulted due to the implementation of EBR in Physics teaching. Also, COV dimension reached high category, after the EBR application, (proportional and probabilistic reasoning)reached moderate category and (correlational and hypothetical-deductive reasoning)reached low category. Lastly, no significant difference was evident in the amount of increment of each SR dimension between the 4 groups of the study. Erlina and colleagues (2018) also claimed that Students' motivation needs to be grown at beginning of the learning to attract their attention to the presentation of premises. They emphasized that EBR needs to put emphasis on the clarification process. Precise clarification on the explanation of the learned concepts and interpretation of the result of empirical verification needs to be undertaken before the application, leading eventually to an accurate claim. So, instruction needs to be coupled with logical interpretation (Erlina et al., 2018)

In line with Erlina and others (2018) study, it has been revealed in ***Table 20*** of this study that, "N-Gain of COV" due to Explore-Instruct design showed no significant difference($p > 0.05$) Between the different pre-instruction Categories, so regardless of the beginning SRA of the participant, Explore-instruct approach can achieve similar N-Gain of COV for all SR levels.

Correlation tests between *N-Gain in COV* and the study group were performed, and partial correlation test while controlling for the starting category revealed association between the N-Gain and the teaching strategy, with more N-Gain associated with the Explore-Instruct.

Finally, *regression analysis* test has been performed to study two relationships, one was between Post-test COV and Pre-test COV in each study group, the other one was between the teaching strategy and the N-Gain in COV.

reflected that *pre-test COV* was a significantly high predictor of *post-test COV* in the IS (control) group, it explains around 18.2% of the post-test COV in that group. In contrast, pre-test COV in the EI group; was not a significant predictor of the post-test CoV, it explains only 3.2% of the variance in post-test CoV. Beck and Bliwise, (2014) interpreted that when a significant interaction exist between pre-test score and experimental group on post-test score, that indicates differential impacts of a treatment based on participant's preparation. But the opposite is happening in this study, which means that the impact of EI approach on the *post-test CoV*, is not based on student's preparation or level of *pre-test CoV.*

Categories been affected:

The distribution of students among the different levels of the SR and their mean score in each category, after experiencing the two teaching styles have been analysed. And it has been observed that after the physics course, the Concrete participants in both groups experienced growth in SRA and move out of the concrete category more than the transitional and formal participants, but the transitional and formal categories did not experience such shifting and mostly stayed within the same category after instruction.

Crosstabulation of frequencies and percentages of students in each level of SR (concrete, transitional, or formal) before (Categ. -PRE) and after (Categ. -POST) the intervention was presented in the results section. The null hypothesis in this contingency table, H0: "There is no significant relationship between the starting percentage of SR category and the percentage of the post-instruction category". The crosstabulation indicated that; in the experimental group; 36 % of the concrete SR had moved out to the transitional category and 5% to the formal category. In contrast, in the control group, 36% also, moved out of the concrete category to the transitional and no one had moved out of this category to the formal. The formal reasoner participant stayed as a formal reasoner. Transitional reasoners mostly stayed in the transitional category in both groups, but in the control group, 12% of those transitional reasoners shifted to the formal category. and formal reasoners are in the ceiling already, so cannot be investigated for category raising. comparing that with surveyed data of a population of 231 high school students in Indonesia in the academic year 2017/2018 studied by Effendy and others (2018), Concrete Operations percentage was 33.11%, Transitional 30.41%, and Formal Operations 27.81%. The results of (Marušić & Sliško, 2012) of three tried teaching methods, presented in the below table and compared with the data of the Explore-Instruct (EI) and Instruct-Solve (IS) approaches in this study.

Table 25 : *Percentages of different approaches' students Pre- and Post-intervention in each category*

	Concrete	Transitional	Formal
EI group (%)			
Pre	80.0	20.0	0.0
Post	18.0	76.0	6.0
IS group (%)			
Pre	49.0	49.0	2.0
Post	21.0	61.0	18.0
RPQ group (%)			
Pre	26.4	57.1	16.5
Post	24.2	47.3	28.6
ED group (%)			
Pre	27.1	52.9	20.0
Post	15.3	45.9	38.8
TM group (%)			
Pre	31.3	49.4	19.4
Post	29.0	51.6	19.4

EI: Explore Instruct approach. **Pre**: Pre-intervention. **Post**: Post-intervention. **RPQ**: Read Present and Question. **TM**: traditional method. **ED**: Experimenting and Discussion.

In (Marušić & Sliško, 2012) the traditional method (TM) differs from the Read Present and Question (RPQ) and the Experimenting and Discussion (ED) in that the instructor wholly utilized oral presentation in presenting the lessons, the students wrote notes from the board in their notebooks and sometimes used the textbook. Numerical questions had been carried out on board. Unless there was some abstraction in the presented material there was no discussion during the session, and explanations were only given by the instructor. There was a strict organization of students' seating arrangement, and no communication was between them. In the contrast the ED approach was organized by the teacher in such a way that at every session a simple experiment was carried out. The experiment was described at the beginning of the session, to the students without really performing it. Students were then asked to expect the possible results of that experiment. Both the physical explanation and their predicted results had to be written down in their notebooks. They were then, asked to reveal and define their personal explanations of the predicted results. Once that was done, the students with the same views were grouped together, students were able then to present their explanations and debate. The debate made it possible for the instructor and students themselves to clearly recognize their level of SR. The experiment had been carried out post the debate by the instructor and the results were recorded. Unexpected experiments' results always produced students' positive emotions and enjoyment and they even often requested to re-carry out the experiment themselves as they felt the results was not believable for them. Obviously, the teacher then required the students to always conduct the experiment themselves. Following the performance of the experiment there was another debate supported by the reasons for expecting specific results. This discussion, has been helped and guided by the teacher, resulted in the formation of a better justification of the observed phenomenon. In this ED approach the seating was not strictly arranged, especially during conducting the experiment, the students grouped around the experiment to closely observe and participate in currying it out (Marušić & Sliško, 2012).In the RPQ teaching and learning design by Marušić and Sliško (2012), each section was broken down into three teams, to encourage analysis and discussion of the suggested topics from the field of the modern Physics. Each of the following tasks was assigned for one of the three teams in the section:

(1) presentation of the questions and problems that emerged from the first topic,

(2) presentation of the questions and problems that emerged from the second topic,

(3) questioning presenters, and critically analysis and evaluation of the reading materials.

The students in the RPQ approach selected their teams themselves, based on their interests and also based on the level of mastery in physics. The instructor assigned students to a proper team when the choice was doubtful, the instructor also assigned a team leader who was responsible for distributing reading materials and in charge of setting the team for their task in the project and for presenting the given topic.

Discussing the impact of the teaching approach on the concrete reasoners; the researcher observed that the biggest difference between pre- and post – the intervention was with the EI approach as revealed in table 25 . But of course, the population of this study is different from the one in Marušić and Sliško (2012) which have been presented in the same table for discussion purposes. Most of the transitional reasoners stayed in the transitional category in both groups, but 12% of those transitional reasoners shifted to the formal category in the control group. and formal reasoners are in the ceiling already, so cannot be investigated for category raising. These results of the contingency table of categories presented in the results section of this study, showed that there is a significant relation between the categories post-instruction and the beginning categories. with an effect size shown in the contingency table indicated practically significant results in both the experimental (Gamma = .840) and the control groups (Gamma = 0.961). the higher gamma value and the closer it is to value of 1, means the closer the statistical results to the real practical life results (Johnson & Christensen, 2014). Comparing those results with a study by Wuriyudani and his colleagues (2018) in the Indonesian context; they found that as their teaching approach of "problem solving heuristic method" showed a positive impact on the SR of students, the concrete level of SRA has decrease by 34.38%. the middle level of SRA (transitional category) has increased by 31,25% and the highest category of SRA (formal category) has increased by 3.13%.

So, the concrete participants had the most tendency among the other participants to shift toward a higher category when experiencing Physics teaching via Explore-Instruct or Instruct-Solve teaching approach in this study and even in other studies like the one of Wuriyudani and his colleagues (2018). The answer for sub-question 3 though (q_3): "Is there a significant difference in tendency to shift toward higher categories between the two followed teaching strategies?" is "No difference", because the crosstabulation did not reveal a clear difference between the two followed teaching style in the percentage of one category ending up with higher category. Comparing these findings of impact on categories with a study by Lagubeau, Tecpan & Hernández (2020); Lagubeau and the others compared the failure rate in introductory physics for the different categories and found that, transitional category, which was the major category in their context, benefited the most from their tried active learning innovation. So, the researcher is expecting that there may be a factor in the concrete reasoners which is highly affected by Explore-Instruct and Instruct-solve approaches in this study and in Wuriyudani and the others (2018) study, while in Lagubeau, Tecpan & Hernández (2020) the tried approach affected a factor in the transitional reasoners that decrease the failure rate in Physics for this category more than the other two categories.

Formal reasoners are in the ceiling already, so cannot be investigated for raising category, but the researcher is expecting a change in a smaller scale. The smaller scale change in the formal level of SR may be suggested as a topic to focus at in future studies. So, in this stage of our research, we can discard our third hypotheses that states H_3 : There is a significantly higher tendency of improving the category of SRA in the Explore-Instruct group than the Instruct-Solve group. And generally, as per Blumer and Beck (2019), with guided-inquiry laboratory courses, student SR skills in their study, did not progress pre-test to post-test. But, when divided into 4 quarters based on the pre-test score in each course, the lowest quartile students had significant gains in both groups. Nonetheless, these findings suggest that courses with guided-inquiry laboratory activities can foster the development of basic SR skills for students who are with minimum preparation throughout a range of class levels.

4.3.1.2 Explore-Instruct and CoV

Finally, *regression analysis* reflected a statistically and practically significant relation between *N-Gain in CoV and* the Explore-Instruct *teaching strategy. R*esults indicated significant impact of the Explore-Instruct teaching strategy where the achieved CoV post-test, is determined by the achieved normalized gain in CoV regardless of the starting score of CoV. But in the absence of Explore-Instruct teaching strategy, the achieved normalized gain which may be due to the nature of the Physics subject, is determined and predicted by the starting CoV pre-test. This interpretation encourages reasoners with low levels of CoV to expect enhancement in the pattern of CoV after experiencing Explore-instruct approach. And that prediction is via practically significant relation between *N-Gain in CoV and* the *EI teaching strategy,* wher*e* it appeared that the teaching strategy has been accounting for around 6.6% of the variance in N-Gain in CoV.

CoV dimension of SR if at all instructed in a college science course is likely to follow the traditional way of teaching science process as per Boudreaux et al., (2008), which is often taught as a set of standardized techniques to be memorized and pursued. Students are often unable to reason outside of the class context when faced with context that differ from the specific examples solved in instruction. Boudreaux and others found that even pupils with a relatively high mathematical skill in courses such as the calculus-based physics, find it difficult to link particular inferences with the related experimental results that are backing them. Science courses offer the students an opportunity to stimulate production of inferences based upon evidence and logical reasoning. The approach Boudreaux and colleagues have adopted in assisting students and guide them on how to apply COV is to monitor them through the needed reasoning in several contexts across an extended period. Their teaching experience and research indicated that students would develop capacity in certain skill when instruction in reasoning is embedded in a coherent frame of content and students would also strengthen their conceptual understanding of the related content. The procedure of going over the reasoning needed to produce inferences out of experimental data allows students to differentiate between the things they came to know because they were told and the things they understood. Boudreaux and others taught COV utilizing the context of sinking and floating to K–8 teachers and have concluded that the experience has enhanced their confidence and competence to do, learn, and instruct science. And as Boudreaux and colleagues' study focused mainly on the reasoning essential for making decisions of whether certain variable impacts the performance of a system or not. Even though they found that most students in their study could articulate the necessity to control variables while designing an experiment, Boudreaux and others found that many students struggled with the underlying COV reasoning. Boudreaux and colleagues then administered specifically designed activities to obtain comprehensive insights of student thinking, they found that a lot of students attempted to draw inferences about the impact of a variable even when numerous variables have been changed. On the other hand, some students drew inferences on whether a variable affects the system performance even if that variable have not been changed. Other students found to be utilizing the reasoning necessary for deciding that certain variable predicts the behaviour of a system when being requested to decide on the possibility of that variable impacting the behaviour of that system. Some students appeared to trust that only one variable can impact the behaviour of a system. Also there has been found a confusion between independent and dependent variables in addition to many difficulties resulted from certain beliefs about the causal processes related to a specific system. In general, Boudreaux and colleagues found that COV reasoning skill are challenging for all academic levels' students. Because designing experiments and results interpretation are key to the scientific process, anyone might expect that students of the calculus-based courses of physics have facility with that aspect of SR. But the activities used in Boudreaux and others' study were still challenging for that population, and that was because most of the students, do not appear to reflect on the skills of reasoning that underlies the COV on their own, their struggles need to be *explicitly addressed*, Moore and Rubbo, (2012) also agree with Boudreaux and others in this explicit presentation of SR in any of the elements of the course, for example, in the lab session of a course. As per Boudreaux and others, understanding the COV is important and necessary mainly for precollege instructors. Yet even instructors who have been previously teaching the topic of sinking and floating had significant difficulty with the activities used in Boudreaux and others' study. Those findings have strong consequences when developing an educated community, since data that are presented publicly are not easy to be interpreted to identify the effect of a specific variable. And it is crucial not only to be capable to interpret the data presented to draw inferences, but also to acknowledge the circumstances in which inferences are not warranted (Boudreaux et al., 2008)

4.3.1.3 The relationship between SR. and Explicit instruction

As per Moore and Rubbo, (2012) there is a strong association between pre-instruction SR scores and N-Gain on concept inventories. That association was strongest mostly for theoretical content, which lack direct observed examples, and mostly weak for descriptive content, where direct observed examples are abundant. significant gains in theoretical content (such as force and energy) were challenging for the population of students in Moore and Rubbo (2012) study, even though the implemented pedagogy was interactive research-verified engaging pedagogy which is expected to result in gain in content knowledge. Moore and Rubbo (2012) also observed that SR gains do not significantly happen without explicit instruction of SR patterns. This has many consequences for teaching, such as the essential need for explicit teaching of SR and the possible need for a re-evaluation of the established topics' sequence in astronomy and conceptual physics. But Explicit discussions about SR in a guided-inquiry classroom should be executed with care, as it is easy for example that struggling students to seemed to make connections, but they are simply copying by memorization the SR of other students in their group. This have been revealed to Moore and Rubbo, (2012) as they found students poorly performing on individual SR tasks after the instruction, while they were correctly approaching the tasks requiring those SR skills during the group discussion.

4.3.1.4 SR., Interpretation and Judgment.

The researcher expected development in SRA of both study groups because of the nature of the Physics as a scientific experimental subject, Physics raise SR as per (i.e., L. Bao et al., 2009; Effendy et al., 2018; Erlina et al., 2018), that is why the researcher selected the two groups design, to differentiate between SRA enhancement as a result of the taught subject and the enhancement as a result of the tried teaching approach Explore-Instruct.

As a result of the Explore-Instruct teaching approach in the experimental group; the researcher proposed $\boldsymbol{H_2}$:" There would be a significantly higher Normalized Gain in all dimensions of SR in the experimental group than in the control group". That hypothesis could be validated only for *CoV*. Explore-Instruct approach could enhance the dimension of Controlling and identification of variables and accordingly enhance the skill of judgment and decision making to control or vary a variable and evaluate what will happen to other variables affected by it. If that skill could be enhanced in scientific domain, it will accordingly be the basic for Control of Multivariable (Fyfe et al., 2014). Multivariable contexts are more than the one variable in the real daily life. In a study by Kuhn, all participants categorized as being successful in utilizing the strategy of control-of-variables, were able at the end of an intervention period, to identify multiple causal and non-causal effects functioning in the multivariable context explored by them in that study(Kuhn et al., 2008).That will lead to enhancement of the interpretation and judgment process in the social domain also (Kuhn et al., 2008).Because human interpretation in the social domain needs to be managed to enhance objectivity over subjectivity, while in the scientific domain needs to be exist and optimized to advance science according to Kuhn and colleagues who stated that:

> "In the social domain, then, the major challenge is to come to terms with the concern that human interpretation plays an unmanageable, overpowering role, while in the science domain, the major challenge is to recognize that human interpretation plays any role at all" (p.15)

4.3.1.5 Value of interpreted quantitative results

. In this study, interpreted results and findings show that traditional Physics teaching approaches if replaced with an inquiry-based approach will improve the aspect of identifying and controlling of variables which was found by recent studies (Boudreaux et al., 2008) to be challenging for students at all levels even students in scientific majors. and as per Drummond and Fischhoff (2017) little is known about the skills, which needed by the non-scientists to peruse and assess scientific evidence, so the findings of this study provide data in that area for curriculum developers, policy makers and teachers working in a context close to this study's context.

As the design of the study was pre-test post-test two groups design, with a group considered control group, the findings are more related to the tested Explore-Instruct teaching approach than the findings of other studies of - one group pre-test post-test design- such as the one by Erlina and others (2018).

Lagubeau and colleagues (2020) findings suggest a possible increment that may reach up to 150 students pass the course each year. But that study did not forget that the infrastructure of universities in the developing countries, is often consist of medium size classrooms of 50 students. Consequently, applying active learning will not increase sections' number, making it financially durable: the primary investment of a single classroom renovation, which is equivalent to about 8 student fees as per Lagubeau and colleagues (2020), would be reimbursed in less than a year through the drop of academic risk.

4.3.2 Discussion and interpretations of qualitative part- question 2

The interviews were via emails because the researcher and participants were in different cities and the only mean of contact with the participating students was the email. After receiving the students' responses and given that the sample size was only four students, the researcher did engage in a manual coding and therefore data of the interviews was manipulated by analysing the raw data of each response, to generate initial codes, related to the Physics pedagogies. These codes were then clustered together to form the themes in the following domains: Teaching approach and engagement domain, Problem-solving activities domain, Student's stress, and motivation domain.

4.3.2.1 Teaching approach and engagement theme:

For students' responses of: "Through scientific experiments I benefited a lot from the new concepts and in my opinion the explanation first then the attempt is good for establishment of the information". And "Attempts to solve questions then the lecture is good for applying what has been learned and in my viewpoint enough questions should be available related to each concept in the lectures".

Researcher's Interpretation and connection to literature: *For laboratory experiment***s** , instruct-solve is better, *but for theoretical lectures* and concepts; explore-instruct is preferred in a condition of having enough questions for each theoretical information to be delivered, and the reasons may be as expected by the researcher due to the differences between Lab. and lecture, in that there is no assessment of answers to questions provided before the instruction, but a submission of the experiment report is compulsory and will be graded. So, students feeling tensed about being assessed may urge them to choose instantaneous high grades instead of far deep learning.

For students' response of: *"I learned a lot* from the information in the physics course *and new definitions* on basics in magnetism and laws of Newton *and many new concepts***"** Researcher's Interpretation and connection to literature comes as follow: The student feel that she learned big amount of content knowledge, she described these contents as being new; even though it has been studied in high school. So, may be:

- The student perceived the topic in a different way than high school.
- The instructor did not connect the topics with what is already known by the student about those topics.
- More data was presented in the specific topic in college than higher school.
- The different language of instruction made the subject look new (most students had Physics in Arabic in the high school)

To connect these interpretations to theoretical framework and literature, it worth to recall Qadir (2016) theoretical aspects, that different cognitive constructions include various approaches inside the cognitive cycle, and the application of certain cognitive framework impact the related reasoning within. So as per the participants in this study the EI is preferred in lectures but in practical labs, where they are expected to conduct practical experiments, they prefer IS. Preference here reflects less mind conflicts in need of being resolved. so, laboratory sessions are expected by the researcher to expose the students to more mind conflicts. As per Koenig and others (2019), the more the recurrence of progressive learning cycles and accordingly more mind conflicts, the higher will be the development of SR, which is the case with the EI approach but rarely happens in IS approach, where the student is passively learning. Koenig and others (2019) showed that in the practical physics laboratory, progression of COV happens in the first four weeks, most probably, the students see that progression as receiving of new concepts and information even though they received this knowledge before in the high school. While experiencing the Explore-Instruct approach, the researcher expects the occurrence of more "phase sequences" formation, that starts with *procedural memory* cell-assemblies prior to *declarative memory* cell-assemblies (Phase sequence theory by O.Hebb in 1946), the thing that makes the explore-instruct approach cognitively closer to the process of scientific way of thinking. According to the abductive theory of scientific method (ATOM) by Haig, (2005) "phenomena serve the important function of prompting the Search for their own understanding by constructing relevant explanatory theories", that is consistent with the explore-instruct, where the observation in the laboratory or natural phenomena in daily life is noticed or experienced first (explore), followed by the student’s mind analysing it to reach a solution or an explanation, ending up by elaboration and discussion of theories during the (instruction phase) lecture.

Based on the previously proposed theoretical framework, which is supported by the cognitive learning theories (students’ understanding), and with the assistant of constructivist theories' teaching approaches (Barz & Achimaș-Cadariu 2016); That cognitive work explains the enhancement in COV dimension of scientific reasoning with EI approach more than the traditional Instruct-Solve approach. The researcher expects that; EI enables the students to experience more phase sequences that start with scientific thinking; SR mechanisms may then precede other cognitive processes when experiencing daily life situations on the long term.

4.3.2.2 Problem-solving activities ‘theme

Interpreting the students’ statement **“***Work sheets* were very helpful”: the researcher expects that solving questions independently was something that gives confidence for exam readiness and deepen the understanding of the received information. The thing that was showed and proven in previous studies such as the one by Ersoy and Başer (2014) which concluded that Problem-Based Learning (PBL) procedures has a commitment to the improvement of students’ creative thinking aptitudes. Not like the conventional teaching strategies, PBL teaching techniques build the expertise of the person's creative thinking which is one of the higher-order thinking abilities. Higher-order thinking abilities are so vital particularly these days, there is an urgent need for people having the capacity to think creatively. According also to a study by Schuster and others (2007), via teaching problem solving in Physics, the learners apply the physics’ principals while passing over and working out the solutions, and that enrich the learner’s thinking when tackling real-life issues. As an experimental science, Physics surrendered itself to the natural enhancement of the hypothetico-deductive reasoning; this reasoning is widely used in science practice and it involves utilizing a hypothesis to explain observed facts, designing an experiment and predict its outcomes using the proposed hypothesis, carrying the experiment out and then compare the outcomes with the predictions and accordingly accept or reject the hypothesis (Etkina et al., 2007).

4.3.2.3 Student's stress, and motivation theme

In an exit questionnaire was adopted from Örnek and others (2008) research and completed by a sample of 25 students from the population of this study, but these 25 students were not participants in the sample of this study. Causes of difficulty in the subject of Physics from the students' points of view were revealed. As per 54% of the students took the exit questionnaire, not having enough examples and problem solving in class, particularly conceptual examples and real-life applications was the strongest factor among the other factors that make physics difficult subject, followed by the factor of the lack of physics background (with 50% of the participants' selection) and the third factor in order is the factor of too much material to learn and not studying enough.

In discussing the results of the qualitative part of this study; the researcher reviewed an exploratory descriptive study by Wong and Kowitlawakul (2020). Individual face-to-face semi-structured interviews were carried out in the academic year 2018-2019. A sample of 20 nursing students from a university in Singapore was recruited. Those interviews were audio-recorded and then transcribed verbatim. Seven themes emerged from the thematic analysis which was used to analyse the data. The themes were: "1) essentials for nursing practices, 2) linking theory to practice, 3) individual thought process, 4) stimulating strategies, 5) classroom environment, 6) clinical environment, and 7) students' attributes" p. 104600**.** Nursing students considered clinical reasoning and critical thinking as necessary for nursing practices and defined these skills as the skills of connecting theory to practice. Students' critical thinking and clinical reasoning were found to be stimulated by Strategies like case-studies, simulation, guidance from clinical instructors and physical clinical experiences. Barriers as per Wong and Kowitlawakul (2020) to improving the critical thinking contained classroom settings, such as the teaching approached and student-instructor ratios, and students' attributes toward learning. Wong and Kowitlawakul stated that their findings provide points to be considered to improve the recent nursing practices and education to support the nursing students better in improving their clinical reasoning and critical thinking skills (Wong & Kowitlawakul, 2020)

Students' perceptions of pedagogies of the Physics laboratory have been investigated by (Koenig et al., 2019) during a semester in 2018 where a study took place and on an exit questionnaire, completed by a sample of 400 out of 742 students; two thirds of the participating students revealed that the lab course was enjoyable for them, while one quarter of the sample was neutral about enjoying it. Regarding the collaborative group work 85%liked that, and around two thirds found it valuable to get the chance to construct their own experiments, then utilize Excel in generating and interpreting graphs, and participate in SR to support a claim. On the other hand, the use of recommended skills of determining uncertainties in calculations and measurements was appreciated by only one third of the students. When asked about improvement, the main statement was more time in the laboratory. The lab time was 2 hours weekly, and considering the exploratory nature of the lab activities, students felt hurried to complete. Students also stated that they liked the labs in which they were requested to design their own experiments (challenge labs), such as the activity of designing the windmill blade system and they asked for more lab sessions like this. In a study by Erlina and others (2018), Students showed positive reaction to Evidence-Based Reasoning, that they admitted it facilitated for them being engaged in SR in Physics learning.

4.3.3 Integrating Quantitative and Qualitative results.

EI has been concluded by the researcher to enhance objectivity over subjectivity and SR over intuitive reasoning: because students with the low N-gain were in the transitional level of SR and they ended up in the higher transitional edge of SR, their responses showed that they are looking at the general benefits of the EI approach related to raising their reasoning on the long term, regardless of the difficulty they face at the beginning of the laboratory or lectures sessions and regardless of the unguaranteed grades (Objectivity in their opinions). Students with the medium N-gain were in the concrete level and improved to be in the transitional level. But the answers of these students focused more on the factor of completing the laboratory reports without missing data, and also having the lecture first so they can imagine how the problems are to be resolved and how the experiments to be conducted (Subjectivity in their opinion). This practice as viewed by the researcher, does not improve the imagery as one of the students responded. It actually provides the students with a ready maid imagery that they try to follow in their practices during the course. Quantitatively it was found by Stephens and Clement (2010) that there is a conjunction between scientific imagery and scientific creative reasoning, also that the enhancement of CoV dimension is associated with the enhancement of

Critical thinking. Critical thinking consequently improves the objectivity of the reasoner, and critical thinking also associated with the scientific imagery (Stephens and Clement, 2010). So, is there a lag period between the progression in CoV and critical thinking? Or critical thinking is affected as soon as the CoV experience a change? This may be addressed in future studies.

Themes were proposed based on the collected responses of the interviewed students: Teaching approach and engagement theme and that was clear quantitatively when the EI teaching approach resulted in higher COV than the traditional IS approach. Problem-solving activities' theme at which it was proved by students' perspectives and by previous studies that problem-solving procedures has a commitment to the improvement of students' creative thinking aptitudes, and that was clear with the enhanced COV which leads to progression in expert-like strategies in solving problems as been proven by Burkholder and others (2020). Student's anxiety and motivation theme, needs to be studied in the future for its correlation with the growth of COV dimension. Students admitted being tensed and not confident of submitting a good laboratory report if they did not receive the theoretical instruction related to the expected outcomes at the end of the experiments. The researcher is proposing a positive correlation between the motivating situations or contexts, such as having submission deadlines and the work to be graded, and the occurrence of optimum exploration by the students before the instruction take place by the instructor.

4.4 Research limitations

The limitations of this study are due to limitations of the instruments, the sampling technique used in each part (qualitative or quantitative), the period of the study and the sample size.

SR test by Lawson which has been edited in 2003 is widely being used and proved for its validity by majority of studies dealing with measuring the SR cognitive ability (Lawson, 1978, 1990, Lawson & Worsnop 1992), but still the student is being tested in the same questions post treatment, the thing that through a doubt of student's memory storing some answers and recall it when getting the same questions again. To reduce such possibility, the order of choices for each question had been changed from pre- to post-test and from one student to the other.

Sampling limitation is due to the purposive sampling technique of the students taking the test is reducing the generalizability of the test scores when compared to the random sampling technique which was not feasible here due to the need for controlling independent variables (prior knowledge (Prior Physics curriculum), course topics (current Physics curriculum), etc.) to ensure the reliability of the results to the dependent variable (the teaching style). Also, few students of the control group students in the two assigned groups which is close to 150 was not being taught by the same instructor, due to the system of the college that put a capacity limit of 30 students per lecture session and 15 students per Laboratory and tutorial session, thus setting limited load to each faculty member, the thing that will lead to losing control of the instructor factor. So, a class observation was done at the beginning of the semester to better accomplish consistency of instruction style between the two participating instructors.

The population size had an effect in some statistical analysis tests needed to be done in this study. The participant's nationality impact on the SRA is one example. Some nationalities had only 1, 2 or less than 10 participants, and for the ANOVA test more sample size is needed to give a real result. The number of participants in each category also is another example of insufficient sample size to perform comparison tests between categories.

Limitations related to the qualitative part of this study are demonstrated by Flick (2014) and include the following: the interpretation of the collected data is not having a clear procedure, even though the coding technique is found to fit best. Also, the generalizability of the concluded subjective theories in the qualitative part is not possible. The email interview method has its own limitations of missing the body language expressions, but in this study the body language does not reflect a lot compared to the written points of views.

The Period of tracking of the SR development in this study was 4 months during which 45 teaching hours of Physics took place with a rate of 3 hours/ week. Even though there was a noticeable gain in SR Within one semester which was the case also in other previous studies ((L. Bao et al., 2009), but following it for a longer period may reveal more detailed results as per (Novia & Riandi, 2017).

4.5 Implications and Research recommendations

4.5.1 Implications:

Results of this study, demonstrated in detail the impact of flipping students' timing of exploration and the instructor timing of instruction or guidance during the real or virtual classroom on student's SRA. These two timings, until the moment I am writing these lines come in the order of, instructor then the student. But before starting to instruct, and once starting a new topic, class teachers should realize the importance of what students bring with them in relation to the different stages of their intellectual development. Add to this the importance of apprehending declarative knowledge of scientific thinking. The importance of meaningful and active education lies in practicing various forms of science, where students face challenges represented in having buzzling problems or cases and then begin to participate in explaining their contexts and testing hypotheses - and here some of their ideas may conflict with the evidence and arguments presented by other students. All this is preferred to happen during student's timing, followed then by instructor guidance. Even advocates of flips close to the one in this study, who have been documented here by reviewing their studies such as (Fyfe et al., 2014, Daniel L. Schwartz & John D. Bransford, 1998) couldn't persuade most educators in the Middle East region particularly and world wild in general, to shift from the traditional teaching approach that starts with the teacher instruction. As per Dancy & Henderson (2010), Promoting change in pedagogical practices is difficult and poorly understood. It may be given according to Dancy & Henderson a careful focus based on research, similar to the focus been given while developing effective curriculum leading to effective pedagogies. One of the reasons in the case of pedagogies promoting SR in general or COV in particular may be the un-presented threshold of progression in CoV dimension with simply stimulating the students' minds and then starting the instruction, and accordingly threshold occurrence of progression in expert-like strategies in solving problems as proven by Burkholder and others (2020). That progression was proven to start happening in this study with a simple flip in the timings of instruction and student' application, without big change in the instructing phase or the application phase of the traditional way of teaching, and without changes in the learning resources. That resulted into progression in CoV, and for stronger progression, more inquiry-based classes, problem-based activities, and assessments may be added to the class and the curriculum.

The personal gains for the researcher in this study, were as follow:

First, she may confidently practice and disseminate Explore-instruct pedagogy knowing its positive impact on COV regardless of the starting level of the student's SR. It seems that researsher shares this concern and personal gains with (Zelkowski & Whitmire, 2018) "I (Author Zelkowski) saw a mindset shifting in (author Whitmire) in that for productive beliefs to continue evolving in light of student resistance, she would need to resist herself and find hard evidence of improving student outcomes. We share this journey." p. 1

The researcher can also determine the requirements or the design of a future research to empower the results, conduct explore-instruct in a different population and context and study the impact focusing on the other SR dimensions (other than COV). The researcher can also target and study the progression of student objectivity, acquiring of scientific knowledge and mastering of critical thinking, while teaching via explore-instruct teaching approach putting in mind its relation to progression in COV.

4.5.2 Future Research

If results of this study are to be scaled with future studies to analyse the impact of EI approach in CoV particularly and other dimensions of SR in general, for longer period of time, bigger sample size, and with much sample randomization, then the magnitude and limits of effects will be clearly explored, strongly evidenced and can be shared with education stakeholders, instructors, principles, educational coordinators, and decision-makers, to inform new directions in STEM teaching pedagogies that has major impact on human judgment, decision making and consequently critical thinking.

4.6 Conclusion

Countries all over the world are facing lots of challenges in all sectors, such as pandemics in health sectors, climate change in environmental sector, economic collapse in economic sector, etc. Those challenges are parallel in time to huge advances in technology, but that advance didn’t lead for solutions to the challenges. Only good human thinking that depends on scientific moral reasoning, can face these challenges. This study concluded that, all dimensions of SR can be uplifted by learning subjects that share features of the subject of Physics, that was evidenced when using either of the two approaches, the Explore-Instruct and the traditional Instruct-solve approaches, and it (SR progression) happened in all levels of SR, but the concrete operational reasoners found to benefit the most among the other level categories of reasoners, putting in mind the ceiling effect of LCTSR. It was also evidenced that, at all levels of SRA, the explore-instruct teaching approach significantly enhances the SR dimension of identification and control of variables more than the traditional IS approach. Identification and control of variables is an essential skill for raising human objectivity over subjectivity while judging and making decisions. The Control-of-Variables Strategy (CVS) is considered a trademark within the improvement of SR (Schwichow et al., 2020). It maintains that educational experiments have to be controlled and contrastive. Earlier prove recommends that CVS is associated with acquiring the scientific content knowledge (Schwichow et al., 2020).

Students’ perspectives of the tried EI approach, showed resistance from some students who were concrete reasoners before the interventions, but they were the most among the other levels of the reasoners to benefit from the EI approach and shift cognitively to the transitional level of SR. If these results are to be shared with education researchers and education stakeholders, instructors, principles, coordinators, and decision-makers it would inform new directions in STEM teaching pedagogies that has major impact on human judgment, decision making and consequently critical thinking.

REFERENCES

Abedi, J., & Ewers, N. (2013). *Accommodations for English learners and students with disabilities: A research based decision algorithm*. Retrieved from http://www.smarterbalanced. org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-representedstudents. pdf

Adair, J. E. (2007). *Decision Making & Problem-Solving Strategies*. NY: Kogan Page Publishers

Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association : JMLA*, vol. 103(3), pp.152–153.

Ahmad, W. M. A. W., Ahmad, B., Abdullah, S., & Mohamad, N. A. A. (2018). *Statistical Analysis using SPSS for Health Sciences* . Malaysia: Penerbit USM.

Al-Balushi, S. M., Al-Musawi, A. S., Ambusaidi, A. K., & Al-Hajri, F. H. (2017). The Effectiveness of Interacting with Scientific Animations in Chemistry Using Mobile Devices on Grade 12 Students' Spatial Ability and Scientific Reasoning Skills. *Journal of Science Education and Technology*, vol. 26(1), pp. 70–81.

Alfaro-LeFevre, R. (2017). Critical thinking, clinical reasoning, and clinical judgment. 6th edn. USA: Elsevier-Health Sciences.

Al Hawi, N. (2014). *Trust and Change: A case study of a University in the UAE Higher Education Sector*. 2014. PhD. The British University In Dubai.

Alzahmi, R. A., & Imroz, S. M. (2012). *A look at factors influencing the UAE* workforce education and development system. vol. *5*(8), 23.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* (pp. ix, 290). Oxford: Oxford University Press.

Arnoult, M. D. (1953). Transfer of pre differentiation training in simple and multiple shape discrimination. *Journal of Experimental Psychology*, vol. 45(6), pp. 401–409.

Badano, G., & Bonotti, M. (2020). Rescuing Public Reason Liberalism's Accessibility Requirement. *Law and Philosophy*, vol *39*(1), pp. 35–65.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang, Y., Li, L., & Wu, N. (2009). Physics: Learning and Scientific Reasoning. *Science*, vol. 323(5914), pp. 586–587.

Bao, L., Fang, K., Cai, T., Wang, J., Yang, L., Cui, L., Han, J., Ding, L. & Luo, Y. (2009). Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison. *American Journal of Physics*, vol. 77 (12), pp. 1118-1123.

Bao, Lei, Xiao, Y., Koenig, K., & Han, J. (2018). Validity evaluation of the Lawson classroom test of scientific reasoning. *Physical Review Physics Education Research*, vol. 14(2), 020106.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2015). Correction to: Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, vol. 44(5), pp. 1748–1748.

Barz, D. L., & Achimaș-Cadariu, A. (2016). The Development of Scientific Reasoning in Medical Education: A Psychological Perspective. *Clujul Medical*, vol. 89(1), 32.

Bawazir, O. A.-R. (2014). The holistic approach to clinical problem using Fishbone format: A simple way for clinical reasoning. *Education in Medicine Journal*, vol. 6(3).

Beatty, E. L., & Thompson, V. A. (2012). Effects of perspective and belief on analytic reasoning in a scientific reasoning task. *Thinking & Reasoning*, vol. 18(4), pp. 441–460.

Beck-Winchatz, B., & Parra, R. D. (2013). Finding Out What They Really Think: Assessing Non-Science Majors' Views of the Nature of Science. *College Teaching*, vol. 61(4), pp. 131–137.

Beichner, R., Bernold, L., Burniston, E., Dail, P., Felder, R., Gastineau, J., Gjertsen, M., & Risley, J. (1999). Case study of the physics component of an integrated curriculum. *American Journal of Physics*, vol. 67(1), pp.16–24.

Bellolio, C. (2019). The Quinean Assumption. The Case for Science as Public Reason. *Social Epistemology*, vol. *33*(3), pp. 205–217.

Berrett, D. (2012).How flipping the classroom can Improve the traditional lecture. *Chronicle of Higher Education*.

Bhat, M. (2019). Learning Styles in the Context of Reasoning and Problem-Solving Ability: An Approach based on Multivariate Analysis of Variance. *International Journal of Psychology and Educational Studies*, vol. 6(1), pp. 10–20.

Bomide, G. S. (1991). Comments on "the uses of Lawson's test of formal reasoning in the Israeli science education context." *Journal of Research in Science Teaching*, vol. *28*(5), pp. 463–464.

Bonotti, M. (2019, July 11). *Response to "The Quinean Assumption. The Case for Science as Public Reason."* Social Epistemology Review and Reply Collective.

Bottge, B. A., Rueda, E., Kwon, J. M., Grant, T., & LaRoque, P. (2009). Assessing and tracking students' problem-solving performances in anchored learning environments. *Educational Technology Research and Development*, vol. 57(4), pp. 529–552.

Boudreaux, A., Shaffer, P., Heron, P., & C. McDermott, L. (2008). Student understanding of control of variables: Deciding whether or not a variable influences the behaviour of a system. *American Journal of Physics - AMER J PHYS*, vol. 76, pp. 163–170.

Bradshaw, W. S., Nelson, J., Adams, B. J., & Bell, J. D. (2017). Promoting the Multidimensional Character of Scientific Reasoning. *Journal Of Microbiology & Biology Education*, vol. *18*(1).

Brandon, E., & Sibbratthie, N. (1996). Logical Reasoning as a Curriculum Area in School. *Education in the West Indies: Developments and Perspectives*, 11.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, vol. *3*(2), pp. 77–101.

Brogt, E. (2009). *Pedagogical and curricular thinking of professional astronomers teaching the Hertzsprung-Russell diagram in introductory astronomy courses for non-science majors*. PhD. The University Of Arizona.

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A Framework for Analysing Scientific Reasoning in Assessments. *Educational Assessment*, vol. 15(3–4), pp. 142–174.

Burkholder, E., Blackmon, L. & Wieman, C. (2020). Characterizing the mathematical problem-solving strategies of transitioning novice physics students. Physical Review Physics Education Research, vol. 16 (2).

Busher, H. & James, N. (2012). in S. Delmont. In cyberspace: qualitative methods for educational research. Cheltenham: Edward Elgar Publishing.

Carmel, J. H., Jessa, Y., & Yezierski, E. J. (2015). Targeting the Development of Content Knowledge and Scientific Reasoning: Reforming College-Level Chemistry for Nonscience Majors. *Journal of Chemical Education*, Vol *92*(1), pp. 46–51.

Chen, N.-S., Kinshuk, D., Wei, C.-W., & Liu, C.-C. (2011). Effects of matching teaching strategy to thinking style on learner's quality of reflection in an online learning environment. *Computers & Education*, vol. *56*,pp. 53–64.

Chen, Z., & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development*, vol. *70*(5), pp. 1098–1120.

Choi, E., Lindquist, R., & Song, Y. (2014). Effects of problem-based learning vs. Traditional lecture on Korean nursing students' critical thinking, problem-solving, and self-directed learning. *Nurse Education Today*, vol. *34*(1), pp. 52–56.

Chua, J. S. M., & Lateef, F. A. (2014). The Flipped Classroom: Viewpoints in Asian Universities. *Education in Medicine Journal*, vol. *6*(4).

Clark, R., & Taylor, D. (1992). Training Problem Solving Skills Using Cognitive Strategies: Part 1: Novice Versus Expert Problem Solvers. *New Directions for Philanthropic Fundraising*, vol. 31, pp. 1–5.

Coletta, V. (2014). *Thinking in Physics* (1st edition). New York: Pearson.

Coletta, Vincent P. (2017). Reaching More Students Through Thinking in Physics. *The Physics Teacher*, vol. *55*(2), pp. 100–105.

Coletta, Vincent P., Phillips, J. A., Singh, C., Sabella, M., & Rebello, S. (2010). *Developing Thinking & Problem-Solving Skills in Introductory Mechanics*. pp. 13–16.

Coletta, V.P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analysing pre instruction and post instruction scores on concept inventories. *Physical Review Physics Education Research*, vol. 16(1).

Cone, C., Godwin, D., Salazar, K., Bond, R., Thompson, M., & Myers, O. (2016). Incorporation of an Explicit Critical-Thinking Curriculum to Improve Pharmacy Students' Critical-Thinking Skills. *American Journal of Pharmaceutical Education*, vol. 80(3), 41.

Conlin, L. D., Gupta, A., Scherr, R. E., Hammer, D., Hsu, L., Henderson, C., & McCullough, L. (2007). *The Dynamics of Students' Behaviours and Reasoning during Collaborative Physics Tutorial Sessions*. AIP Conference Proceeding vol. 951 (1). pp.69–72.

Cramer, K., & Post, T. R. (1993). Connecting Research To Teaching Proportional Reasoning. *Mathematics Teacher*, Vol. *86*, pp. 404–407.

Creativity in Conceptual Change: A Cognitive-Historical Approach. (2008). In N. J. Nersessian, *Creating Scientific Concepts*. Cambridge: The MIT Press.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. California: SAGE.

Croker, S., Morris, B., Zimmerman, C. & Masnick, A. (2021). "The Emergence of Scientific Reasoning", in H. Kloos, B. Morris and J. Amaral (ed.). *Current Topics in Children's Learning and Cognition*. InTech.

Dai, W.-Z., Xu, Q., Yu, Y., & Zhou, Z.-H. (2019). *Bridging machine learning and logical reasoning by abductive learning*. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver: Canada

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics - AMER J PHYS*, vol. 78(1).

Daniel L. Schwartz, & John D. Bransford. (1998). A Time for Telling. *Cognition and Instruction*, vol. 16(4), pp. 75–522.

Daniel L. Schwartz, & Taylor Martin. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction*, vol. *22*(2), pp. 129–184.

DeHaan, R. L. (2011). Teaching Creative Science Thinking. *Science*, vol. 334(6062), pp. 1499–1500.

Demirdamar, R., & Toklu, H. (2013). The evaluation of prescription dispensing scores of the pharmacy students before and after the problem-based "rational drug use" course: Results of the two years' experience. *MARMARA PHARMACEUTCAL JOURNAL*, vol. *3*(17), pp. 175–175.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interviews. *Medical Education*, vol. 40(4), pp. 314–321.

Ding, L., Wei, X., & Mollohan, K. (2016). Does Higher Education Improve Student Scientific Reasoning Skills? *International Journal of Science and Mathematics Education*, vol. 14(4), pp. 619–634.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher*, vol. 35(3), pp. 125–129.

Dolan, E., & Grady, J. (2010). Recognizing Students' Scientific Reasoning: A Tool for Categorizing Complexity of Reasoning During Teaching by Inquiry. *Journal of Science Teacher Education*, vol. 21(1), pp. 31–55.

Dorner, D. (1997). *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations* (1st edition). New York: Basic Books.

Drachsler, H., & Kirschner, P. (2011). *Learner Characteristics* (p. 2). https://doi.org/10.1007/978-1-4419-1428-6_347

Drummond, C., & Fischhoff, B. (2017). Development and Validation of the Scientific Reasoning Scale. *Journal of Behavioural Decision Making*, vol. 30(1), pp. 26–38.

Druyan, S. & Levin, I. (1996). The Differential Contribution of Field-dependent and Field-independent Cognitive Styles to Socio-cognitive Transaction in Promoting Scientific Reasoning. International Journal of Behavioural Development, vol. 19 (4), pp. 831-850.

Dunbar, K. N., Sabella, M., Henderson, C., & Singh, C. (2009). *The Biology Of Physics: What The Brain Reveals About Our Understanding Of The Physical World*. AIP Conference Proceedings. vol. 1179 (15). pp. 15–18.

Duncker, K., & Lees, L. S. (1945). *On problem-solving*. Washington, DC: American Psychological Association.

Effendy, S., Hartono, Y., & Ian, M. (2018). *The Ability of Scientific Reasoning and Mastery of Physics Concept of State Senior High School Students in Palembang City*. Proceedings of the International Conference on Science and Education and Technology 2018 (ISET 2018). Paris: Atlantis Press. pp.504-509 pp. 404-509

Elder, L., & Paul, R. (2003). *A Miniature Guide for Students and Faculty to Scientific Thinking*. Online: Foundation for Critical Thinking.

Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, vol. *22*(5–6), pp. 333–349.

Erlina, N., Susantini, E., Wasis, W., Wicaksono, I. & Pandiangan, P. (2018). The effectiveness of evidence-base reasoning in inquiry-based physics teaching to increase student's scientific reasoning. Journal of Baltic Science Education, vol. 17 (6), pp. 972-985.

Ersoy, E., & Başer, N. (2014). The Effects of Problem-based Learning Method in Higher Education on Creative Thinking. *Procedia - Social and Behavioural Sciences*, vol. 116(1), pp. 3494–3498.

Etkina, E., Gentile, M., Karelina, A., Ruibal-Villasenor, M. R., Suran, G., Sabella, M., Henderson, C., & Singh, C. (2009). *Searching for "Preparation for Future Learning" in Physics*. pp. 141–144.

Etkina, E., Karelina, A., & Villasenor, M. R. (2007). *Studying Transfer Of Scientific Reasoning Abilities*. vol. *883*, pp. 81–84.

ETS. (2010). Linking TOEFL iBT ™ Scores to IELTS® Scores. ETS. [Accessed 15 May 2021], from https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Evans, J., & Frankish, K. (2009). *In Two Minds: Dual Processes and Beyond* (J. Evans & K. Frankish, Eds.). Oxford: Oxford University Press.

Evans, J. St. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, vol. *6*.

Fabby, C., & Koenig, K. (2015). *Examining the Relationship of Scientific Reasoning with Physics Problem Solving*. Vol. 16(4), 8.

Flick, U. (2014). *An Introduction to Qualitative Research* (Fifth edition). New York: SAGE Publications Ltd.

Fuller, T. (2013). Is scientific theory change similar to early cognitive development? Gopnik on science and childhood. *Philosophical Psychology*, vol. 26(1), pp. 109–128.

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology*, vol. *84*(3), pp. 502–519.

Gavrilovska, L., Atanasovski, V., Macaluso, I. & DaSilva, L. (2013). Learning and Reasoning in Cognitive Radio Networks. IEEE Communications Surveys & Tutorials, vol. 15 (4), pp. 1761-1777.

Hussain, Liaquat & Nawaz, Allah & Abbas, Asif & Khan, Allah & Khan, Muhammad. (2014). A Gender-based comparative study of teaching Physics through CAI and Ordinary Lecture Method. Gomal University Journal of Research. vol. 30. pp. 44-55. _

Gibson, E. (1969). *Principles of Perceptual Learning and Development* (First printing edition). New York: Pearson College Div.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, vol. 62(1), pp. 32–41.

Giere .R. N. (1984). Causal Models with Frequency Dependence . *Journal of Philosophy*. vol. 81 (7), pp. 384-391

Giere, R., Bickle, J. & Mauldin, R. (2006). Understanding scientific reasoning. Belmont, Calif.: Thomson/Wadsworth.

Gillani, B. (2010). Inquiry-Based Training Model and the Design of E-Learning Environments. *Issues in Informing Science and Information Technology*, vol. 7(1), pp. 001–009.

Gillani, B. B. (2003). *Learning Theories and the Design of E-Learning Environments*. Maryland: University Press of America.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of Grounded Theory: Strategies for Qualitative Research*. England: Routledge.

Golanbari, M., & Garlikov, R. (2008). *Employing socratic pedagogy to improve engineering students' critical reasoning skills: Teaching by asking instead of by telling*. ASEE Annual Conference and Exposition, Conference Proceedings.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, vol. 10(4), pp. 371–388.

Haig, B. D. (2008). Scientific method, abduction, and clinical reasoning. *Journal of Clinical Psychology*, vol. 64(9), pp. 1013–1018.

Hake, R. (2002). Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pre-test Scores on Mathematics and Spatial Visualization. The Physics Education Research Conference; [online]. Boise, Idaho. [Accessed 7 September 2020]. Available at: http://www.arxiv.org

Han, J. (2013). Scientific Reasoning: Research, Development , and Assessment. PhD. The Ohio State University.

Handhika, J., Cari, C., Soeparmi, A., & Sunarno, W. (2016). *Student conception and perception of Newton's law*. AIP Conference Proceedings vol. 1708, 070005.

Hazen, R. M., Trefil, J., & Washington, C. I. of. (1991). *Science Matters—Achieving Scientific Literacy* (Second Printing edition). New York: Doubleday.

Heit, E., & Feeney, A. (2007). *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*. Cambridge: Cambridge University Press.

Hitt, G. W., Isakovic, A. F., Fawwaz, O., Bawa'Aneh, M. S., El-Kork, N., Makkiyil, S., & Qattan, I. A. (2014). Secondary implementation of interactive engagement teaching techniques: Choices and challenges in a Gulf Arab context. *Physical Review Special Topics - Physics Education Research*, *10*(2). Scopus. https://doi.org/10.1103/PhysRevSTPER.10.020123

Holyoak, K. J., & Morrison, R. G. (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge: Cambridge University Press.

Howard, A. (Ed.). (1995). *The Jossey-Bass social and behavioural science series. The changing nature of work.* Jossey-Bass.

Hoyningen-Huene, Paul. (2008). Systematicity: The Nature of Science. Philosophia. vol. 36. pp. 167-180.

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, vol. 15(9), pp. 1277–1288.

Hsu, C.-Y., Kalyuga, S., & Sweller, J. (20150511). When should guidance be presented in physics instruction? *Archives of Scientific Psychology*, vol. *3*(1), 37.

Hudicourt-Barnes, J. (2003). The Use of Argumentation in Haitian Creole Science Classrooms. *Harvard Educational Review*, vol. 73 (1), pp. 73-93.

Huizingh, E. (2007). *Applied Statistics with SPSS*. New York: SAGE Publications.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods*, vol. 18(1),pp. 3–20.

James, N. (2016). Using email interviews in qualitative educational research: Creating space to think and time to talk. *International Journal of Qualitative Studies in Education*, vol. 29(2), pp. 150–163.

Jao, L. (2018). *Practice-Based Pedagogy in Mathematics and Science Teaching Methods: Challenges and Adaptations in Context*. vol. 10.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the Test…or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding. *Educational Psychology Review*, vol. 26(2), pp. 307–329.

Jensen, Jamie L., Neeley, S., Hatch, J. B., & Piorczynski, T. (2017). Learning Scientific Reasoning Skills May Be Key to Retention in Science, Technology, Engineering, and Mathematics. *Journal of College Student Retention: Research, Theory & Practice*, vol. 19(2), pp. 126–144.

Jensen, Jamie Lee, & Lawson, A. (2011). Effects of Collaborative Group Composition and Inquiry Instruction on Reasoning Gains and Achievement in Undergraduate Biology. *CBE—Life Sciences Education*, vol. 10(1), pp. 64–73.

Johnson, R. and Christensen, L. (2014). *Educational Research*. 5th ed. Los Anglos: Sage publications.

Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, vol. 14(4), pp. 308–312.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, vol. 40(4), pp. 651–672.

Kapur, M. (2018). Examining the preparatory effects of problem generation and solution generation on learning from instruction. *Instructional Science*, vol. 46(1), pp. 61–76.

Kartal, T., & Kartal, B. (2019). Examining Strategies Used by Pre-service Science Teachers in Stoichiometry Problems in Terms of Proportional Reasoning. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*,vol. 48(1), pp. 910–944.

Kempert, S. & Hardy, I. (2015). Children's scientific reasoning in the context of bilingualism. International Journal of Bilingualism, vol. 19 (6), pp. 646-664.

Khan, B. (2014). Undergraduate students' perception about current lecturing practices, Pak. Armed Forces Med. J. vol. 64(2). pp. 319-327

Kim, K. H., VanTassel-Baska, J., Bracken, B. A., Feng, A., & Stambaugh, T. (2014). Assessing Science Reasoning and Conceptual Understanding in the Primary Grades Using Standardized and Performance-Based Assessments. *Journal of Advanced Academics*, vol. 25(1), pp. 47–66.

Kim, Y. (2006). *Students' Cognitive Conflict and Conceptual Change in a Physics by Inquiry Class*. vol.818(1). pp. 117–120.

Kind, P., & Osborne, J. (2017). Styles of Scientific Reasoning: A Cultural Rationale for Science Education? *Science Education*, vol. 101(1), pp. 8–31.

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*, vol. 41(2), pp. 75–86.

Knol, M. H., Dolan, C. V., Mellenbergh, G. J., & van der Maas, H. L. J. (2016). Measuring the Quality of University Lectures: Development and Validation of the Instructional Skills Questionnaire (ISQ). *PLOS ONE*, vol. 11(2), e0149163.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, vol. *36*(5), pp. 757–798.

Koenig, K., Wood, K., Bortner, Larry, & Bao, L. (2019). Modifying Traditional Labs to Target Scientific Reasoning. *Journal of College Science Teaching*, vol. 48, pp. 28–35.

Koenig, K., E Wood, K., Bortner, L. & Bao, L. (2019). Modifying Traditional Labs to Target Scientific Reasoning. Journal of College Science Teaching, vol. 48 (5), pp. 28-35.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, vol. 41(4), pp. 212–218.

Kryjevskaia, M., & Stetzer, M. R. (2013). *Examining inconsistencies in student reasoning approaches*. AIP Conference Proceedings, vol. 1513.pp. 226–229.

Kuhn, D. (2011). What is scientific thinking and how does it develop? In *The Wiley-Blackwell handbook of childhood cognitive development, 2nd ed* (pp. 497–523). Hoboken: Wiley-Blackwell.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, vol. 23(1), pp. 435–451.

Kuhn, D., & Jr, D. D. (2004). Connecting Scientific Reasoning and Causal Inference. *Journal of Cognition and Development*, vol. *5*(2), pp. 261–288.

Kuhn, D. & Modrek, A. (2017). Do reasoning limitations undermine discourse? Thinking & Reasoning, vol. 24 (1), pp. 97-116.

Kurian, R. E. R., & Al-Assaf, Y. (2020). Impact of high school curriculum on student performance at university. *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, pp. 1–7. https://doi.org/10.1109/GHTC46280.2020.9342924

Lack, C. W., & Rousseau, J. (2016). *Critical Thinking, Science, and Pseudoscience: Why We Can't Trust Our Brains*. New York: Springer Publishing Company.

Lagubeau, G., Tecpan, S., & Hernández, C. (2020). Active learning reduces academic risk of students with nonformal reasoning skills: Evidence from an introductory physics massive course in a Chilean public university. *Physical Review Physics Education Research*, vol. 16(2).

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, vol. 15(1), pp. 11–24.

Lawson, A.E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, vol. 42(6), pp. 716–740.

Lawson, A. E. (1995). Science Teaching and the Development of Thinking. Belmont, CA: Watsworth Publishing Company.

Lawson, A. E. (2002). What Does Galileo's Discovery of Jupiter's Moons Tell Us About the Process of Scientific Discovery? *Science & Education*, vol. 11(1), pp. 1–24.

Lawson, A. E. (2003). *The Neurological Basis of Learning, Development and Discovery: Implications for Science and Mathematics Instruction*. Hidelberg: Springer Netherlands.

Lawson, Anton E., & Weser, J. (1990). The rejection of non-scientific beliefs about life: Effects of instruction and reasoning skills. *Journal of Research in Science Teaching*, vol. 27(6), pp. 589–606.

Lawson, Anton E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. *Journal of Research in Science Teaching*, vol. 29(2), pp. 143–166.

Lawson, A. E. (2004). The Nature and Development of Scientific Reasoning: A Synthetic View. *International Journal of Science and Mathematics Education*, vol. 2(3), 307.

Lazonder, A. W., & Wiskerke-Drost, S. (2015). Advancing Scientific Reasoning in Upper Elementary Classrooms: Direct Instruction Versus Task Structuring. *Journal of Science Education and Technology*, vol. 24(1), pp. 69–77.

Lee, O., Quinn, H. & Valdés, G. (2013). Science and Language for English Language Learners in Relation to Next Generation Science Standards and with Implications for Common Core State Standards for English Language Arts and Mathematics. *Educational Researcher*, vol. 42 (4), pp. 223-233.

Bao, L., Fang, K., Cai, T., Wang, J., Yang, L., Cui, L., Han, J., Ding, L. & Luo, Y. (2009). Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison. American Journal of Physics, vol. 77 (12), pp. 1118-1123.

Leijser, J., & Spek, B. (2021). Level of clinical reasoning in intermediate nursing students explained by education year and days of internships per healthcare branches: A cross - sectional study. *Nurse Education Today*, vol. *96.*

Leonard, D. C. (2002). *Learning Theories: A to Z: A to Z*. California: ABC-CLIO.

Lepine, J. A., Colquitt, J. A., & Erez, A. (2000). ADAPTABILITY TO CHANGING TASK CONTEXTS: EFFECTS OF GENERAL COGNITIVE ABILITY, CONSCIENTIOUSNESS, AND OPENNESS TO EXPERIENCE. *Personnel Psychology*, vol. *53*(3), pp. 563–593.

Liao, Y.-W., & She, H.-C. (2009). Enhancing Eight Grade Students' Scientific Conceptual Change and Scientific Reasoning through a Web-based Learning Program. *Educational Technology & Society*, vol. 12(4), pp. 228-240.

Litan, R. E., Wyckoff, A. W., Fealing, K. H., Panel on Developing Science, T., and Innovation Indicators for the Future, Committee on National Statistics, Division of Behavioural and Social Sciences and Education, Board on Science, T., and Economic Policy, Policy and Global Affairs, & National Research Council. (2014). *Capturing Change in Science, Technology, and Innovation: Improving Indicators to Inform Policy*. Washington, D.C.: National Academies Press.

Lloyd, G. E. R. (2013). Reasoning and Culture in a Historical Perspective. *Journal of Cognition and Culture*, vol. 13(5),pp. 437–457.

Locke, J. (1996). *An Essay Concerning Human Understanding* (K. P. Winkler, Ed.; Abridged edition). Indiana: Hackett Publishing Company, Inc.

Loehr, A. M., Fyfe, E. R., & Rittle-Johnson, B. (2014). Wait for it . . . Delaying Instruction Improves Mathematics Problem Solving: A Classroom Study. *The Journal of Problem Solving*, vol. *7*(1).

Lorch, R., Lorch, E., Calderhead, W., Dunlap, E., Hodell, E., & Freer, B. (2010). Learning the Control of Variables Strategy in Higher and Lower Achieving Classrooms: Contributions of Explicit Instruction and Experimentation. *Journal of Educational Psychology*, vol. 102, pp. 90–101.

Luo, M., Wang, Z., Sun, D., Wan, Z. & Zhu, L. (2020). Evaluating scientific reasoning ability: The design and validation of an assessment with a focus on reasoning and the use of evidence. Journal of Baltic Science Education, vol. 19 (2), pp. 261-275.

Marrison, C. (2002). *The Fundamentals of Risk Measurement* (1st edition). New York: McGraw-Hill Education.

Marušić, M., & Sliško, J. (2012). Influence of Three Different Methods of Teaching Physics on the Gain in Students' Development of Reasoning. *International Journal of Science Education*, vol. 34(2), pp. 301–326.

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, vol. 29(1), pp. 43–55.

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist*, vol. 59(1), pp. 14–19.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, vol. 63(2), pp. 81–97.

Mintzes, J. (2006). Handbook of college science teaching. Arlington, Va: NSTA Press.

Moll, R. F., & Milner-Bolotin, M. (2009). The effect of interactive lecture experiments on student academic achievement and attitudes towards physics. *Canadian Journal of Physics*, vol. 87(8), pp. 917–924.

Moore, D. S., Notz, W. I., & Fligner, M. A. (2015). *The Basic Practice of Statistics* (Seventh edition). New York: W. H. Freeman.

Moore, J. C., & Rubbo, L. J. (2012). Scientific reasoning abilities of nonscience majors in physics-based courses. *Physical Review Special Topics - Physics Education Research*, vol. 8(1).

Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The Emergence of Scientific Reasoning. In *Current Topics in Children's Learning and Cognition*. London: Intech Open.

Muijs, D. (2010). *Doing Quantitative Research in Education with SPSS* (2nd edition). New York: SAGE Publications Ltd.

Murphy, P. K., Firetto, C. M., & Greene, J. A. (2017). Enriching Students' Scientific Thinking through Relational Reasoning: Seeking Evidence in Texts, Tasks, and Talk. *Educational Psychology Review*, vol. 29(1), pp. 105–117.

National Research Centre. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. Washington, D.C.: National Academies Press.

Ndihokubwayo, K., Ndayambaje, I., & Uwamahoro, J. (2021). Analysis of Lesson Plans from Rwandan Physics Teachers. *International Journal of Learning, Teaching and Educational Research*, vol. 19(12)Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memoryoriented training promotes memory for training. *Memory & Cognition*, vol. 19(6), pp. 543–557.

Nersessian, N. (2008). Creativity in Conceptual Change: A Cognitive-Historical Approach. In N. J. Nersessian, *Creating Scientific Concepts*. Cambridge. The MIT Press.

Netz, R. (1999). *The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History*. Cambridge: Cambridge University Press.

Nosich, G. (2011). *Learning to Think Things Through: A Guide to Critical Thinking Across the Curriculum* (4th edition). London: Pearson.

Novia, N. & Riandi, R. (2017). The Analysis of Students Scientific Reasoning Ability in Solving the Modified Lawson Classroom Test of Scientific Reasoning (MLCTSR) Problems by Applying the Levels of Inquiry. Jurnal Pendidikan IPA Indonesia, vol. 6 (1).

N. Torres, H., & L. Zeidler, D. (2002). The Effects of English Language Proficiency and Scientific Reasoning Skills on the Acquisition of Science Content Knowledge by Hispanic English Language Learners and Native English Language Speaking Students. *The Electronic Journal for Research in Science & Mathematics Education*. https://ejrsme.icrsme.com/article/view/7683

Obizoba, C. (2015). *Instructional Design Models—Framework for Innovative Teaching and Learning Methodologies*, vol. 2(1), 12.

Örnek, F., Robinson, W., & Haugan, M. (2008). What makes physics difficult. *International Journal of Environmental & Science Education*, vol. 18 (3), pp. 30–34.

Paas, F., Renkl, A., & Sweller, J. (2010). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, vol. 38(1), pp. 1–4.

Palmer, P. J. (1998). *The Courage To Teach: Exploring the Inner Landscape of a Teacher's Life.* San Francisco, CA : Jossey-Bass Inc., Publishers,

Patton, M. Q. (2002). Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective. *Qualitative Social Work*, vol. 1(3), pp. 261–283.

Paul, R. (1993). *Critical Thinking: How to Prepare Students for a Rapidly Changing World* (1st edition). US: Foundation for Critical Thinking.

Paul, R., & Elder, L. (2006). *Critical Thinking: Learn the Tools the Best Thinkers Use*. New Jersey: Pearson/Prentice Hall.

Paul, R., Niewoehner, R., & Elder, L. (2006). *The Thinker's Guide to Engineering Reasoning*. US: Foundation for Critical Thinking.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, vol. 2(3), pp. 176–186.

Piaget, J. (1973). *To Understand is to Invent: The Future of Education* (First Edition). Grossman Publishers.

Piaget, J. (1976). Need and Significance of Cross-Cultural Studies in Genetic Psychology. In B. Inhelder, H. H. Chipman, & C. Zwingmann (Eds.), *Piaget and His School: A Reader in Developmental Psychology* (pp. 259–268). Springer. https://doi.org/10.1007/978-3-642-46323-5_19

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*(2), 153–179. https://doi.org/10.1111/j.2044-835X.2012.02082.x

Piraksa. (2014). Effect of Gender on Student's Scientific Reasoning Ability: A Case Study in Thailand. *Procedia - Social and Behavioural Sciences*, *116*, 486–491. https://doi.org/10.1016/j.sbspro.2014.01.245

Potter, W., Webb, D., Paul, C., West, E., Bowen, M., Weiss, B., Coleman, L., & De Leone, C. (2014). Sixteen years of collaborative learning through active sense-making in physics (CLASP) at UC Davis. *American Journal of Physics*, vol. 82(2), pp. 153–163.

Prosser, M., & Trigwell, K. (2014). Qualitative variation in approaches to university teaching and learning in large first-year classes. *Higher Education*, vol. 67(6), pp. 783–795.

Public and scientists' views on science and society. (2015). Washington, D.C.: Pew Research centre. [Accessed 15 May 2021], http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Putri, H., Siahaan, P., & Hernani. (2020). Students' scientific reasoning and argumentative abilities through levels of inquiry models based on socio-scientific issue. *Journal of Physics: Conference Series*, vol. 1521, 042100.

Pyper, B. A. (2012). Changing scientific reasoning and conceptual understanding in college students. *AIP Conference Proceedings*, vol. 1413(1), pp. 63–65.

Qadir, J. (2016). Artificial intelligence based cognitive routing for cognitive radio networks. *Artificial Intelligence Review*, vol. 45(1), pp. 25–96.

Quinn, H., Lee, O. and Valdés, G., 2012. Language demands and opportunities in relation to Next Generation Science Standards for English language learners: What teachers need to know. *Commissioned papers on language and literacy issues in the Common Core State Standards and Next Generation Science Standards*, vol. *94*(2012), pp.32-32.

Ramful, A., & Ho, S. Y. (2015). Quantitative reasoning in problem solving. *Australian Primary Mathematics Classroom*, vol. 20(1), pp. 15–19.

Rawls, J. (2005). *Political Liberalism: Expanded Edition* . New York: Columbia University Press.

Redekop, B. W. (2009). Common sense in philosophical and scientific perspective. *Management Decision*, vol. 47(3), pp. 399–412.

Redish, E. (1994). Implications of cognitive studies for teaching physics. *American Journal of Physics - AMER J PHYS*, vol. *62*, pp. 796–803.

Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 40(5), pp. 1432–1440.

Ridenour, J., Feldman, G., Teodorescu, R., Medsker, L., & Benmouna, N. (2013). *Is conceptual understanding compromised by a problem-solving emphasis in an introductory physics course*? Physics Education Research Conference. AIP Conference Proceedings, vol. 1513(1), pp. 338–341.

Ridge, N., Kippels, S., Farah, S., & Research, S. (2017). Curriculum Development in the United Arab Emirates. *Sheikh Saud Bin Saqr Al Qasimi Foundation Policy Paper Series*, vol. 18.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses: structural topic models for survey responses. *American Journal of Political Science*, vol. 58(4), pp. 1064–1082.

Romine, W. L., Sadler, T. D., & Kinslow, A. T. (2017). Assessment of scientific literacy: Development and validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR). *Journal of Research in Science Teaching*, vol. 54(2), pp. 274–295.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of Undergraduate Science Course Innovations on Learning. *Science*, vol. 331(6022), pp. 1269–1270.

Rusli, A. (2012). *Science and scientific literacy vs science and scientific awareness through basic physics lectures: A study of wish and reality*. pp. 169–173.

Ryder, J. (2001). Identifying Science Understanding for Functional Scientific Literacy. *Studies in Science Education*, vol. 36(1), pp. 1–44.

Saiz, C. & Rivas, S.F. (2016). New teaching techniques to improve critical thinking. The Diaprove methodology. Educational Research Quarterly, vol. 40 (1), pp. 3-36

Schiefer, J., Golle, J., Tibus, M., & Oschatz, K. (2019). Scientific Reasoning in Elementary School Children: Assessment of the “Inquiry Cycle.” *Journal of Advanced Academics*, vol. 30(2), pp. 144–177.

Schuster, D., Undreiu, A., Adams, B., Hsu, L., Henderson, C., & McCullough, L. (2007). *Multiple Modes of Reasoning in Physics Problem Solving, with Implications for Instruction*. pp. 184–187.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, vol. 103(4), pp. 759–775.

Schwichow, M., Osterhaus, C., & Edelsbrunner, P. A. (2020). The relation between the control-of-variables strategy and content knowledge in physics in secondary school. *Contemporary Educational Psychology*, vol. *63*.

Scientific Reasoning Test. (2000). Retrieved August 25, 2020, from http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Mathematics%20Assessments/Scientific%20Reasoning%20Test.pdf

Sella, A. C., Mendonça Ribeiro, D., & White, G. W. (2014). Effects of an Online Stimulus Equivalence Teaching Procedure on Research Design Open-Ended Questions Performance of International Undergraduate Students. *The Psychological Record*, vol. 64(1), pp. 89–103.

She, H.-C., & Liao, Y.-W. (2010). Bridging scientific reasoning and conceptual change through adaptive web-based learning. *Journal of Research in Science Teaching*, vol. 47(1), pp. 91–119.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, vol. 26(2), pp. 145–159.

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting Validity in the Assessment of English Learners. *Review of Research in Education*, *39*(1), 215–252. https://doi.org/10.3102/0091732X14557003

Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. *Frontiers in Psychology*, vol. 6.

Sokolowski, A. (2013). *Inductive Reasoning and the Theory of Equation*. vol. 4.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, vol. 46(6), pp. 610–631.

Stacey, K., & Vincent, J. (2009). Modes of reasoning in explanations in Australian eighth-grade mathematics textbooks. *Educational Studies in Mathematics*, vol. *72*(3), pp. 271–288.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning* (pp. xvi, 296). Mahwah: Lawrence Erlbaum Associates Publishers.

Stephens, A. L., & Clement, J. J. (2010). Documenting the use of expert scientific reasoning processes by high school physics students. *Physical Review Special Topics - Physics Education Research*, vol. 6(2), 020122.

Sternberg, R., Todhunter, R., Litvak, A. & Sternberg, K. (2020). The Relation of Scientific Creativity and Evaluation of Scientific Impact to Scientific Reasoning and General Intelligence. Journal of Intelligence, vol. 8 (2), p. 17.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, vol. 12(2),pp. 257–285.

Taconis, R., Ferguson-Hessler, M. G. M., & Broekkamp, H. (2001). Teaching science problem solving: An overview of experimental work. *Journal of Research in Science Teaching*, vol. *38*(4), pp. 442–468.

Tajudin, N. Mohd., Saad, N. S., Rahman, N. A., Yahaya, A., Alimon, H., Dollah, Mohd. U., & Abd Karim, Mohd. M. (2012). Mapping the level of scientific reasoning skills to instructional methodologies among Malaysian science-mathematics-engineering undergraduates. *AIP Conference Proceedings*, vol. 1450, pp. 262–265.

Tala, A.-Z., al-shara, I., & Mohammad, A.-S. (2009). The Scientific Reasoning Level of Students' in the Faculty of Science in Al-Hussein Bin Talal University and Its Affection of Gender, Teaching level, and Specialization. *An-Najah University Journal for Research - Humanities*, vol. 33, pp. 401–437.

Tanişli, D., & Dur, M. (2018). Quantitative Reasoning: Reflections on Solving Real-World Problems. *Nicel Muhakeme: Gerçek Yaşam Problemlerinin Çözüm Sürecinden Yansımalar.*, vol. 47(1), pp. 60–108.

Tateo, L. (2014). Science at the Supermarket: Multiplication, Personalization and Consumption of Science in Everyday Life. *Integrative Psychological and Behavioural Science*, vol. 48(2), pp. 161–175.

Teherán Sermeño, P. D. C., Carriazo, J. G., & León Luque, J. C. (2010). Blended Learning Applied to an Introductory Course on Conceptual Physics. *International Journal of Online Engineering (IJOE)*, vol. *6*(3).

Ticu, M. (2015). *Critical thinking in development of creativity*. vol. 7.

Trickett, S. B., & Trafton, J. G. (2007). "What if…": The Use of Conceptual Simulations in Scientific Reasoning. *Cognitive Science: A Multidisciplinary Journal*, vol. 31(5), pp. 843–875.

Turkan, S., & Liu, O. L. (2012). Differential performance by english language learners on an inquiry-based science assessment. *International Journal of Science Education*, *34*(15), 2343–2369. Scopus. https://doi.org/10.1080/09500693.2012.705046

Vallier, K. (2011). Against Public Reason Liberalism's Accessibility Requirement. *Journal of Moral Philosophy*, vol. *8*(3), pp. 366–389.

Vallier, K. (2014). *Liberal Politics and Public Faith: Beyond Separation*. England: Routledge.

Vallier, K. (2016). In Defence of Intelligible Reasons in Public Justification. *The Philosophical Quarterly*, vol. *66*(264), pp. 596–616.

van der Graaf, J., van de Sande, E., Gijsel, M. & Segers, E. (2019). A combined approach to strengthen children's scientific thinking: direct instruction on scientific reasoning and training of teacher's verbal support. International Journal of Science Education, vol. 41 (9), pp. 1119-1138.

Wati, M., Mahtari, S., Hartini, S., & Amelia, H. (2019). A Rasch Model Analysis on Junior High School Students' Scientific Reasoning Ability. *International Journal of Interactive Mobile Technologies (IJIM), vol.* 13 (07), pp. 141–149.

Weitz, D., O'Shea, G., Zook, N., & Needham, W. (2011). Working Memory and Sequence Learning in the Hebb Digits Task: Awareness Is Predicted by Individual Differences in Operation Span. *The American Journal of Psychology*, vol. 124(1), pp. 49–62.

Weld, J., Stier, M., & McNew-Birren, J. (n.d.). The Development of a Novel Measure of Scientific Reasoning Growth Among College Freshmen: The Constructive Inquiry Science Reasoning Skills Test. *Research and Teaching*, vol. 8.

Wilkins, S. (2010). Higher education in the United Arab Emirates: An analysis of the outcomes of significant increases in supply and competition. *Journal of Higher Education Policy and Management*, vol. *32*(4), pp. 389–400.

Wollin, J. A., & Fairweather, C. T. (2012). Nursing education: A case study of a Bachelor of Science Nursing programme in Abu Dhabi, United Arab Emirates: Nursing education in Abu Dhabi. *Journal of Nursing Management, vol. 20*(1), pp. 20–27.

Wolpert, L. (1994). *The Unnatural Nature of Science*. Cambridge: Harvard University Press.

Wong, S. & Kowitlawakul, Y. (2020). Exploring perceptions and barriers in developing critical thinking and clinical reasoning of nursing students: A qualitative study. *Nurse Education Today,* vol. 95, p. 104600.

Wuriyudani, H., Wiyanto, W. & Darsono, T. (2018). Problem Solving Heuristic to Develop Scientific Reasoning. *Physics Communication*, vol. 3 (1), pp. 1-9.

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*, vol. *84*(5), pp. 204–212.

Yin, R. K. (2003). *Case Study Research: Design and Methods*. California: SAGE.

Yoder, G., & Cook, J. (2014). *Rapid conversion of traditional introductory physics sequences to an activity-based format*. vol. 15(2), 9.

Yuksel, I. (2019). The Effects of Research-Inquiry Based Learning on the Scientific Reasoning Skills of Prospective Science Teachers. Journal of Education and Training Studies, vol. 7 (4), p. 273.

Yulianti, E., Mustikasari, V. R., Hamimi, E., Rahman, N. F. A., & Nurjanah, L. F. (2020). *Experimental evidence of enhancing scientific reasoning through guided inquiry model approach*. AIP Conference Proceedings 2215.

Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of Research in Science Teaching*, vol. 47(9), pp. 1064–1093.

Zelkowski, J., and Whitmire, B. (2018). *Impact of Explore-Instruct Pedagogy on Student Learning: Solidifying teacher belief changes to productive!* Alabama Journal of Mathematics. vol. 4(2018), pp. 1–7.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, vol. *27*(2), pp. 172–223.

APPENDICES

APPENDIX I. : LETTERS OF ETHICAL CLEARANCE

الجامعة البريطانية في دبي

The British University in Dubai

www.buid.ac.ae

20 February 2016

Institute of Applied Technology
Abu Dhabi, United Arab Emirates

This is to certify that Ms Fatima Ahmed Abazar with ID number 2014121016 is a registered part-time student on the Doctor of Education programme in The British University in Dubai since September 2014.

Ms Abazar is currently working on her research titled "The Impact of Shifting from Instruct-solve to Explore-instruct Teaching Approach on the Students' Scientific Reasoning". She is required to gather data through surveys, and interviews. Your permission to conduct her research in your organisation is hereby requested. Further support provided to her in this regard will be highly appreciated.

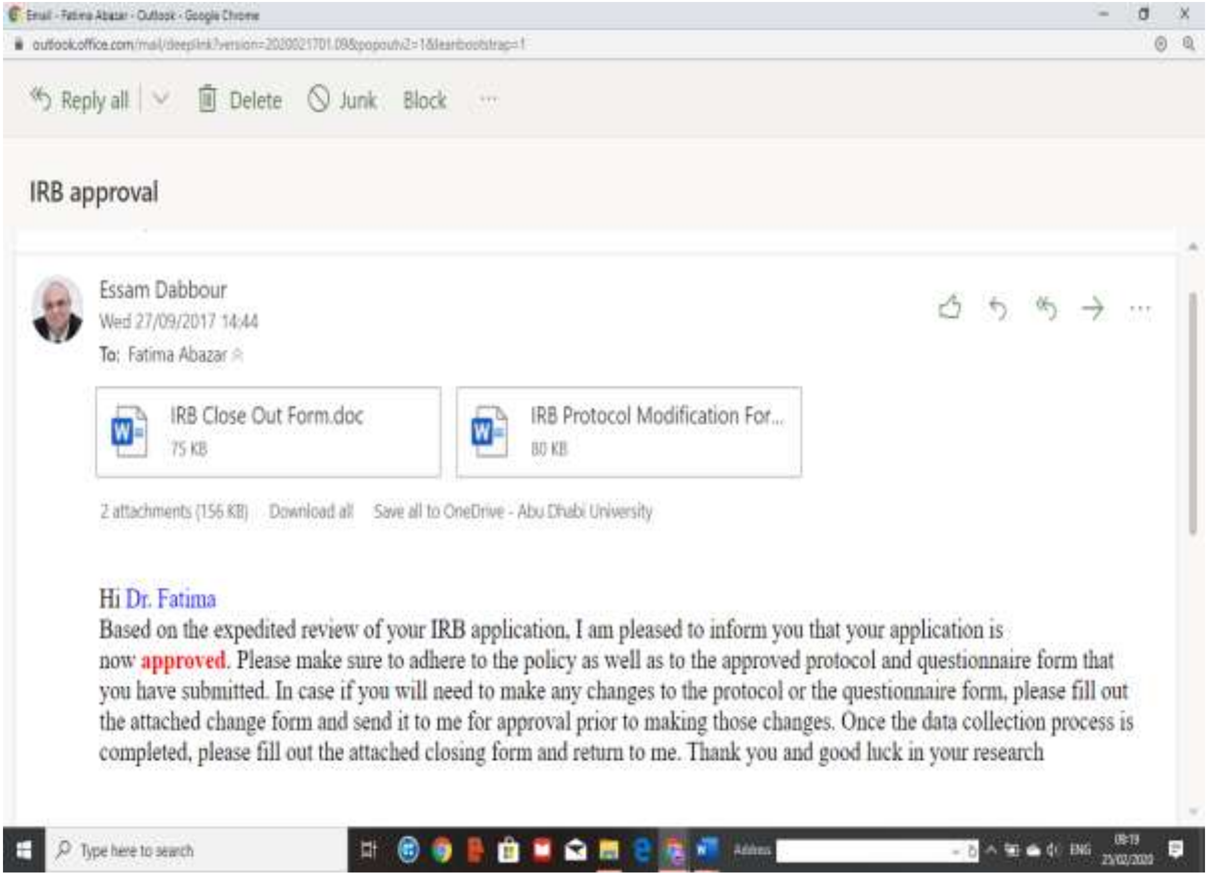
This letter is issued on Ms Abazar's request.

Yours sincerely,

Amer Alaya
Head of Student Administration

The British University in Dubai Student Administration

PO Box 345015 • Block 11 Dubai International Academic City Dubai U A E • T +971 4 279 1400 • F +971 4 279 1490
FB.com/BUiD.Team BUiD_Team youtube.com/BUiDadmin @BUiD_Team BUiD



Email - Fatima Abazar - Outlook - Google Chrome

Reply all | Delete | Junk | Block

IRB approval

Essam Dabbour
Wed 27/09/2017 14:44
To: Fatima Abazar

IRB Close Out Form.doc
75 KB

IRB Protocol Modification For...
80 KB

2 attachments (156 KB) Download all Save all to OneDrive - Abu Dhabi University

Hi Dr. Fatima
Based on the expedited review of your IRB application, I am pleased to inform you that your application is now **approved**. Please make sure to adhere to the policy as well as to the approved protocol and questionnaire form that you have submitted. In case if you will need to make any changes to the protocol or the questionnaire form, please fill out the attached change form and send it to me for approval prior to making those changes. Once the data collection process is completed, please fill out the attached closing form and return to me. Thank you and good luck in your research

Scientific Reasoning Test 2000

I know the purpose of this study is to check the impact of the utilized teaching approach on the Scientific Reasoning of the students and I have decided to volunteer as a research participant for this study.

	CRN: 2072	SEC 1-LAB	Signature	Date
			[illegible]	10-1-2017
			[illegible]	10-Jan-2017
			Aya	10-Jan-2017
			[illegible]	
			[illegible]	10-1-2017
			[illegible]	10-Jan-2017
			Fatima	10-1-2017
			[illegible]	10/1/2017
			[illegible]	10.1.17
			[illegible]	10-Jan-2017
			[illegible]	10-1-2017
			Jawaher	10-Jan-2017
			Khawla	10-1-2017
			[illegible]	10, 1, 2017
			[illegible]	10-1-2017
			[illegible]	10-Jan-2017
			[illegible]	
			[illegible]	10-1-2017
			[illegible]	10-Jan-2017
			[illegible]	January, 10, 20
			[illegible]	10-1-2014
			[illegible]	10-1-2017
			[illegible]	10-1-2017
			Afraa	9-1-2017
			[illegible]	9-1-2017
			[illegible]	9-1-2017
			[illegible]	9-1-2017
			[illegible]	9-1-2017
			[illegible]	9,1,2017
			[illegible]	9-1-2017
			[illegible]	9-1-2017
			[illegible]	9-1-2017

APPENDIX II. : LCTSR

CLASSROOM TEST OF

SCIENTIFIC REASONING

Multiple Choice Version

Directions to Students:

This is a test of your ability to apply aspects of scientific and mathematical reasoning to analyze a situation to make a prediction or solve a problem. Make a dark mark on the answer sheet for the best answer for each item. If you do not fully understand what is being asked in an item, please ask the test administrator for clarification.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

Revised Edition: August 2000 by Anton E. Lawson, Arizona State University. Based on: Lawson, A.E. 1978. Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1): 11-24.

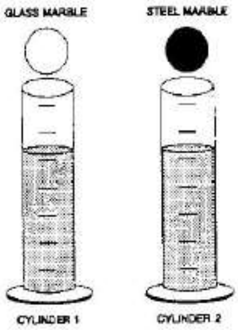
1. Suppose you are given two clay balls of equal size and shape. The two clay balls also weigh the same. One ball is flattened into a pancake-shaped piece. *Which of these statements is correct?*

   a. The pancake-shaped piece weighs more than the ball
   b. The two pieces still weigh the same
   c. The ball weighs more than the pancake-shaped piece

2. *because*

   a. the flattened piece covers a larger area.
   b. the ball pushes down more on one spot.
   c. when something is flattened it loses weight.
   d. clay has not been added or taken away.
   e. when something is flattened it gains weight.

3. To the right are drawings of two cylinders filled to the same level with water. The cylinders are identical in size and shape.

   Also shown at the right are two marbles, one glass and one steel. The marbles are the same size but the steel one is much heavier than the glass one.

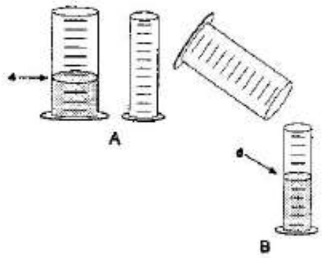
   When the glass marble is put into Cylinder 1 it sinks to the bottom and the water level rises to the 6th mark. *If we put the steel marble into Cylinder 2, the water will rise*

   a. to the same level as it did in Cylinder 1
   b. to a higher level than it did in Cylinder 1
   c. to a lower level than it did in Cylinder 1

4. *because*

   a. the steel marble will sink faster.
   b. the marbles are made of different materials.
   c. the steel marble is heavier than the glass marble.
   d. the glass marble creates less pressure.
   e. the marbles are the same size.

5. To the right are drawings of a wide and a narrow cylinder. The cylinders have equally spaced marks on them. Water is poured into the wide cylinder up to the 4th mark (see A). This water rises to the 6th mark when poured into the narrow cylinder (see B).



Both cylinders are emptied (not shown) and water is poured into the wide cylinder up to the 6th mark. *How high would this water rise if it were poured into the empty narrow cylinder?*

a. to about 8
b. to about 9
c. to about 10
d. to about 12
e. none of these answers is correct

6. *because*

a. the answer can not be determined with the information given.
b. it went up 2 more before, so it will go up 2 more again.
c. it goes up 3 in the narrow for every 2 in the wide.
d. the second cylinder is narrower.
e. one must actually pour the water and observe to find out.

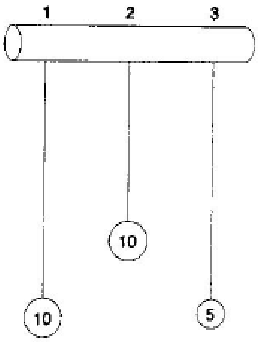
7. Water is now poured into the narrow cylinder (described in Item 5 above) up to the 11th mark. *How high would this water rise if it were poured into the empty wide cylinder?*

a. to about 7 1/2
b. to about 9
c. to about 8
d. to about 7 1/3
e. none of these answers is correct

8. *because*

a. the ratios must stay the same.
b. one must actually pour the water and observe to find out.
c. the answer can not be determined with the information given.
d. it was 2 less before so it will be 2 less again.
e. you subtract 2 from the wide for every 3 from the narrow.

9. At the right are drawings of three strings hanging from a bar. The three strings have metal weights attached to their ends. String 1 and String 3 are the same length. String 2 is shorter. A 10 unit weight is attached to the end of String 1. A 10 unit weight is also attached to the end of String 2. A 5 unit weight is attached to the end of String 3. The strings (and attached weights) can be swung back and forth and the time it takes to make a swing can be timed.



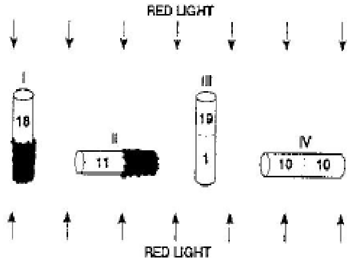
Suppose you want to find out whether the length of the string has an effect on the time it takes to swing back and forth. *Which strings would you use to find out?*

a. only one string
b. all three strings
c. 2 and 3
d. 1 and 3
e. 1 and 2

10. *because*

a. you must use the longest strings.
b. you must compare strings with both light and heavy weights.
c. only the lengths differ.
d. to make all possible comparisons.
e. the weights differ.

11. Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to red light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing.



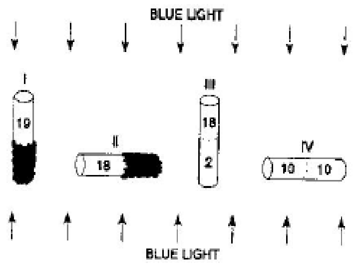
This experiment shows that flies respond to (respond means move to or away from):

a. red light but not gravity
b. gravity but not red light
c. both red light and gravity
d. neither red light nor gravity

12. *because*

a. most flies are in the upper end of Tube III but spread about evenly in Tube II.
b. most flies did not go to the bottom of Tubes I and III.
c. the flies need light to see and must fly against gravity.
d. the majority of flies are in the upper ends and in the lighted ends of the tubes.
e. some flies are in both ends of each tube.

13. In a second experiment, a different kind of fly and blue light was used. The results are shown in the drawing.



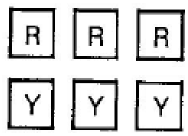
These data show that these flies respond to (respond means move to or away from):

a. blue light but not gravity
b. gravity but not blue light
c. both blue light and gravity
d. neither blue light nor gravity

14. *because*

a. some flies are in both ends of each tube.
b. the flies need light to see and must fly against gravity.
c. the flies are spread about evenly in Tube IV and in the upper end of Tube III.
d. most flies are in the lighted end of Tube II but do not go down in Tubes I and III.
e. most flies are in the upper end of Tube I and the lighted end of Tube II.

15. Six square pieces of wood are put into a cloth bag and mixed about. The six pieces are identical in size and shape, however, three pieces are red and three are yellow. Suppose someone reaches into the bag (without looking) and pulls out one piece. *What are the chances that the piece is red?*

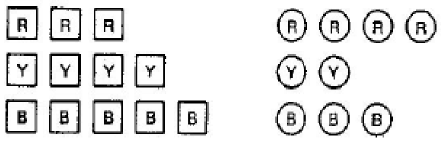


a. 1 chance out of 6
b. 1 chance out of 3
c. 1 chance out of 2
d. 1 chance out of 1
e. cannot be determined

16. *because*

a. 3 out of 6 pieces are red.
b. there is no way to tell which piece will be picked.
c. only 1 piece of the 6 in the bag is picked.
d. all 6 pieces are identical in size and shape.
e. only 1 red piece can be picked out of the 3 red pieces.

17. Three red square pieces of wood, four yellow square pieces, and five blue square pieces are put into a cloth bag. Four red round pieces, two yellow round pieces, and three blue round pieces are also put into the bag. All the pieces are then mixed about. Suppose someone reaches into the bag (without looking and without feeling for a particular shape piece) and pulls out one piece.



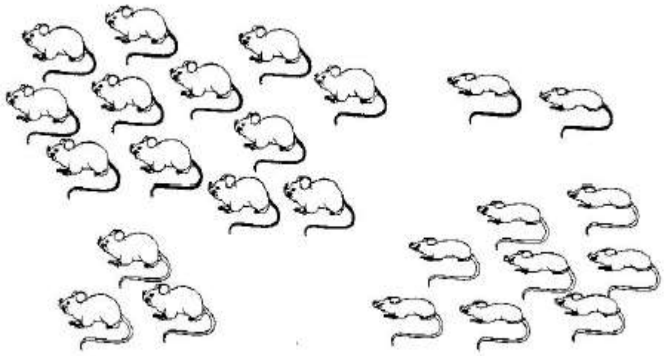
What are the chances that the piece is a red round or blue round piece?

a. cannot be determined
b. 1 chance out of 3
c. 1 chance out of 21
d. 15 chances out of 21
e. 1 chance out of 2

18. *because*

a. 1 of the 2 shapes is round.
b. 15 of the 21 pieces are red or blue.
c. there is no way to tell which piece will be picked.
d. only 1 of the 21 pieces is picked out of the bag.
e. 1 of every 3 pieces is a red or blue round piece.

19. Farmer Brown was observing the mice that live in his field. He discovered that all of them were either fat or thin. Also, all of them had either black tails or white tails. This made him wonder if there might be a link between the size of the mice and the color of their tails. So he captured all of the mice in one part of his field and observed them. Below are the mice that he captured.



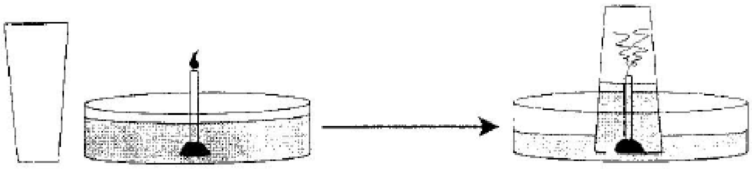
Do you think there is a link between the size of the mice and the color of their tails?

a. appears to be a link
b. appears not to be a link
c. cannot make a reasonable guess

20. *because*

a. there are some of each kind of mouse.
b. there may be a genetic link between mouse size and tail color.
c. there were not enough mice captured.
d. most of the fat mice have black tails while most of the thin mice have white tails.
e. as the mice grew fatter, their tails became darker.

21. The figure below at the left shows a drinking glass and a burning birthday candle stuck in a small piece of clay standing in a pan of water. When the glass is turned upside down, put over the candle, and placed in the water, the candle quickly goes out and water rushes up into the glass (as shown at the right).



This observation raises an interesting question: Why does the water rush up into the glass?

Here is a possible explanation. The flame converts oxygen into carbon dioxide. Because oxygen does not dissolve rapidly into water but carbon dioxide does, the newly formed carbon dioxide dissolves rapidly into the water, lowering the air pressure inside the glass.

Suppose you have the materials mentioned above plus some matches and some dry ice (dry ice is frozen carbon dioxide). *Using some or all of the materials, how could you test this possible explanation?*

a. Saturate the water with carbon dioxide and redo the experiment noting the amount of water rise.
b. The water rises because oxygen is consumed, so redo the experiment in exactly the same way to show water rise due to oxygen loss.
c. Conduct a controlled experiment varying only the number of candles to see if that makes a difference.
d. Suction is responsible for the water rise, so put a balloon over the top of an open-ended cylinder and place the cylinder over the burning candle.
e. Redo the experiment, but make sure it is controlled by holding all independent variables constant; then measure the amount of water rise.

22. What result of your test (mentioned in #21 above) would show that your explanation is probably wrong?

a. The water rises the same as it did before.
b. The water rises less than it did before.
c. The balloon expands out.
d. The balloon is sucked in.

23. A student put a drop of blood on a microscope slide and then looked at the blood under a microscope. As you can see in the diagram below, the magnified red blood cells look like little round balls. After adding a few drops of salt water to the drop of blood, the student noticed that the cells appeared to become smaller.

Magnified Red Blood Cells After Adding Salt Water

This observation raises an interesting question: Why do the red blood cells appear smaller?

Here are two possible explanations: I. Salt ions (Na+ and Cl-) push on the cell membranes and make the cells appear smaller. II. Water molecules are attracted to the salt ions so the water molecules move out of the cells and leave the cells smaller.

To test these explanations, the student used some salt water, a very accurate weighing device, and some water-filled plastic bags, and assumed the plastic behaves just like red-blood-cell membranes. The experiment involved carefully weighing a water-filled bag, placing it in a salt solution for ten minutes and then reweighing the bag.

What result of the experiment would best show that explanation I is probably wrong?

a. the bag loses weight
b. the bag weighs the same
c. the bag appears smaller

24. *What result of the experiment would best show that explanation II is probably wrong?*

a. the bag loses weight
b. the bag weighs the same
c. the bag appears smaller

APPENDIX III. : ÖRNEK AND OTHERS (2008) FACTORS OF PHYSICS DIFFICULTY

Student-Controlled Factors

	% of Students
Lack of motivation and interest	
Not studying more	
Not reading the textbook	
Not completing CHIP assignment	
Not doing practice many problems	
Working only assigned problems	
Not doing homework	
Lack of previous experience	
Lack of physics background	
Lack of higher level mathematics	

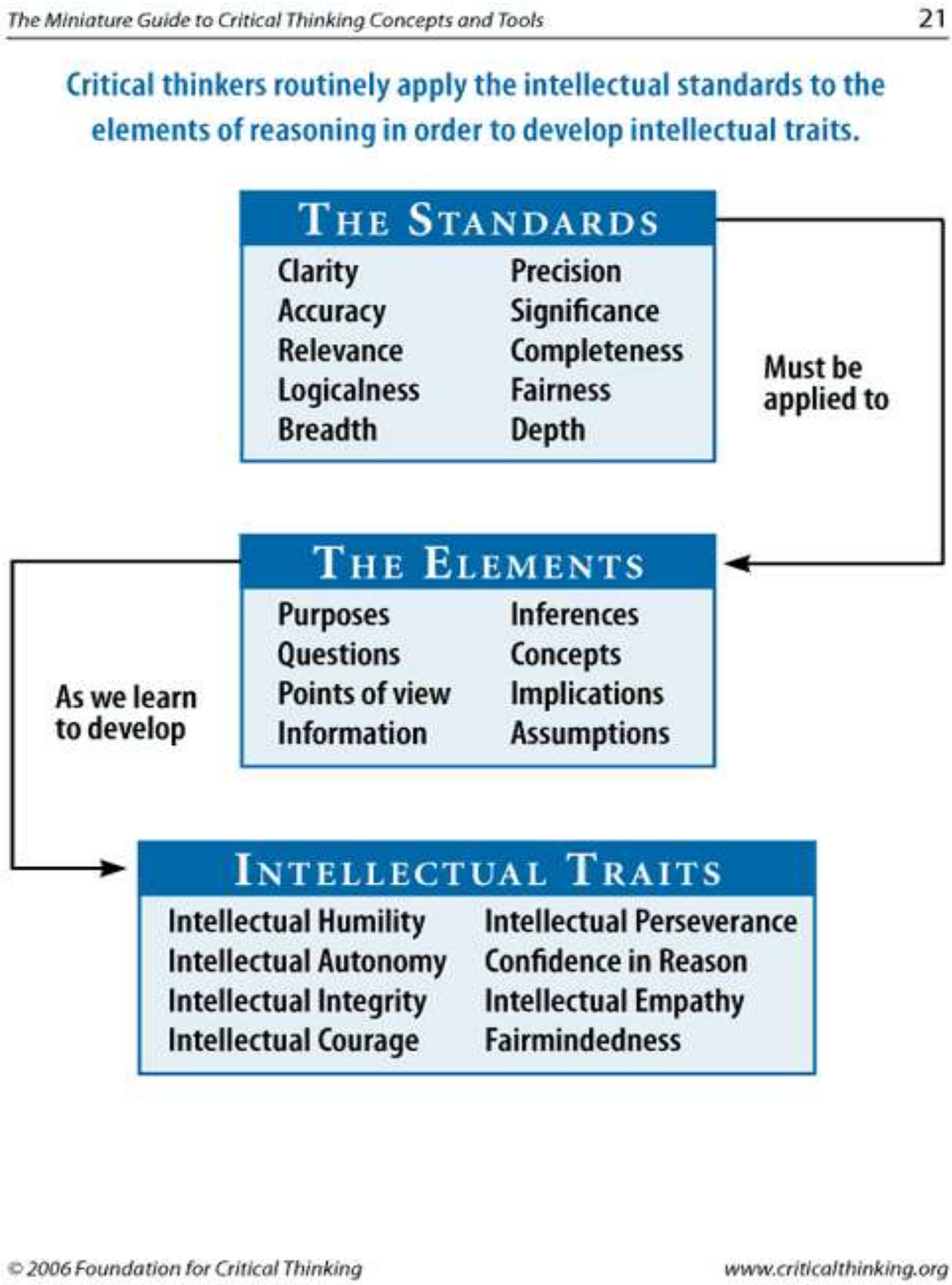
Course-Controlled Factors

	% of Students
Too much work	
Hard CHIP homework	
Lack of consistency between the lab/tutorial/lecture and homework	
Textbooks, lectures, CHIP homework questions are too complicated	
Tutorial sections are not useful	
Not enough examples, real life applications, and problem solving especially conceptual questions in class	
Being so picky on grading	
Hard questions on the exams and were not related to what solved in the class	
Poor professors	
Poor TAs	

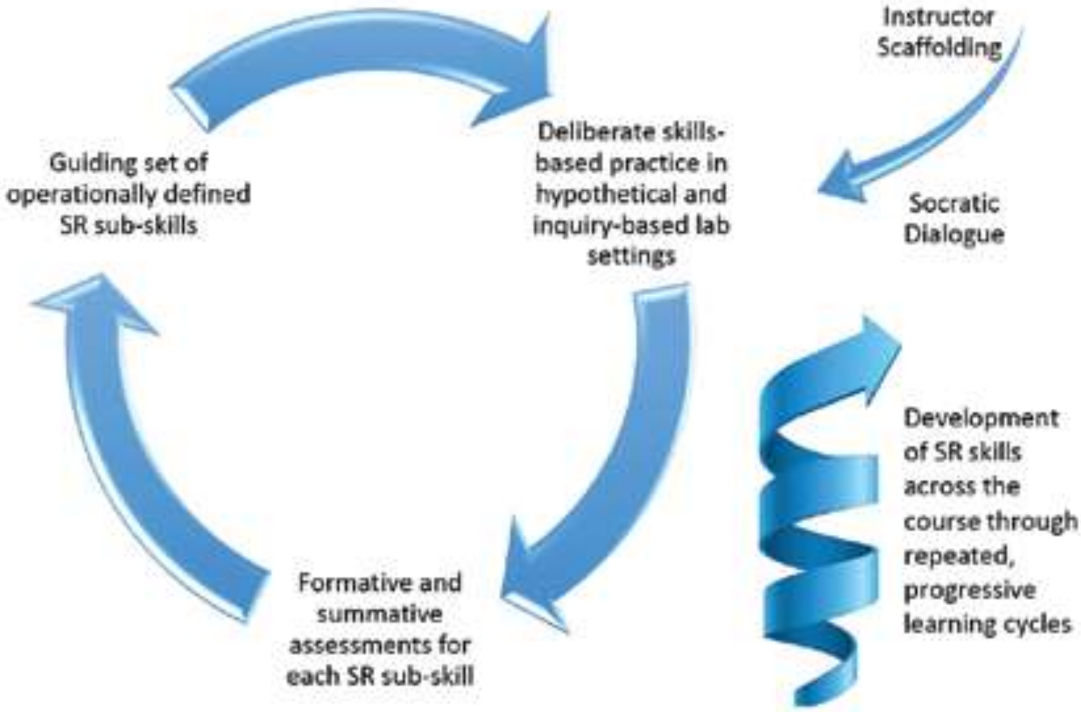
Factors Related to the Nature of Physics

	% of Students
Physics is cumulative. If you miss one concept, it is hard to grasp the next one	
Physics is very difficult subject	
There is too much material to learn	
Physics is very abstract	
Physics requires good mathematics	
Physics has too much theory	
Physics has too many formulas to be learned	
Physics has too many laws and rules	
Physics is not interesting enough	
Physics cannot be learned without mathematics background	

APPENDIX IV. THE STANDARD APPLIED TO ELEMENTS OF REASONING TO DEVELOP INTELLECTUAL TRAITS.



APPENDIX V. FEATURES OF SR CURRICULA FRAMEWORK ESTABLISHED BY (Koenig et al., 2019, P.29)


Instructor
Scaffolding
Guiding set of
operationally defined
SR sub-skills
Deliberate skills-
based practice in
hypothetical and
inquiry-based lab
settings
Socratic
Dialogue
Development
of SR skills
across the
course through
repeated,
progressive
learning cycles
Formative and
summative
assessments for
each SR sub-skill

APPENDIX VI: FRAMEWORK FOR LAB CURRICULUM FOR COV

Framework for designing lab curriculum that targets abilities in control of variables (COV).

COV skill domains	Learning outcomes (COV subskills)
Planning controlled experiments	Design a controlled experiment with 2 variables. (Low)
	Design a controlled experiment with 3 variables. (Low)
	Design a controlled experiment to determine if a variable is *influential* when it is testable based on conditions in the provided scenario. (Intermediate)
Identifying controlled experiments	Decide if a variable is *testable* when it is, in fact, testable based on conditions in the provided scenario. (Low)
	Decide if several variables are *testable* when some are influential and hidden relations exist in the provided scenario. (High)
	Decide if several variables are *influential* when some are influential and hidden relations exist in the provided scenario. (High)
Interpreting controlled experiments	Given experimental results, interpret evidence to determine if a variable is *influential* when it is, in fact, testable and influential in the provided scenario. (Intermediate)
	Use control of variables to determine if sufficient conditions exist in causal relations. (High)
Understanding indeterminacy of confounded experiments	Given experimental results, decide if several variables are *testable* when they are not testable based on conditions in the provided scenario. (Intermediate)

Retrieved from : (Koenig et al., 2019. p. 30)

APPENDIX VI. HIGH SCHOOL STREAM OF PARTICIPATING STUDENTS

High School Curriculum			Frequency	Percent
Unknown	26 %	Unknown	26	26 %
Private High Schools	24 %	American	6	5.0 %
		British	4	4.0 %
Public High Schools	50 %	Science	58	57.4 %
		Art	4	4.0 %
		Institutes of Applied technology	3	3%
		Total	101	100.0

APPENDIX VII. IELTS AND EQUIVALENT TOEFL SCORES

TOEIC	TOEFL Paper	TOEFL CBT	TOEFL IBT	IELTS
0 - 250	0 - 310	0 - 30	0 - 8	0 - 1
	310 - 343	33 - 60	9 - 18	1 - 1.5
255 - 400	347 - 393	63 - 90	19 - 29	2 - 2.5
	397 - 433	93 - 120	30 - 40	3 - 3.5
405 - 600	437 - 473	123 - 150	41 - 52	4
	477 - 510	153 - 180	53 - 64	4.5 - 5
605 - 780	513 - 547	183 - 210	65 - 78	5.5 - 6
	550 - 587	213 - 240	79 - 95	6.5 - 7
785 - 900	590 - 637	243 - 270	96 - 110	7.5 - 8
905 - 990	640 - 677	273 - 300	111 - 120	8.5 - 9
Top Score	Top Score	Top Score	Top Score	Top Score
990	677	300	120	9

Retrieved from https://sites.google.com/site/b3fink4urself/Home/ielts-score

APPENDIX VIII. STRUCTURE OF THE COGNITIVE PROCESS' DIMENSION OF THE REVISED TAXONOMY by (Krathwohl, 2002, p. 215).

1.0 Remember – Retrieving relevant knowledge from long-term memory.
1.1 Recognizing
1.2 Recalling
2.0 Understand – Determining the meaning of instructional messages, including oral, written, and graphic communication.
2.1 Interpreting
2.2 Exemplifying
2.3 Classifying
2.4 Summarizing
2.5 Inferring
2.6 Comparing
2.7 Explaining
3.0 Apply – Carrying out or using a procedure in a given situation.
3.1 Executing
3.2 Implementing
4.0 Analyse – Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
4.1 Differentiating
4.2 Organizing
4.3 Attributing
5.0 Evaluate – Making judgments based on criteria and standards.
5.1 Checking
5.2 Critiquing
6.0 Create – Putting elements together to form a novel, coherent whole or make an original product.
6.1 Generating
6.2 Planning
6.3 Producing

APPENDIX IX.

Summary of Recommended Accommodations for English Learners From Abedi and Ewers (2013)

Accommodation	Risk	Access
Traditional glossary with Spanish translations (content-related terms removed)		
Traditional glossary with Spanish translations and extra time (content-related terms removed)		
Customized dictionary/glossary in English (content-related terms removed)		Yes
Customized dictionary in English (content-related terms removed) and extra time		Yes
Computer-based test		Yes
Pop-up glossary (computer-based test; content-related terms excluded)		Yes
Modified (simplified English)		Yes
Extra time within the testing day (not combined with another accommodation)		Yes
Read-aloud of test directions in student's native language	Minor	
Picture dictionary (alone; combined with oral reading of test items in English; and combined with bilingual glossary)	Minor	
Bilingual dictionary	Minor	
Test break	Minor	Yes
Test in a familiar environment with other English language learners	Minor	
Small group setting	Minor	Yes
Read-aloud of test questions (math, science, history/social science) to student by teacher or electronic media	Minor	Yes
Spanish translation of test	Moderate	
Dual-language translation of test	Moderate	
Read-aloud of test questions (English language arts) to student by teacher or electronic media	High	
Commercial dictionary/glossary in English	High	

APPENDIX X. Representation of Interactive Engagement Teaching Approach in Khalifa University-UAE. (Hitt et al., 2014, p.12)