EFFECTIVENESS OF CONCEPTUAL CHANGE STRATEGIES IN SCIENCE EDUCATION: A META-ANALYSIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÇAĞATAY PAÇACI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
MATHEMATICS AND SCIENCE EDUCATION

JUNE 2022

Approval of the thesis:

EFFECTIVENESS OF CONCEPTUAL CHANGE STRATEGIES IN SCIENCE EDUCATION: A META-ANALYSIS

submitted by **ÇAĞATAY PAÇACI** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Mathematics and Science Education Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ___________

Prof. Dr. Erdinç Çakıroğlu
Head of the Department, **Mathematics and Science Edu.** ___________

Assoc. Prof. Dr. Ömer Faruk Özdemir
Supervisor, **Mathematics and Science Edu. Dept., METU** ___________

Assist. Prof. Dr. Ulaş Üstün
Co-Supervisor, **Mathematics and Science Edu. Dept.,**
Artvin Çoruh University ___________

Examining Committee Members:

Prof. Dr. Bilal Güneş
Mathematics and Science Education Dept., Gazi University ___________

Assoc. Prof. Dr. Ömer Faruk Özdemir
Mathematics and Science Education Dept., METU ___________

Prof. Dr. Ali Eryılmaz
Mathematics and Science Education Dept., METU ___________

Prof. Dr. Esen Uzuntiryaki Kondakçı
Mathematics and Science Education Dept., METU ___________

Prof. Dr. Nejla Yürük
Mathematics and Science Education Dept., Gazi University ___________

Date: 29.06.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Çağatay PAÇACI

Signature:

ABSTRACT

EFFECTIVENESS OF CONCEPTUAL CHANGE STRATEGIES IN SCIENCE EDUCATION: A META-ANALYSIS

Paçacı, Çağatay
Doctor of Philosophy, Department of Mathematics and Science Education
Supervisor: Assoc. Prof. Dr. Ömer Faruk Özdemir
Co-Supervisor: Assist. Prof. Dr. Ulaş Üstün

June 2022, 403 pages

There is extensive literature focusing on students' misconceptions in various subject domains. Several conceptual change approaches have been trying to understand how conceptual change occurs to help learners handle these misconceptions. This meta-analysis aimed to integrate studies investigating the effectiveness of three types of conceptual change strategies; cognitive conflict, cognitive bridging, and ontological category shift in science learning. We conducted a random-effects meta-analysis to calculate an overall effect size in Hedges' g with a sample of 218 primary studies, including 18,051 students. Our analyses resulted in a large overall effect size (g=1.10, 95% CI [1.01, 1.19], k=218, p<.001). We also performed a Robust Bayesian Meta-analysis to calculate an adjusted effect size, which specified a large effect (adjusted g=0.93, 95% CI [0.68, 1.07], k=218, p<.001). Results are also consistent across the conceptual change strategies of cognitive conflict (g=1.10, 95% CI [0.99, 1.21], k=150, p<.001), cognitive bridging (g=1.06, 95% CI [0.84, 1.28], k=30, p<.001), and ontological category shift (g=0.88, 95% CI [0.50, 1.26], k=9, p<.001). However, a wide-ranging prediction interval [0.19, 2.38] pointed out a high heterogeneity in the distribution of effect sizes. Thus, we investigated the moderating effects of several variables using simple and multiple meta-regressions. The final meta-regression model we created explained 35% of overall heterogeneity. This meta-analysis provides robust evidence that conceptual change strategies significantly enhance students' learning in science.

Keywords: Conceptual Change, Cognitive Conflict, Cognitive Bridging, Ontological Category Shift, Meta-analysis, Science Achievement.

ÖZ

FEN EĞİTİMİNDE KAVRAMSAL DEĞİŞİM STRATEJİLERİNİN ETKİNLİĞİ: META-ANALİZ

Paçacı, Çağatay
Doktora, Matematik ve Fen Bilimleri Eğitimi Bölümü
Tez Yöneticisi: Doç. Dr. Ömer Faruk Özdemir
Ortak Tez Yöneticisi: Dr. Öğr. Üyesi Ulaş Üstün

Haziran 2022, 403 sayfa

Öğrencilerin farklı konulardaki kavram yanılgılarına odaklanan, kapsamlı bir alan yazın bulunmaktadır. Çeşitli kavramsal değişim yaklaşımları, öğrencilerin bu kavram yanılgılarıyla başa çıkmalarına yardımcı olmak için kavramsal değişimin nasıl meydana geldiğini anlamaya çalışmaktadır. Bu meta-analizde, üç kavramsal değişim stratejisi türünün etkinliğini araştırarak çalışmaların sonuçlarını bütünleştirmeyi amaçlıyoruz. Bunlar, bilişsel çatışma, bilişsel köprüleme ve ontolojik kategori değişimi yaklaşımlarıdır. 18,051 öğrenciyi içeren 218 birincil çalışmadan oluşan bir örneklemle Hedges' g indeksinde genel etki büyüklüğünü hesaplamak için rastgele etkiler modeline göre bir meta-analiz gerçekleştirildi. Analizlerimiz büyük genel etki büyüklüğü ile sonuçlandı (g=1.10, %95GA [1.01, 1.19], k=218, p<.001). Aynı zamanda, düzeltilmiş etki büyüklüğünü hesaplamak için Robust Bayesian Meta-analizi gerçekleştirildi (düzeltilmiş g=0.93, %95GA [0.68, 1.07], k=218, p<.001). Sonuçlar ayrıca bilişsel çatışma (g=1.10, %95GA [0.99, 1.21], k=150, p<.001), bilişsel köprüleme (g=1.06, %95GA [0.84, 1.28], k=30, p<.001) ve ontolojik kategori değişimi (g=0.88, %95 GA [0.50, 1.26], k=9, p<.001) gibi kavramsal değişim stratejileri arasında da tutarlıdır. Bununla birlikte, geniş tahmin aralığı [0.19, 2.38], etki büyüklüklerinin dağılımında yüksek düzeyde heterojenliğe işaret etmektedir. Bu nedenle, basit ve çoklu meta-regresyon kullanarak birçok değişkenin etkilerini araştırdık. Elde ettiğimiz son meta-regresyon modeli, toplam heterojenliğin %35'ini açıklamaktadır. Bu meta-analizin sonuçları, kavramsal değişim stratejilerinin öğrencilerin fen bilimleri konularında öğrenmelerini önemli ölçüde geliştirdiğine dair sağlam kanıtlar sunmaktadır.

Anahtar kelimeler: Kavramsal Değişim, Bilişsel Çatışma, Bilişsel Köprüleme, Ontolojik Kategori Değişimi, Meta-analiz, Fen Başarısı.

To my wife, who has always supported me to pursue my goals, to my son, who is the source of my energy in life, and to my family, who raised me in peace and love...

ACKNOWLEDGMENTS

I want to thank my supervisor, Assoc. Prof. Dr. Ömer Faruk Özdemir and my co-supervisor Assist. Prof. Dr. Ulaş Üstün for their support and guidance throughout the study, and I would like to extend my thanks and appreciation to my committee members, Prof. Dr. Ali Eryılmaz and Prof. Dr. Nejla Yürük, for their valuable guidance.

Sharing my experiences about the process with my twin brother Timur Paçacı and getting feedback from him has always motivated me. I extend my thanks to him.

I want to thank my friends Belkıs Garip and Dilber Demirtaş for providing valuable assistance with coding and feedback. I also would like to thank my friends Hasan Yücel Ertem and Öskan Öcalan for their support.

I should also thank Assist. Prof. Dr. Ertuğrul Özdemir and Assoc. Prof. Dr. Nurcan Cansız for their feedback during the coding process.

There are too many authors whose studies were used in this study and supported me in my coding process. I extend my appreciation, especially to Michelene T. H. Chi and James D. Slotta, for their extra support.

Finally, I would like to express my gratitude to my devoted wife, Arzu Elif Paçacı, for her dedication and patience and raising a lovely boy Doruk Vefa.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CCS: Conceptual Change Strategy

CCM: Conceptual Change Model

CMA: Comprehensive Meta-analysis

TFM: Trim and Fill Method

RRM: Random-Random Model

FRM: Fixed Random Model

AR: Agreement Rate

ES: Effect Size

CI: Confidence Interval

PI: Prediction Interval

k: Number of Samples

FSN: Fail-safe N

SSCI: Social Sciences Citation Index

PQDT: ProQuest Dissertations and Theses

NTC: National Thesis Center

VIF: Variance Inflation Factor

RoBMA: Robust Bayesian Meta-analysis

PET: Precision-Effect Test

PEESE: Precision-Effect Estimate With Standard Errors

CHAPTER 1

INTRODUCTION

Different learning theories have focused on various aspects of learners while explaining how students learn science. Some theories have focused on the interactions with social environments, such as the sociocultural theory of learning, and others have focused on the interactions with physical environments, such as the constructivist view of learning. Nevertheless, there seems to be a consensus among the theories that students bring to class has a vital role in what students get out of the course. Maybe the most explicit reference to what students bring to class was made by Ausubel (1968). In meaningful learning theory, Ausubel argues that the essential factor for learning new concepts is learners already existing knowledge structure. Learners should relate new knowledge with the relevant concepts that they already know. For meaningful learning, new knowledge should interact with learners' prior knowledge.

Learning was defined as constructing new knowledge on early cognitive structures shaped by experiences, observations (Karpudewan et al., 2017), or interactions with other people (Leach & Scott, 2002). Therefore prior knowledge obtained in different ways is an essential tool for learning. The importance of prior knowledge while learning new conceptions triggers researchers to understand students' prior knowledge. Since the late 1970s, an immense body of literature has focused on what students know and how they influence further learning. The literature accumulated throughout the years explicitly showed that students' prior knowledge may not always be consistent with scientific claims. In other words, prior knowledge may not help learners construct the new knowledge. Furthermore, prior knowledge may construct mis-ideas or mis-reasoning while constructing new knowledge. For this reason, all learning processes cannot lead learners to grasp scientific knowledge (Illeris, 2018).

Previous researches have consistently shown that students have different experiences and ideas before entering formal instruction, which may contradict scientific claims (Abdullah et al., 2017; Barke et al., 2009; Heng & Karpudewan, 2017). Therefore an extensive body of research has focused on understanding students' pre-instructional ideas. These ideas are generally called misconceptions. However, different researchers have adopted different terminologies for pre-instructional ideas because of different epistemological explanations of learning. Some of the most common terminologies are alternative conceptions (Driver & Easley, 1978; Hewson & Hewson, 1989, Mungsing, 1993), naïve conceptions (Baillargeon, 2004; Caramazza et al., 1981; Vosniadou, 1994), initial conceptions, misconceived knowledge (Chi et al., 1994; 2008), pre-instructional beliefs (Chinn & Brewer, 1993), misconceptions (Griffiths & Preston, 1992; Posner et al., 1982; Vosniadou, 1994), intuitive knowledge (diSessa et al., 1998), spontaneous reasoning, and children's science (Karpudewan et al., 2017). Misconceptions are commonly defined as student conceptions producing systematic patterns of error (Vosniadou, 2019). The student misconceptions may stem from textbooks, instructional materials, analogies, scientific language, teachers' misconceptions, drawings, students' experiences, peers, friends, or parents (Barke et al., 2009; Karpudewan et al., 2017; Sinatra & Pintrich, 2002). They cause learning difficulties and block comprehending scientific ideas (Chinn & Brewer, 1993; Smith et al., 1993). This is why overcoming misconceptions became one of the major objectives of instruction. However, it is evident that overcoming misconceptions requires more effort than detecting them.

In traditional instructions (usually in lecture, direct, or expository instruction), it is common to use students' already existing conceptions and experiences to help them make sense of new knowledge (Chamber & Andre, 1995). Students use their preconceptions to construct new knowledge even if they are not compatible with the accepted scientific explanations (Seyedmonir, 2000). Since traditional instruction is inadequate to present the inconsistencies between what students know and scientific ideas, more specific instructional models were needed.

Defining students' conceptions before the instruction may provide more effective pedagogies to remediate misconceptions for teachers. Hence, once pre-instructional concepts are identified, instruction must be geared toward overcoming misconceptions (Mason et al., 2017). In this sense, the most common model, the conceptual change model, was proposed by Posner et al. (1982), which was derived from Kuhn's (1970) theory about the development of science and Piaget's (1964) theory of learning described in terms of assimilation and accommodation processes. According to Posner et al. (1982), there are four conditions before a conceptual change is likely to occur. The first condition is expressed as dissatisfaction with prior knowledge. It is critical to provide dissatisfaction with prior knowledge by presenting anomalous data. Anomalous data is expected to help students realize their existing conceptions are unproductive.

The second condition is that presented knowledge must be intelligible. To consider an alternative conception, the learner should understand what the new knowledge means in a given context. Thirdly, a new concept must be plausible for students to give full credit to the new conception. Plausibility can be explained as the capacity of presented knowledge to solve problems. Fourthly, in addition to resolving issues, the presented knowledge should be fruitful and suggest new insights and discoveries.

After the conceptual change model (CCM) proposition, more attention was given to understanding students' prior knowledge structures (Duit et al., 2008; Özdemir, 2004). In the literature, there seem to be three general frameworks about the nature of students' prior knowledge as knowledge-as-theory perspective (coherent knowledge), knowledge-as-elements perspective (knowledge in pieces), and knowledge as ontological categories perspective (ontological categories). These perspectives do not entirely reject each other but lead to divergence in exploring the possible roles of prior knowledge in the conceptual change process (Özdemir & Clark, 2007; Vosniadou, 2008).

The traditional CCM mainly stresses the coherent nature of knowledge. This model accepts knowledge embedded in science and mathematics frameworks that are not fragmented and may be confirmed or refuted by new data. Students' knowledge coherence may also be confirmed or refuted by daily experiences when they face confirming or conflicting situations. In order to foster conceptual change, the CCM claims that students must become aware of the inconsistencies between their prior knowledge and scientific knowledge. The CCM proposes assimilation and accommodation to create an expected change in students' conceptions. This model has wide support within the science education community (Chinn & Brewer, 1993; Duit et al., 2008; Limon, 2001; Vosniadou, 1994; Vosniadou et al., 2008).

Regarding instructional implications of CCM, it is expected that the learner will become dissatisfied with prior conceptions when faced with conflicting knowledge (Posner et al., 1982). Therefore it makes sense to create instruction over the dissatisfaction process to change misconceptions. One of the most effective instructional strategies that emerged from the CCM is the cognitive conflict strategy (Duit et al., 2008). Specifically, the strategy of creating dissatisfaction with students' prior knowledge is a 'classical approach' in conceptual change literature (Vosniadou, 2017). This strategy proposes to remediate misconceptions during the instructional intervention by satisfying four fundamental conditions of CCM, which were (1) there must be dissatisfaction with prior knowledge; (2) new conception must be intelligible; (3) the new conception must be plausible and (4) the new conception should suggest a fruitful solution. Linnenbrink and Pintrich (2003) argued that, during this strategy, students could reveal their misconceptions and recognize the discrepancy between new knowledge and their current knowledge structure. Cognitive conflict aims to dissatisfy students with their nonscientific prior knowledge to achieve conceptual change. Several research studies proved that cognitive conflict is more effective in acquiring advanced conceptions and facilitating conceptual change than traditional strategies. (Carlsen, 1989; Guzzetti et. al., 1993; Jensen et al, 1996; Launey, 1995; Liao & She, 2009; Liu, 2008; Loon et al., 2015; Mason et al., 2019; Sanger & Greenbowe, 2000; Sungur et al., 2001; Uzuntiryaki & Geban, 2005).

Nevertheless, there are also a significant number of critiques on the effectiveness of cognitive conflict. For example, Dreyfus et al. (1990) revealed that successful students could benefit from cognitive conflict, but unsuccessful students could develop negative attitudes toward the conflict process. Vosniadou et al. (2008) introduce the major limitation of cognitive conflict as the assumption that conceptual change happens in a short period. She argues that cognitive conflict requires a long instructional time. Supporting Vosniadous's arguments, Chan et al. (1997) showed that even if students face anomalous data, they do not necessarily experience conflict and resistance to changing their prior knowledge. Smith et al. (1993) stated that cognitive conflict is based primarily on inappropriate prior knowledge to generate conceptual change. It is incompatible with the constructivist perspective. New knowledge can only be constructed upon prior knowledge. Several research studies also show that cognitive conflict is not necessarily an effective method even when compared to traditional methods (Saigo, 1999; Seyedmonir, 2000; Södervik et al., 2015; Tsai, 2003; Windschitl & Andre, 1996; Yilmaz, 2007; Zohar & Kravetsky, 2005).

On the other hand, the knowledge-in-pieces perspective opposes the idea of the coherent nature of prior knowledge. diSessa (2002) argues that students' prior knowledge is composed of knowledge in pieces rather than a strong coherent knowledge structure. Therefore, he opposes taking dissatisfaction with prior knowledge as the main focus of conceptual change. Instead, productive knowledge pieces should be in the foreground. diSessa describes the fragmented knowledge elements held by students in terms of phenomenological primitives (p-prims). P-prims are defined as the intuitive equivalent of physical laws. However, p-prims cannot be categorized as right or wrong; it depends on the context. This is why p-prims are interpreted as productive or unproductive according to their appropriateness for particular contexts. diSessa considers p-prims the smallest cognitive structure entity that creates more complex systems. Knowledge-in-pieces perspective describes the conceptual change as the revision, refinement, and reorganization of the current cognitive structure by using the productive p-prims (Brown, 1995; diSessa, 1993; Smith et al., 1993; Özdemir & Clark, 2007).

The major implication of the knowledge-in-pieces perspective is that conceptual change can be achieved by using the relevant p-prims or e-prims (explanatory primitives) through analogies, models, and classroom interactions. Therefore, instruction should be based on productive intuitive knowledge pieces rather than unproductive ones, such as in the case of cognitive conflict strategy (diSessa et al., 1993; Smith et al., 1993). Sharing similar concerns about the incompatibility of cognitive conflict strategy with the constructivist perspective, some researchers focused on constructing a bridge rather than a conflict with prior conceptions to promote conceptual change (Yaman, 2013). For example, Clement et al. (1989) propose that students not only have misconceptions but also have productive conceptual resources enabling them to bridge prior and new knowledge. Student prior knowledge also allows the appropriate evaluations for new contexts. Therefore, as an instructional implication, bridging analogies and explanatory models can be effective conceptual change strategies to overcome misconceptions by providing connections between prior and new knowledge (Clement, 1993). Cognitive bridging aims to use students' productive prior knowledge rather than dissatisfied ones to overcome misconceptions and trigger conceptual change as an instructional strategy. Several research studies provided evidence that cognitive bridging is more effective in acquiring advanced conceptions and facilitates conceptual change than traditional strategies (Clement, 1993; Çetingül & Geban, 2011; Diakidoy & Kendeou, 2001; Gokhale, 1996; Köseoğlu & Bayır, 2012; Li, 2008; Stavy, 1991; Woloshyn et al., 1994; Yaman, 2013; Yilmaz et al., 2006).

Nevertheless, some other studies also show that cognitive bridging is not necessarily effective compared to traditional methods (Gayeta & Caballes, 2017; Sota, 2012). Another critique of the classical approach was put forward by Chi et al. (1994) by focusing on the ontological nature of students' misconceptions. According to Chi, prior knowledge may conflict with new knowledge, but conceptual change may not be realized solely by the dissatisfaction process. Students' misconceptions may be deeply rooted in how they ontologically categorize their conceptions regarding the matter, process, or mental states. Chi (2008) stated that since the nature of students' categorical structures may differ from the scientists', the dissatisfaction process is insufficient to promote conceptual change. This is because dissatisfaction only leads to a partial change in beliefs and does not promote a change in ontological categories. Therefore, Chi claims that the traditional CCM is not necessarily effective in achieving conceptual change without tapping students' ontological categories. Similarly, Henderson et al. (2018) believe that learners keep the incorrect ontological categories even if contradictory or refuting knowledge (anomalous data) is presented to achieve conceptual change. Learners will remain committed to that ontological classification if a concept is perceived with an incorrect ontology. For example, some learners describe an object losing heat as the loss of "hot particles," where the object cools down over time as its total number of "hot particles" decreases. In such a case, trying to convince learners by presenting contradictory or refuting information would be futile for achieving conceptual change.

As an instructional implication, the ontological perspective proposes a more direct process of conceptual change which is called ontological category shift. For an ontological category shift, it is critical to delineate the ontological nature of students' prior knowledge. According to Chi (1994), the learning process includes the categorization of concepts in terms of ontological characteristics, and robust misconceptions mainly stem from the miscategorization of core conceptions. During this process, learners should reorganize their prior knowledge at the ontological/categorical level. Ontological category shift aims to change students' miscategorized conceptions to achieve conceptual change as an instructional strategy. Several research studies have proved that ontological category shift is more effective in acquiring advanced concepts and facilitates conceptual change than traditional strategies (Akgül, 2010; Chiu & Lin, 2005; Çoruhlu & Çepni, 2015; Slotta & Chi, 2006; Uzuntiyaki & Geban, 2005). Nevertheless, some research studies also show that ontological category shift is not necessarily effective, even concerning traditional methods (Charles, 2003; Erdmann, 2001; Yang et al., 2012).

1.1 The Rationale and the Purpose of the Study

Scientific literature means every written, audio, or spoken material covering scientific data on a specific field collected throughout the years. It provides a way to summarize and evaluate findings, compare and contrast different authors' views, highlight exemplary studies, concludes, analyze the included studies, generates new knowledge, and provide gap analysis in the field (Lasserson et al., 2019). But, in order to obtain the above gains, literature should provide reliable, relevant, and up-to-date research evidence for scientists, academicians, learners, practitioners, and policymakers. On the other hand, the results of any study can lead to false impressions or inappropriate conclusions due to generalization, reliability, or validity problems. Therefore, it is possible to reach confusing results by using studies one by one, even for a specific subject of literature (Petticrew and Roberts, 2006). At the same time, the rapidly growing volume of literature makes it impossible to assess the vast number of primary studies simultaneously. Lasserson et al. (2019) argue that obtaining an up-to-date summary of the knowledge is blocked by an increasing number of studies so that it can be derived from the literature just by systematic assessment of primary studies. Therefore, combining the whole body of literature is critical to make up-to-date, reliable, and justifiable decisions on conceptual change literature.

Another dimension of the conceptual change literature review is that, as the preceding paragraphs show, there are different conceptual change perspectives about the nature of students' pre-instructional knowledge. These perspectives do not completely reject each other but lead to a variance in the possible roles of prior knowledge in the conceptual change process. Different knowledge perspectives propose different types of instructional interventions to achieve conceptual change. Nevertheless, empirical studies on the effect of conceptual change strategies show contradictory results. Cooper et al. (2019) inform us that scientific literature contains related studies in the same field to verify and extend the previous findings. Results across these studies are not identical, implying the diversity between associated studies. Therefore, there is a need to integrate findings across related studies on the conceptual change literature by using holistic methods.

Comprehensive systematic methods such as research review, systematic review, narrative review, or research synthesis are needed to provide a holistic view. Research synthesis is one of the most common, comprehensive, and practical methodologies for overall analysis in literature. The goal is the most notable difference between research synthesis and other research methods (literature review, systematic review, narrative review). It aims to combine empirical studies to create generalization, the applicability of previous findings, and obtain new knowledge through integrating previous findings (Cooper et al., 2019; Magliocca et al., 2014). Its role is not only to provide a comprehensive comparison between the effectiveness of different perspectives but also to assist with deepening the meaning of the existing knowledge (Hunter & Schmidt, 2004). Research synthesis may provide an opportunity to build more powerful results for studies done on conceptual change literature. Along with the cumulative nature of science, using the opposite conclusions of the studies gains an objective viewpoint of this field. At this point, research syntheses enhance existing studies and provide new dimensions to conceptual change literature.

Accurate statistical evaluation is prominent for review studies also. Even though some disadvantages are argued about research synthesis in literature, it is an effective integrating method to combine primary studies (Cooper & Hedges, 2019). As a quantitative research synthesis method, meta-analysis is one of the most common and comprehensive ones that include well-defined statistical procedures. Primary studies may have some limitations in terms of statistical evaluation for practitioners and application processes. Rosenthal and Dimatteo (2001) criticized that it is common to yield conflicting numerical findings from primary studies about practical issues in education, medicine, and other bio-psychological and socio-psychological disciplines. Researchers commonly study these issues, but the findings are varying and sometimes contradictory. For example, are computer-based methods effective on student attitudes? Do vitamins prevent lung cancer? Does exposure to waves emitted from mobile phones increase the risk of cancer? What is the relationship between gender and math achievement? Does exercise relieve the side effects of medicines? As the importance of meta-analysis, in order to resolve such conflicting quantitative evidence, accurate estimation of descriptive statistics is necessary.

As another limitation of primary study findings, there are significant concerns about statistical significance while evaluating primary studies. Especially the sensitivity of statistical significance for sample size creates confusion about the findings of primary studies. For example, two studies with the same effect size may have very different significance depending on their sample size. Or studies with a large number of sample sizes may have very large effect size values even if they have very similar statistical significance versus small sample size studies. Nevertheless, meta-analysis alleviates these problems by focusing on effect size values with a predefined weighting model. Therefore, meta-analysis plays a crucial role in literature to make comprehensive, valid, and reliable decisions. In this sense, it summarizes the main effects from the primary quantitative studies by using effect sizes. This integrated main effect by numerous studies can provide useful knowledge for policymakers and researchers.

The main purpose of this study is to compare the effectiveness of conceptual change strategies on students' science academic achievement at the elementary, middle, high school, and university levels by conducting a meta-analysis. This study focused on science achievement in experimental studies, comparing the conceptual change strategies with traditional teaching strategies. Studies published or unpublished in English or Turkish languages between 1984 and 2021 years were included. Additionally, the effectiveness of different types of conceptual change strategies on science achievement were examined by considering different moderator variables such as publication type (doctoral dissertations, master theses, journal articles and conference papers), region (Africa, America, Asia, Europe, Turkey), subject (Biology, Chemistry, and Physics), educational levels (elementary, middle school, high school, undergraduate), experimental designs (poor experimental, quasi experimental and true experimental), instrument type (adapted test, preexisting test and researcher developed test), question type (mix type question, objective type question and open-ended type question), teacher training (unstated or stated), material used during intervention (text, computer, hands on material), school type (private school and public school), school location (rural and urban), sampling method (nonrandom sampling and random sampling), researcher effect (researcher is one of the teachers, researcher is not one of the teachers and researcher is the only teacher), teacher effect (Same teacher for control and experimental groups, different teachers for control and experimental groups), number of assessment question tier (one-tier, two-tier and three-tier), conceptual change methods (cognitive bridging, cognitive conflict and ontological category change), treatment verification (unstated or stated), measuring outcome (achievement test and misconception test), year, intervention length, treatment intensity, sample size and class size. Specific research questions generated for this study were as follows.

1.2 Research Questions

1. What is the overall effectiveness of conceptual change strategy on science achievement compared to traditional teaching methods?
2. What is the overall effectiveness of cognitive conflict on science achievement compared to traditional teaching methods?
3. What is the overall effectiveness of cognitive bridging on science achievement compared to traditional teaching methods?
4. What is the overall effectiveness of ontological category shift on science achievement compared to traditional teaching methods?
5. What is the role of design characteristics on the effectiveness of CCS on science achievement?
    5.1 What is the role of experimental design on the effectiveness of CCS on science achievement?
    5.2 What is the role of teacher training on the effectiveness of CCS on science achievement?
    5.3 What is the role of design characteristics on the effectiveness of CCS on science achievement?
    5.4 What is the role of the researcher effect on the effectiveness of CCS on science achievement?
    5.5 What is the role of teacher effect on the effectiveness of CCS on science achievement?
    5.6 What is the role of treatment verification on the effectiveness of CCS on science achievement?
6. What is the role of publication characteristics on the effectiveness of CCS on science achievement?
    6.1 What is the role of publication type on the effectiveness of CCS on science achievement?
    6.2 What is the role of publication year on the effectiveness of CCS on science achievement?

7. What is the role of intervention characteristics on the effectiveness of CCS on science achievement?
    7.1 What is the role of the type of conceptual change strategy on the effectiveness of CCS on science achievement?
    7.2 What is the role of material on the effectiveness of CCS on science achievement?
    7.3 What is the role of the subject domain on the effectiveness of CCS on science achievement?
    7.4 What is the role of intervention length on the effectiveness of CCS on science achievement?
    7.5 What is the role of intervention intensity on the effectiveness of CCS on science achievement?
8. What is the role of subject characteristics on the effectiveness of CCS on science achievement?
    8.1 What is the role of the region in the effectiveness of CCS on science achievement?
    8.2 What is the role of sample size on the effectiveness of CCS on science achievement?
    8.3 What is the role of class size on the effectiveness of CCS on science achievement?
    8.4 What is the role of education level on the effectiveness of CCS on science achievement?
    8.5 What is the role of school location on the effectiveness of CCS on science achievement?
    8.6 What is the role of school type on the effectiveness of CCS on science achievement?
9. What is the role of measurement characteristics on the effectiveness of CCS on science achievement?
    9.1 What is the role of outcome measure on the effectiveness of CCS on science achievement?

9.2 What is the role of instrument type on the effectiveness of CCS on science achievement?

9.3 What is the role of question type on the effectiveness of CCS on science achievement?

9.4 What is the role of the number of tiers on the effectiveness of CCS on science achievement?

1.3 Definition of Important Terms

Science Achievement is the students' level of understanding of scientifically accepted knowledge at the level of knowing, understanding, comprehending, applying, or analyzing by measuring their quantitative scores on a given assessment test.

Conceptual Change Strategy is the process in which students' scientifically wrong conceptions are replaced or reorganized with scientific conceptions through instructional processes. It focuses essentially on creating a scientifically true conceptual understanding by means of eliminating students' pre-instructional scientifically wrong conceptions.

Cognitive Conflict is a type of conceptual change strategy which focuses on students' scientifically wrong conceptions to trigger conceptual change by using dissatisfaction processes. According to this strategy, learners are not satisfied with their knowledge structure when they face a situation that cannot be explained with their prior knowledge. This makes it easy for learners to acquire new knowledge (Hewson, 1992).

Cognitive Bridging is a type of conceptual change strategy that focuses on students' productive conceptual resources enabling them to bridge prior and new knowledge to trigger conceptual change. From a constructive perspective, dissatisfaction with inappropriate prior knowledge is not a proper instructional strategy for constructing new knowledge. Therefore productive prior knowledge can be used to diagnose common misconceptions.

Ontological Category Shift is a type of conceptual change strategy which focuses on reorganizing the students' mis-categories at the ontological/categorical level to overcome misconceptions by replacing them with the proper ones. Since the categorical structures of learners may be formed by intuitive rendition based on experiences, the dissatisfaction or cognitive bridging processes are insufficient to promote conceptual change (Chi, 2008). Therefore, reorganizing the students' ontological categories enables them to overcome misconceptions to achieve conceptual change.

Traditional Teaching Strategy refers to a variety of expository instructions which do not include any type of alternative teaching method.

Meta-analysis was firstly defined by Glass (1976) as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding" p.3. This powerful statistical method provides a comprehensive realization related to literature by using effect sizes as a unit of analysis. Well-established and detailed coding process enables us to resolve questions that cannot be answered with any single study.

Meta-regression is a statistical method to investigate the relationship between moderator variables and a dependent variable. The process is similar to regression analysis, except that moderators are at the level of the primary study rather than the level of the subject, and the dependent variable is effect size rather than sample data (Borenstein et al., 2009).

Effect Size is the degree of strengths for treatment effect. It enables assessing the differences between groups (d family) or measuring the strength of a relationship (r family). The Hedges'*g* is used for the effect size value in this study which is the difference between the treatment group mean and the control group mean scores divided by the pooled standard deviation obtained from primary study scores (Hedges, 1981). This score is recommended if the groups are dissimilar in size by weighting each group's standard deviation concerning its sample size (Ellis, 2010).

1.4 Significance of the Study

There is no perfect study without any errors (Hunter & Schmidt, 2004). Schmidt and Le (2004) argued that primary study imperfections mainly stem from methodological issues. At the same time, meta-analysis eliminates errors in methodology and enables researchers to estimate measurement bias in primary studies. Lipsey and Wilson (2001) noted that meta-analysis as a research synthesis is less vulnerable to methodological errors and free from weaknesses of conventional research review techniques. On the other hand, primary studies in the literature on conceptual change strategies (CCS) also have methodological imperfections like limited sample size, poor research designs, unstandardized data collecting procedures and instruments, subjectively determined limitations of the study, and so on. Therefore, inconsistent and controversial results stemming from methodological imperfection are common in this field also. This study aimed to alleviate methodological imperfections by adopting a meta-analytic perspective to estimate measurement biases in CCS.

Different instructional implications of CCS yield very divergent effectiveness on achievement (Brown, 1995; Slotta & Chi, 2006; Smith et al., 1993; Tsai, 2003; Zohar & Kravetsky, 2005). It is also not surprising to come across statistically significant or nonsignificant results. However, some studies also indicate reverse outcomes; that is, the traditional method is more effective than CCS (Saigo, 1999; Seyedmonir, 2000; Windschitl & Andre, 1996; Zohar & Kravetsky, 2005). As stated, controversial results are so common in conceptual change literature. The significance of meta-analysis is that it is highly preferred to disclose inconsistencies and controversies in the literature concerning the effectiveness of instructional interventions to improve students’ science achievement, attitude, motivation, and skills in science education (Braver, Thoemmes, & Rosenthal, 2014). Since single primary studies do not comprehend whole research fields, subject characteristics, and methodologies, it is impractical to reach a conclusion accepted by all due to instructional implementations. In order to find consistent patterns within conspicuously contradictory results across single studies for CCS, meta-analysis can enhance comprehension to provide a more detailed investigation. Therefore it can be used to resolve apparent inconsistencies and controversies in CCS literature concerning the effectiveness of instructional interventions (Glass, 1976; Hunter & Schmidt, 1990).

Conceptual change literature relates to misconceptions, epistemological and ontological knowledge perspectives, and learning issues. Therefore very detailed process needs to detect possible characteristics of CCS due to its nature. Since any single study may focus on searching a few predefined research questions, the effects of moderators that exist due to the nature of the study may be neglected by researchers. For example, how much degree may learn outcomes affect by sample characteristics like education level or socio-economic status? Or is CCS more effective for chemistry than physics concepts? Or is the conflict perspective more effective than the bridging perspective? Any single study cannot identify these study-specific moderators. Rather, meta-analysis is an effective statistical technique to inform about essential moderating characteristics of primary studies by investigating moderator variables. Rosenthal and DiMatteo (2001) introduced that meta-analysis highlights the patterns of findings by examining the relationship between main questions and moderators of interest. Therefore, meta-analysis allows us to formulate potential characteristics of studies that may not be noticed in single primary studies.

Borenstein et al. (2009) underline that conventional reviews cannot synthesize the p-values with different statistical results. Researchers often prefer to report p-values that imply keeping mostly statistical significance in perspective. This tendency mainly stems from the fact that significant results are good and nonsignificant are trivial. Moreover, it is common to reach both statistically significant and non-significant results in the field of CCS (Sanger & Greenbowe, 2000; Tsai, 2003). On the other hand, the goal in any science literature is to verify and extend the findings by using the cumulative nature of knowledge. Rosenthal and DiMatteo (2001) stated that cumulating very small effects could draw drastic results, especially in medical studies. Similarly, this tendency leads to inconsistent or controversial evaluations of the effectiveness of CCS. Since meta-analysis allows combining small effects that researchers neglect, it provides an opportunity to reach a summary effect that includes both significant and non-significant findings for the conceptual change literature without losing data.

Another advantage of practicing meta-analysis on CCS is analyzing the strength of practical significance. Statistical significance is so sensitive to sample size rather than focusing on the degree of treatment effect. Moreover, the sample sizes in the literature are highly different from each other. Therefore, using statistical significance, comparing and reaching a common treatment effect across primary studies isn't very sensible. On the other hand, a meta-analysis that informs us to report standardized effect sizes for any primary study may show the degree of effectiveness for CCS. Especially in science education literature and medical science, the degree of treatment effect is highly preferred rather than whether it exists (Borenstein et al., 2009; Rosenthal & DiMatteo, 2001).

Gap analysis is another strength of meta-analysis studies in the literature. Some fields are studied intensely, but researchers do not focus on some, and the growing number of studies makes it difficult to notice gaps in the literature. Therefore, a systematic and comprehensive method constitutes a general perspective to define gaps in the literature. In this sense, the conceptual change literature is composed mainly of three fields of perspectives as cognitive conflict, cognitive bridging, and ontological category shift (Yaman, 2013). But, the dominant perspective that researchers commonly prefer is cognitive conflict (Duit et al., 2008; Limon, 2001). This should not imply that researchers might neglect the other perspectives. Therefore, a comprehensive gap analysis is a need for CCS. In this sense, a meta-analysis on CCS lead researchers to study areas that can contribute more to the literature with less effort.

This meta-analysis also identified the weaknesses of discrete types of conceptual change strategies that work on how students deal with their prior knowledge and new knowledge. In this sense, this meta-analysis enabled CCS to know how efficient student achievement is. Beyond the comprehensive analysis of literature, another strength of the meta-analysis study is investigating the weaknesses and problems concerned with areas of science education. Researchers have examined the role of disclosing weaknesses. Chan (1993) pointed out that weaknesses and problems defined thanks to conflicting results enable the growth of scientific knowledge. Borenstein et al. (2009) defined that research synthesis focuses on the idea of resolving conflicts in the literature and attempting to identify underlying issues for future research.

It is valuable to state that there is no comprehensive meta-analysis study on CCS; instead, researchers focused on a specific instructional method of conceptual change. For example, Armağan (2011) mainly focused on conceptual change texts as the only source of CCS. Similarly, Gelen (2015) reviewed just assisted conceptual change materials like concept maps, concept cartoons, and conceptual change texts. Guzzetti et al. (1993) investigated text structures and discussion methods. Mufid et al. (2020) investigated the effectiveness of conceptual change texts, and Schroder and Kucera (2021) stressed the impact of cognitive conflict on science. The common point for previous meta-analyses is that they mainly focused on the Posner et al. (1982) description of cognitive conflict for the scope of CCS.

On the other hand, there are severe criticisms that there are already existing conceptual change strategies derived from different knowledge perspectives except for the dissatisfaction process as described in the literature review part. This is mainly due to the different epistemological and ontological approaches that make it more complex to clearly define conceptual change as an instructional strategy in science education. But these strategies have such discrete effectiveness in summarizing the field efficiently. In this sense, detailed and comprehensive reviews must be performed to collect conceptual change studies. Thus, this meta-analysis study fills the prominent gap in review studies done about conceptual change literature.

Additionally, it must be clearly stated that there is no multiple meta-regression on the CCS. In many ways, the reasons for dispersion on effect sizes are masked by unconsidered moderators that are effective on true variance in mean effects. The potential pitfall of simple meta-regression is that acknowledging unisolated moderators causes an overly simplistic assessment of the effect value. Multiple meta-regression analyses tried to determine the possible confounders to estimate the effect of the independent variables on dependent variables.

From a methodology perspective, previous meta-analyses have ignored the influence of traditional instruction. Controlling such a variable related to design also causes significant changes in the results. In other words, managing the effects of conventional instruction by ascribing its attributes to the control group is one of the essential features of this study to achieve more standardized results.

Apart from the above evaluations, this meta-analysis emphasized how effective conceptual change strategies are in science education. Thus, the findings could help science educators to enhance notions about CCS that could meet the students' needs of remediating their non-scientific prior knowledge. In this sense, the main idea behind this meta-analysis is to compare the effectiveness of conceptual change and traditional strategies on student achievement. We also interpreted the moderator variables, which are conceptual change strategies, publication type, region, subject domain, educational levels, experimental designs, instrument type, question type, teacher training, material type, school type, school location, sampling method, researcher effect, teacher effect, number of tiers, treatment verification, outcome measure type, publication year, intervention length, intervention intensity, sample size, and class size. This meta-analysis will provide comprehensive guidance for further studies on conceptual change literature.

CHAPTER 2

LITERATURE REVIEW

Osborne (1982) stated that it is common for teachers and students to bring previously developed conceptions and beliefs conflicting with scientifically accepted ones into the instructional settings. Since the early 1970s, researchers' interest in students' non-scientific concepts has grown exponentially (Pfundt & Duit, 1991). Defining, revealing, and reconstructing non-scientific knowledge became critical in science education. Especially, Piaget's (1964) arguments about assimilation and accommodation processes have provided the opportunity to study more to define non-scientific knowledge. Piaget's theoretical constructs of assimilation, accommodation and equilibration processes reveal the importance of understanding prior knowledge structures while acquiring new knowledge. On the other hand, due to different epistemological and ontological perspectives, there is no consensus on the nature of students' prior knowledge structure. Although many phrases refer to non-scientific prior knowledge, there is not a comprehensive expression. Rather, non-scientific prior knowledge is called by different names due to different theoretical arguments like alternative conception (Driver & Easley, 1978; Hewson & Hewson, 1989), naïve conceptions (Caramazza et al., 1981), preconceptions, alternative frameworks, naïve perceptions (Driver & Easley, 1978; Linder, 1993; Wandersee, Mintzes & Novak, 1994), intuitive conceptions (Burbules & Linn, 1988; Clement et al., 1989), and critical barriers (Hawkins, 1985). Nevertheless, the misconception is one of the most common labels to refer to students' non-scientific knowledge structure. In this sense, Fisher (1985) has defined misconceptions as the prior beliefs and notions inconsistent with current scientific viewpoints. They are the natural phenomena that exist before being expressed in the classroom by learners (Posner, Strike, Hewson, & Gertzog, 1982). Martin et al. (2002) defined misconceptions more comprehensively as nonscientific beliefs, naïve theories, mixed conceptions, conceptual misunderstandings, or preconceived notions. Clement et al. (1989) give a more constructivist voice to misconception as “creative constructions of the individual” (p.555). Additionally, science education researchers have warned against overusing the term misconceptions. Strike (1983) pointed out that misconceptions are confused with any kind of student’s errors; instead, they are the source of errors that conflict with scientific conceptions. Vosniadou (1988) used the terms ‘experiential beliefs’, ‘preconceptions’, and ‘misconceptions’ to refer to students’ misunderstandings. Shortly, different theoretical perspectives bring about different definitions of student misconceptions, but it is critical to understand the source of this type of knowledge in the learning process.

In literature, it is highly accepted that students’ naive knowledge may be the origin of misconceptions. Chan (1993) gives an example of a misconception on free-falling objects that any learner may hold the pre-Newtonian conception that the weights of objects increase as they come closer to the ground (since learner thought that gravity increases), speed increases, and so acceleration occurs. Or misconceptions can stem from the interaction with the social environment. For Piaget, since parents and teachers are considered credible sources of knowledge, learners may not question the factual information but rather directly acquire them. Therefore, undefined expressions within a community may lead to the development of misconceptions among children.

Driver and Easley (1978) introduced that misconceptions may be constructed during the formal instructions due to misunderstandings or teachers’ misconceptions. Textbooks, student background knowledge, erroneous simulations or demonstrations (Helier & Finley, 1992; Ivowi, 1984), student experiences, and everyday language (Kaltakcı & Eryılmaz, 2010; Nelmes, 2005) can be possible sources of student misconceptions. Vosniadou (1988) noted that, during the intervention, incorrect assimilation of new knowledge could also account for the construction of misconceptions. New concepts introduced during instructions can be easily manipulated by students holding naive theories, which generates synthetic models, another form of misconceptions.

Reconstruction of misconceptions is critical in that student misconceptions may responsible for incompatible evaluations and judgments of scientifically accepted ones during the knowledge construction process. Therefore, several theoretical approaches see misconceptions as a barrier that makes it difficult to learn science. Zietsman and Hewson (1986) stressed that students have difficulties acquiring new conceptions because they have deeply-rooted personal experiences that lead to misconceptions. Eggen and Kauchak (2004) pointed out that misconceptions have the form related to prior knowledge and are difficult to change. Nelmes (2005) also addressed that students are reluctant to share their prior knowledge if they are unsure. Therefore it is critical to elicit and eliminate misconceptions through well-designed instructional practices. But, these knowledge and beliefs are difficult to reveal, reorganize or eliminate. Therefore current literature on science education mainly tries to identify misconceptions and overcome them by designing more effective instructional interventions.

2.1 Conceptual Change Strategy as a Teaching Method

Based on the empirical studies on students' conceptions, students have often been viewed as holding flawed ideas that are strongly held, that interfere with learning, and that instruction must confront and replace (Smith et al., 1993). Carey (2008) poses that one must not lose sight of the issue that each of the later conceptual systems is definable to the initial ones. This is one of science education's central tenets of the knowledge reconstruction process. As an instructional vision for this issue, the content of the instructional process shifts along with the preinstructional knowledge structure. Therefore, effective instruction should consider both acquiring scientific knowledge without leading to any misconception and overcoming already existing misconceptions. In a more assertive expression, conceptual change aims at an intervention process of obtaining new knowledge by overcoming non-scientific knowledge structure without resulting in misconceptions.

The conceptual change process is firstly clarified by Piaget as introducing assimilation, accommodation, and equilibration concepts in the context of the evolution of knowledge and became the focus of attention by researchers. His stage theory about cognitive development also contributes to the knowledge restructuring process. He states that there is no doubt that 8-year-olds have more cognitive resources than 4-year-olds to consider more aspects of phenomena. He emphasizes the importance of prior knowledge structure concerning the knowledge construction process as well as experiences and explicit processes.

Thomas Kuhn's historical account of conceptual change in scientific theories also enhances the historical perspective stated in his book 'The Structure of Scientific Revolution (Kuhn, 1970). He stressed the historical development of scientific conceptions by showing transitions from Aristotelian view to Galilean, from Galilean to Newtonian, or from Newtonian to quantum physics (Özdemir, 2004; 2015). After the 1980s', it became clear that there was a need to reflect the conceptual change model for educational purposes. Consequently, more instruction-oriented approaches have been put forth. Based on initial arguments, Posner et al. (1982) introduced the Conceptual Change Model (CCM) based on initial arguments. This model basically attempts to create dissatisfaction with pre-instructional misconceptions to acquire new knowledge.

In the late 1990s, diversified theoretical and methodological explanations came to clarify the mechanism of conceptual change research by using psychology, philosophy, and the history of science. The rapidly evolving area of conceptual change suggests discrete models explaining how misconceptions are formed, revealed, and remediated with instructional interventions.

From a coherent framework perspective, Vosniadou (1988; 1989) defines the conceptual change in the light of cultural and social effects. According to Vosniadou et al. (2008), representations of the physical world cannot be fragmented; rather, it is a coherent system that is continuously re-confirmed by everyday experiences in the light of the social and cultural background. Therefore, the change in knowledge structure for a coherent framework perspective is difficult. In this sense, conceptual change mechanisms are based on avoiding internal inconsistency between the incoming information and his/her prior knowledge.

With a supportive argument to Vosniadou, Hewson (1992) describe conceptual change as the knowledge construction process by the extinction of the former state of knowledge structure and adding new knowledge to what is already there. Social experiences enable us to construct knowledge in a coherent and valuable way. But, students' prior knowledge changes the perception and interpretation of the new knowledge. The same events may be perceived and interpreted differently by students with different prior knowledge. The conceptual change mechanism helps to grow awareness of the consistency and tenacity between students' prior knowledge and new knowledge of the physical world. These arguments mainly concentrated on reconstructing students' existing knowledge structure under the constraints of the coherent framework theory.

Contrary to the idea about the coherent or theory-like nature of students' prior knowledge structure, some researchers argue that students' misconceptions are not theory-like but fragmented in nature like diSessa (1993), Smith et al. (1993) and Hammer (1996). diSessa (2008) proposes the knowledge in pieces perspective to disclose the mechanisms for conceptual change. He believes that knowledge structure is not so coherent and is strongly integrated with each other; rather, new knowledge is gradually developed by adding new knowledge to the previous knowledge pieces. Therefore, integrating the fragmented knowledge pieces by using productive prior experiences enables to achieve of conceptual change.

With a supportive argument for the fragmented knowledge perspective, Smith et al. (1993) focused on the productive roles of student experiences with a constructivist view to explain the conceptual change mechanisms. Smith et al. argued that students have both flawed and productive prior knowledge elements, but a constructivist perspective requires the use of productive knowledge elements to make sense of the new knowledge

On the other hand, Chi (1992) adopted an ontological perspective to knowledge acquisition by defining prior knowledge with respect to lateral and hierarchical ontological categories. She argues that when students' misconceptions conflict with new ideas at different category levels, refutation will not promote conceptual change. Misconception and the correct conception should be assigned to the same lateral or hierarchical categories to achieve a successful conceptual change.

The theoretical approaches for knowledge structure suggest different conceptual change mechanisms while defining educational interventions. For example, while the classical conceptual change model proposed cognitive conflict as an instructional strategy (Posner et al., 1982), the knowledge-in-pieces perspective proposes cognitive bridging (diSessa, 1993; Smith et al., 1993). Similarly, ontological perspective proposes ontological category shift (Chi, 1992; 1993; 1995; 2002).

2.2 Conceptual Change Strategies

2.2.1 Cognitive Conflict

The classical conceptual change approach suggests that introducing a new conception may not be sufficient to acquire new knowledge; therefore, instructional practices should pay special attention to students' prior knowledge (Limon, 2001). In this sense, dissatisfaction with prior knowledge has a central role in classical conceptual change (Hewson, 1992). This approach proposes cognitive conflict as a well-structured instructional method to overcome non-scientific prior conceptions (Hewson & Hewson, 1984).

Among various conceptual change models defined in the literature, several research studies propose cognitive conflict as an effective instructional strategy which was firstly introduced by Posner et al. (1982). This strategy is based on Piaget's assimilation and accommodation notions for the cognitive reorganization of knowledge. According to Posner's model, conceptual change is an epistemological model describing the conditions of a successful conceptual change which are eliciting prior knowledge, promoting dissatisfaction, addressing new knowledge, and achieving conceptual change. The work of Posner et al. (1982) became the leading model that guides researchers to advance instructional practices in the field of conceptual change.

The cognitive conflict strategy is the most prevailing instructional strategy emerging from the conceptual change model. Limon (2001) defines cognitive conflict as an instructional strategy to promote conceptual change through anomalous data or contradictory information. Presenting the anomalous data trigger students' dissatisfaction with their prior knowledge. Thus, the dissatisfaction process also triggers reorganizing, restructuring, or changing prior knowledge with the new one. Sinatra and Mason (2008) also define cognitive conflict strategy as a way of stimulating students' understanding and revising inadequate current conceptions about a phenomenon or event by introducing anomalous data.

According to Posner et al. (1982), the radical phase of the restructuring process to achieve conceptual change is accommodation. In order to achieve accommodation, the following four conditions should be satisfied. The first condition is dissatisfaction with existing knowledge. Hewson (1992) put forward the notion that dissatisfaction is a reason for changing the status of prior knowledge. If individuals are satisfied with their current knowledge, they tend to retain their current concepts. Cognitive conflict stimulates learners to question the effectiveness of their prior knowledge. Consequently, cognitive conflict makes students realize the inadequacies of their prior knowledge so that it should be extended or exchanged with scientific knowledge.

The properties of new knowledge are also critical to activating conceptual change. The second condition proposes that new knowledge should be intelligible (the learner should know what the new knowledge exactly is). Posner et al. (1982) stated that intelligibility requires understanding concepts, terms, symbols, or identifying representations of what the functions and theories are saying.

Thirdly, new knowledge should be plausible. Posner et al. (1982) define it as the new knowledge that should be consistent with current scientific knowledge. In other words, it can be defined as the capacity of presented knowledge to solve problems.

Fourthly, new knowledge should be fruitful and able to suggest new insights and discoveries when crossing new situations. When new knowledge is both intelligible and plausible, students may interpret new experiences to resolve problems. Therefore, fruitful new knowledge provides an accommodation process that is more persuasive and permanent for students.

The cognitive conflict perspective aims to trigger the conceptual change process by using anomalous data to force learners to grapple with alternative responses. In order to achieve conceptual change, cognitive conflict has also been used with different instructional materials such as hands-on, text-based, and computer-based materials. For example, hands-on materials like laboratory experiments (Niaz & Chac'o, 2003), inquiry-oriented discussions (Anyanvu, 2008), or fostering cognitive development in simulated conditions (Budiman et al., 2014) are different types of hands-on materials. Text-based materials are also prevalent in conflict perspectives like conflict maps (Sungur et al., 2001; Tsai, 2003), conceptual change texts (Çakır et al., 2002), drawing texts (Launey, 1995), or refutational texts (Diakidoy et al., 2015; Mason et al., 2017). The computer-based materials in cognitive conflict are computer-supported modeling (Li, 2008), computer-assisted instruction (Jensen et al., 1996), or computer simulations (Baser, 2006).

In the literature, several quantitative studies give statistical evidence on the effectiveness of cognitive conflict in science achievement. These studies also clearly inform us about the moderating factors like sample, design and publications characteristics, measurement, and intervention processes. Additionally, the literature also reflects the various instructional implications of cognitive conflict. Therefore, it is essential to refer to some of the cognitive conflict studies we stated below.

Carlsen (1989) studied one hundred four university students (47 male and 57 female) for three days as an experimental study. Participants do not get any physics courses or any other similar study on electricity. The control group has been exposed to traditional instruction that is teacher-oriented. There is no assisting instructional tool except traditional text during the course. The design for the two groups is quasi-experimental. The experimental group has exposed to the conceptual change instruction designed with a cognitive conflict strategy. Both cognitive conflict texts and compute simulations were used during the instruction. Multiple-choice tests are used as an assessment tool. The course duration for the two groups is equal and applied by the same instructor. According to pre-post test scores, this study indicates that a combination of text and computer simulations designed concerning cognitive conflict strategy enables remediate of incorrect pre-instructional conceptions of the electricity concept.

Similarly, Sanger and Greenbowe (2000) studied the effects of cognitive conflict instruction implemented by computer animations for misconceptions about electron flow in aqueous solutions. The study shows that conceptual change instruction enriched with computer animations effectively detects and eliminates misconceptions on electron flow in aqueous solutions. Moreover, students may improve their visual understanding and use symbolic representations better. Therefore, researchers argue that cognitive conflict strategy may enhance conceptual understanding better than traditional strategies.

The above results reveal that computer simulations effectively take an interest in students and may lead to an improved understanding. Therefore, instructional practices taking into account students prior knowledge are an effective instructional tool. Overall results imply that computer-based environments using cognitive conflict strategy improve students' science acquisition by remediating prior knowledge.

In a similar study, Jensen et al. (1996) worked on the effectiveness of cognitive conflict strategy with experimental research on sixty-three students in two different entry-level science courses at a university. They also designed a learning environment based on the conflicting processes for students' common misconceptions about diffusion and osmosis subjects. Results showed that students in the cognitive conflict group outperformed the traditional group students.

Uzuntiyaki (2003) also focused on the effectiveness of cognitive conflict strategy on the chemical bonding concepts. This study consisted of 42 9th-grade students in a private school. In the control group, students were instructed with traditionally designed chemistry texts. In the experimental group, cognitive conflict strategy was used to remediate students' misconceptions. She also used analogies to facilitate student comprehension of the chemical bonding concept. The study shows that cognitive conflict strategy leads to gain significantly better acquisition of chemical bonding concepts and remediate misconceptions than traditionally designed instruction.

Nwankwo and Madu (2014) examined the effect of cognitive conflict over traditional instruction on students' conceptual changes in heat and temperature. The study was conducted with 249 secondary students from two schools. The experimental group received cognitive conflict-based instruction, while the control group instructed traditionally. The study's findings show that the cognitive conflict strategy is significantly more effective in eliminating misconceptions about heat and temperate as compared with the traditional one.

Al Khawaldeh (2013) also designed an experimental study to indicate the role of conceptual change text based on cognitive conflict strategy on 10th-grade students' understanding of genetics concepts by comparing traditionally designed instruction. Researcher examined how cognitive conflict-based instruction help to revise students' prior knowledge and struggle with their misconceptions to achieve conceptual change. Student dissatisfaction with their existing conceptions allows them to think about their prior knowledge and reflect on it. In this way, students may have reasonable time to disclose and express their ideas, examine the plausibility and utility of their prior conceptions, and put into practice new ideas in a context familiar to them.

Çil and Çepni (2012) presented that they observed a significant change in learning for physics (nature of science) when they used conceptual change texts satisfying Posners' assumptions. They also recommended that the conflict-based conceptual change methods are the most effective way of teaching the nature of science.

Another dimension of these studies is their effect sizes for the same statistical data for conflict-based texts. For example, Özmen et al. (2009) found the t value for achievement is ($t(56)=2.195$, $p=.032$), Koparan et al. (2010) found the t value for achievement is ($t(44)=3.003$, $p=.004$), Geban and Bayır (2000) found the t values for achievement is ($t(48)=3.654$, $p<.05$), Çakır et al. (2002) found the t values for achievement is ($t(82)=4.000$, $p<.05$). These studies show the difference in the significant size of the cognitive conflict strategy.

In addition to the above studies, some researchers also empirically show that conflict-based conceptual change texts are more effective than traditional methods (Berber & Sarı, 2009; Beerenwinkel et al., 2011; Çetingül & Geban,2011; Demircioğlu, 2009; Geban & Bayır, 2000; Kıngır et al., 2013; Özkan, 2013; Özmen et al., 2009).

Although several studies confirm the effectiveness of instructions designed with cognitive conflict strategy, contradictory results also exist about the role of this strategy as an instructional strategy.

Seyedmonir (2000) did an experimental study on the text-based conceptual change method, which is the Cognitive Reconstruction of Knowledge Model, based on the cognitive conflict process. A three-week intervention on 425 undergraduate students has been done on Newtonian physics. The study suggests that cognitive conflict-based instruction does not have any superiority over traditional one for students with a prior physics background to advance learner cognitive engagement. Therefore, the researcher stated that text-based instruction based on cognitive conflict strategy might not necessarily eradicate college students' misconceptions about Newtonian law of motion subjects.

As a similar finding, Zohar and Kravetsky's (2005) study was done on about 240 students in the eighth and ninth grades about a biology topic. Although high-level students benefit from cognitive conflict instruction, the results showed that there is no significant effect for low academic level students. Therefore, the study also implies that teaching with cognitive conflict may not be effective for any group.

Chamber and Andre (1995) tested the effectiveness of the classical conceptual change approach in learning science effectively. They found that text-based cognitive conflict strategies do not improve student learning of physics subject (electricity) if the learner is high or low interest, but they should be moderately interested. Also, it is unclear to generalize results to different grades to state the effectiveness of instruction.

Yılmaz (2007) perform a comparative study on conceptual change text and traditional instruction on student achievement about genetic concepts. The conceptual change text provides students with exposure to satisfy four conditions of the cognitive conflict process. The instructional implication indicated that students might not undergo conceptual change concerning the genetics concepts. Therefore, there are significant critics on the issue of the role of cognitive conflict strategy.

Another experimental study on the effectiveness of cognitive conflict strategy is implemented by Tsai (2003) to compare the impact of conflict maps on remediating students' alternative conceptions versus traditional instruction. Ninety-seven eighth graders were assigned to experimental or control groups randomly. The researcher gathered data through a two-tier test. The results show that a conflict map designed concerning four conditions of cognitive conflict has no impact on students to construct conceptual change for electric circuits.

Södervik et al. (2015) examined refutational texts' effect on 171 university students' understanding of photosynthesis. The control group was exposed to the traditional text also. The refutational text aims to refute students' nonscientific prior knowledge by exposing the dissatisfaction process. Later, the new knowledge should be intelligible, plausible, and fruitful (Posner et al., 1982). The researcher stated that cognitive conflict-based texts have no effect on remediating successful students' prior knowledge and understanding of photosynthesis but rather just low achieved students profit from cognitive conflict-based instruction. Therefore, it is difficult to say that the cognitive conflict strategy is effective for all students. Some experimental studies also support these results (Saigo, 1999; Windschitl & Andre, 1996).

The above statements imply that although cognitive conflict strategy may promote the comprehension of scientific understanding, this perspective has contradictory results. Therefore, more constructive oriented conceptual change perspectives emerged for instructional implications.

2.2.2 Cognitive Bridging

Significant progress has been made to advance new theoretical arguments against conflict perspectives in conceptual change literature. The dissatisfaction process of cognitive conflict strategy has the assumption that student knowledge structure is coherent. However, several researchers, such as Smith et al. (1993), argued that the constructivist perspective does not confirm the dissatisfaction with prior knowledge to generate new knowledge but rather, it needs to be utilized to advance learning. diSessa (2002) also argues that the implication of the constructivist perspective places the position that dissatisfaction with existing knowledge is inadequate to allow students to grasp new knowledge. It is critical to take into account students' productive prior knowledge. diSessa (2014) also argues that students' prior knowledge is composed of pieces rather than a coherent framework, and according to the constructivist perspective, it is impractical to take dissatisfaction as the main focus instead productive knowledge pieces should be in the foreground.

As an alternative instructional implication different from the classical approach, the cognitive bridging strategy is based on a knowledge-in-pieces perspective. Cognitive bridging aims primarily to facilitate conceptual change by making use of productive prior knowledge. diSessa believes that since students use their preinstructional and fragmented knowledge structures to construct more advanced scientific knowledge, it is effective to use productive prior knowledge to achieve conceptual change.

Although there is no precise definition for cognitive bridging strategy, it can be shortly defined as using the productive prior knowledge to construct and impose scientific knowledge without focusing on the conflicting processes. In this study, we used the term "cognitive bridging strategy" inspired by Yaman (2013) and Vidak, Odžak, and Mešić (2019) to refer to instructional practices using students' productive knowledge elements to overcome students' misconceptions. This term was chosen because it implies a link between existing knowledge and new knowledge to achieve conceptual change and reveal its position against cognitive conflict strategy. The central assumption is that students come to class with many resources gained from daily life experiences. These resources may provide better acquisition of new knowledge. Several studies have demonstrated the effectiveness of this strategy (Clement, 1993; Gokhale, 1996; Li, 2008; Stavy, 1991; Yaman, 2013).

diSessa (2008) state that some instructional implications should be considered concerning the effectiveness of the cognitive bridging perspective. Firstly, adequate time needs for better conceptual understanding and to achieve profound results from instruction. The cognitive bridging strategy accepts conceptual change as a longer-term process in contrast to the conflict perspective.

Secondly, the richness of conceptual resources should be used productively rather than for dissatisfaction. The bridging perspective implies a link between existing knowledge and new knowledge to achieve conceptual change. Regarding instructional implications, it is critical to attend to students carefully in a classroom environment by using relevant experiences. According to Clement (1993), analogies effectively trigger relevant experiences in the learning process even if there are naïve concepts in some contexts causing misconceptions. The critical argument is that learner should activate their prior knowledge to modify, displace, replace or suppress it. Otherwise, developing new conceptions may not be possible.

Thirdly, one of the main concerns of bridging perspective is that coaching meta-conceptual awareness enables to development of scientific knowledge by constructing on prior knowledge (Vosniadou et al., 2008). In this way, the learner can differentiate productive prior knowledge pieces. An effective coaching process provides more healthy learning for conceptual change.

Bridging perspective constantly cites the constructivist approach and aims to build new knowledge on productive prior knowledge. In order to activate a productive construction process and enhance the acquisition of new concepts, different instructional materials are taken advantage such as hands-on, text-based, and computer-based materials. The different instructional tools can be used for hands-on materials like analogical reasoning during experiments (Stavy, 1991), analogical models (Pekmez, 2010), or comparing models (Aykutlu & Şen, 2011). Additionally, there are also different forms of text-based materials like elaborative text (Woloshyn et al., 1992), discussion texts (Diakidoy & Kendeou, 2001), and analogies on worksheets (Dilber & Düzgün, 2008). Since the 2000s, the facilitating properties of computers have improved instructional material sources like analogy simulations (Şendur et al., 2008), conceptual change analogies (Karakethudaoğlu, 2010), and computer modeling (Li, 2008).

Several quantitative studies give statistical evidence on the effectiveness of cognitive bridging in science achievement in the literature. These studies also inform us about the possible moderating factors like sample, design, publication, measurement, and intervention characteristics.

In current literature, cognitive bridging frame may not be directly used as an instructional strategy. But related studies use this term like a Ph.D. study by Yaman (2013). This study was done on the effects of instructions-based cognitive bridging and cognitive conflict strategies across traditional instruction on 9th-grade students' understanding of force and motion concepts. The study shows that cognitive bridging is a more effective strategy for conceptual understanding of physics concepts than instructions based on traditional lecturing. On the other hand, if we compare cognitive conflict and cognitive bridging strategies, there is no significant difference in the effectiveness of these strategies on the same topic.

Bridging analogies are one of the main instructional implications of the cognitive bridging strategies. These analogies use bridging cases to reach the targeted conception. One of the notable experimental studies in this field is done by Clement (1993) on three areas of mechanics. He designed an experimental study on 205 high school students to overcome alternative conceptions about forces on moving objects. Clement argues that diSessa's knowledge in pieces perspective enables us to use correct anchoring intuitions to bridge with prior knowledge and new conceptions.

Thus, the cognitive bridging strategy can encourage conceptual change and may be thought of as a form of guided constructivism.

One of the studies focusing on the use of cognitive bridging strategy was Stavy (1991), who used analogies to overcome conservation of matter misconceptions. The researcher defines the role of analogies as a knowledge-building process by using students' existing intuitive knowledge. The primary focus is bridging productive prior knowledge with new knowledge. The researcher designed an experimental study on a chemical concept for 74 middle school students to provide evidence for the effectiveness of the bridging strategy. The study reveals that supporting and activating students ' productive intuitive knowledge is an effective strategy to achieve conceptual change.

Similarly, Gokhale (1996) studied anchoring analogies in overcoming misconceptions about electrical quantities in basic electronic circuits through a hands-on experiment. The study was performed on 46 students with 23 in experimental and 23 in control groups at the university level. "The experimental group was given a lab demonstration using a device that was specially designed to explain the particle analogy for the electron. The control group was given a demonstration without it" (p.11). This enables to trigger a cognitive bridging process. The study shows that the cognitive bridging strategy supports eliminating misconceptions and advancing learning on electricity concepts.

Yılmaz et al. (2006) emphasized the efficacy of bridging analogies for remediating misconceptions about mechanics concepts. The study is based on anchoring analogies to reach the targeted concept by profit by analogical reasoning. Researchers stated that the bridging strategy helps students to modify their deeply held prior knowledge to more scientific conceptions. As a result of the study, the percentage of students having misconceptions about gravity significantly decreased. Therefore, it is prominent to signify the role of cognitive conflict strategy in a valuable contribution to remediating students' misconceptions.

Li (2008) argued that the use of cognitive perturbation strategy is enhanced by computer modeling, which yields from the bridging analogies and constructivist perspective. The dominant idea behind this strategy is fostering conceptual change by disclosing prior knowledge through model building and appropriate conceptual anchors. In contrast to cognitive conflict, students are expected to use their productive prior knowledge to construct models. This experimental study indicates that learning science through model building based on bridging strategy enables students to acquire better the nature of science and the practice of scientific knowledge.

Although several studies confirm the effectiveness of instructions designed with a cognitive bridging strategy across other strategies, contradictory results also exist about the role of this strategy as an instructional strategy.

Gayeta and Caballes (2017) conducted an experimental study with 50 major science students to investigate the effect of flipped classroom instruction. They proposed that flipped classroom environment based on a constructivist perspective enables students to use their productive prior knowledge to construct scientific knowledge. Active engagement of the learner is crucial in that prior knowledge is elaborated, modified, and changed with respect to context. But the effectiveness of this strategy can still be criticized in such that identifying misconceptions is possible but correcting misconceptions is not. Therefore, a cognitive bridging strategy should be examined more to decide on certain intervention effects.

2.2.3 Ontological Category Shift

Chi (1992) defines knowledge clusters as categories that are prominent in learning mechanisms. She focuses on the idea that concepts have attributes and inherit features coming from their categories. In this way, the learner may categorize or assign a concept to a category to make inferences about unfamiliar concepts. It is meaningful when knowledge is in practice and consistent with pre-defined properties for these categories. For example, a bird has the property of flying, laying eggs, and having a beak. Any object with these properties should be considered in the bird category. If objects have no common properties, the categories of these objects can be thought of as ontologically different (Slotta et al., 1995). Therefore, we can say that parrot and crow are in the same category, but parrot and monkey are not.

Chi describes the role of categories in the conceptual change process by introducing two assumptions. Firstly, if learners have no obvious category to assign a new concept, they assign it to the next higher level of the appropriate category. For instance, learner encounters a bird with four wings and does not know that it's a kind of bird, the observer would categorize it at the next level up as a kind of animal. This informs us that there is a hierarchical categorization.

Socondly, Chi (1994) proposes that if learners have no obvious category to assign a new concept, they assign it to a new lateral category instead of a hierarchical category. Lateral category is defined as the same level categories (Keil, 1981; Chi, 1997). For example, birds and reptiles are two lateral categories that include subcategories and have a common higher-level category called animal. Their attributes are different, but there is no hierarchy among them. In the above example, when learners encounter a four wings bird they do not know, they assign it to a reptile rather than an animal. The ontological category mistakes also cause mistakes in categorical inferences and attributions. Therefore, ontological category mistakes may account for the existence of robust misconceptions and create a barrier to scientific understanding. In this sense, conceptual change instruction should focus on ontological category mistakes.

Chi (1992) defines conceptual change for ontological category mistakes in two forms: non-radical change (Figure 2.1) and radical change (Figure 2.2). During non-radical change, there is no need to reassign the ontological category rather, there needs to be a reorganization of the knowledge domain in the same category. This is the most common type of change in knowledge for our daily experiences. Chi gives an example of this change process that a shrunk can be thought of as a raccoon with respect to its fur and size. The transformation between them is a type of non-radical change.

On the other hand, an ontological shift is needed in radical change (Figure 2.2). The nodes are represented by different shapes in Figure 2.2 from the original tree. It is a situation where the concept is gradually shifting to another lateral category. For example, confusion between living things with artifacts like a real dog and a toy dog. Chi (1992) uses the reassignment expression to describe this change. Therefore, Chi refers to radical conceptual change when she defines the conceptual change process.

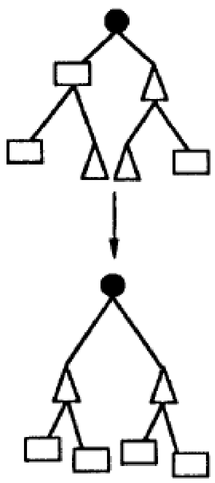


Figure 2.1 Chi's (1992, p.135) denotation of nonradical conceptual change

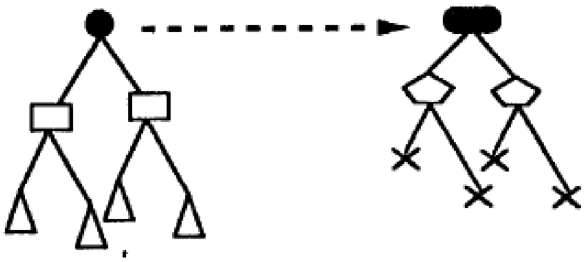


Figure 2.2 Chi's (1992, p.135) denotation of radical conceptual change

As a different perspective from classical conceptual change, the ontological category process proposes that misconceptions arise from incorrectly assigning concepts in a lateral category. In this sense, the ontological category adopts the conceptual change process markedly different from classical conceptual change. Chi and Slotta (1993) indicated that creating conflict may not necessarily provide conceptual change. They also address the idea that conceptual change is possible by removing misconception, which is the miscategorization of knowledge in the absence of a correct lateral category. This is a progressive and gradual process rather than a direct accommodation. Chi and Roscue (2002) define conceptual change as the shift of miscategorized knowledge from one ontological category (miscategory) to a workable (true) ontological category.

The major point of the ontological category process is creating radical conceptual change. Such a change requires transformation between ontological categories. Learners should change their knowledge in miscategory into a scientifically true one. Therefore, instructional practices triggering the conceptual change process should be more structured. The instruction should;

i) begin with describing the attributes of the existing ontological category
ii) Secondly, the attributes of the new ontological category should be defined
iii) Thirdly, the learner should understand the meaning of individual concepts at new ontological categories to advance new conceptual understanding.
iv) Finally, learners should reassign the concept from the previous ontological category to a new ontological category to assimilate the new knowledge (Chi & Slotta, 1993).

It is common to profit from several instructional materials from an ontological perspective also. In the scope of current literature, these materials can be categorized into three groups hands-on (Berg, 2010), text-based (Erdmann, 2000; Uzuntiryaki & Geban, 2005), and computer-based (She & Liao, 2010; Slotta & Chi, 2006; Yang et al., 2012). Each material type also leads to different instructional processes and causes yield different effect values. That is why the possible effect of material should be considered in this study to yield a better understanding of moderator variables.

In the literature, several quantitative studies give statistical evidence on the effectiveness of ontological category shift. These studies also clearly inform us about the moderating factors like sample, design, publication, measurement, and intervention characteristics. Additionally, the literature also reflects the various instructional implications of ontological category shift. Therefore, it is essential to refer to some of the studies reviewed below.

The effectiveness of ontological category shift on the conceptual change process also is a controversial issue. There are comprehensive studies showing the effectiveness of this strategy. For example, Uzuntiyaki and Geban (2005) designed an instruction by using Chi's ontological category change frame. Firstly misconceptions stem from miscategories of knowledge is described to reassign the concepts located into a miscategory into the scientific category for matter concepts. Later, the attributes of the new category were disclosed with the help of conceptual change texts. Thirdly, students were expected to form concept maps to reassign their ontological categories. Finally, discussions on the new ontological categories were held to enhance understanding of the matter concept. The results of this study showed that ontological category shift was an effective strategy to help students change their misconceptions.

Parallel evidence was also provided by the experimental study of Slotta and Chi (2006) on the effectiveness of ontology training on physics concepts for 24 undergraduate students. A computer-based module enhances the instruction. The study aims to change the ontological nature of miscategorized concepts through ontology training. Researchers mainly emphasize that ontological commitments form the learners' understanding of fundamental concepts. Therefore attributions of features or properties of concepts should be in the correct ontological category. The study reveals that ontological category change instruction is critical to learning concepts and achieving conceptual change.

Chiu and Lin (2005) had an experimental study to disclose the role of multiple analogies on conceptual change. The study is mainly concerned with the effectiveness of the multiple analogies, which are theoretically based on an ontological perspective to link students' prior understanding of daily life events to the knowledge of the scientific domain. The study results demonstrated that otological category change enables to overcome of misconceptions and promotes scientific understanding.

Çoruhlu and Çepni (2015) investigated a study on the effect of the 5E model enriched with conceptual change pedagogies on a comet, star drift, and meteor concepts. The study focused on the ontological categories of these concepts to eliminate student misconceptions. The conceptual change text was applied to enhance student comprehension and to provide correct futures of new ontological categories. As a result of the study, instruction designed based on the ontological perspective is an effective strategy.

Another study focused on the effectiveness of ontological distinction questions is examined by Erdmann (2001) to improve conceptual change on the photosynthesis concept in a conceptual change text. Researchers ask questions about the ontological attributes of plants and animals in the text to increase learners' metacognitive awareness. The study reveals that the used conceptual change text design based on ontological distinction enables learners to go through the conceptual change concerning photosynthesis.

Although several studies confirm the effectiveness of instructions designed with ontological category shift strategy, contradictory results also exist. For example, Charles (2003) introduces that students trained with ontological frameworks might not understand problem-solving better, and some are devoid of causality. On the other hand, students trained with traditional instruction may have better transfer knowledge. This is why novice learners may not have a proper ontological framework to construct a scientific category. Secondly, conceptual change needs time and gradual change. This evidence suggests that the new emergent categories were not necessarily stable or coherent as an opposing idea of ontological perspective.

Yang et al. (2012) aimed to help students to develop new ontological attributions about challenging concepts such as diffusion, microfluidics, and heat transfer to achieve conceptual change. They proposed that students can more easily achieve scientific understanding after establishing an appropriate ontological category. As a result of experimental study enhanced with computer modeling, students had moderate development on some subjects but not all. For example, there is no significant difference between control and experimental groups in heat transfer concepts. Therefore, there should be a further investigation as there may be other factors that might have contributed to this result.

2.2.4 Why did We Choose Conceptual Change Strategies for Meta-analysis

Conceptual change occupies a vast place in the literature as a growing field over time. As stated above, there are different instructional implications stemming from distinct knowledge perspectives. The findings of these implications also imply varied strategies to achieve conceptual change, and their effectiveness is also pretty scattered within the framework of conceptual change literature. Therefore, it is critical to comprehensively analyze these findings to specify more valid and reliable implications for instructional purposes. Moreover, many moderators significantly affect outcomes, and their effect on treatment intervention is never observed by single studies or narrative reviews. But, how can we talk about the overall effectiveness of conceptual change strategies? Clearly, there needs to be a holistic view of the effectiveness of different conceptual change strategies. In this sense, meta-analysis can provide practical knowledge to present the overall effectiveness of conceptual change strategies in science education. Additionally, the increasing popularity of meta-analysis studies is obvious. But, there is no meta-analysis study on this field yet due to the more complex nature than other educational strategies and methods. Hence, the gap in this field, the growing number of studies, and the increasing divergence of findings make performing a meta-analysis on conceptual change literature significant.

In science education literature, it is clearly noticed that the number of studies related to conceptual change instruction increased rapidly (Figure 2.3). The below data was collected by searching the “conceptual change” phrase on the title between quotations from 1978-2021 years in more than 100 databases reached by METU library opportunities. The studies were published or unpublished journal articles, master thesis or doctorate dissertations, ebooks, papers, reports, magazines, electronic resources, or new in more than 50 languages. The result of this comprehensive search sign in the growing popularity of this field in literature.

(Some of the searched databases**:** Google Scholar, ProQuest, TR Dizin, Scopus, Web of Science, ScienceDirect, ERIC, Education Source, Academic Search Complete, DergiPark, American Doctoral Dissertations, Social Sciences Citation Index, Science Citation Index, Complementary Index, EBSCO e-Classics Collection (EBSCO), EconLit with Full Text, Education Index, ERIC, GreenFILE, SocINDEX with Full Text, Teacher Reference Center, ULAKBİM Ulusal Veri Tabanları (UVT) – ULAKBIM, Turkish National Databases).

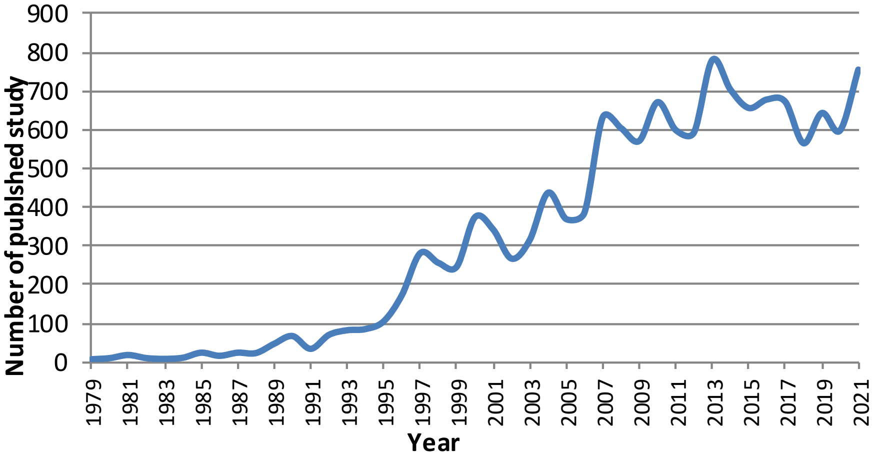


Figure 2.3 The number of studies related to conceptual change for years.

2.3 Meta-Analysis as a Method of Research Synthesis

The cumulative nature of knowledge is important to advancing current scientific knowledge. Today, vast amounts of data have been accumulated in even a single area of science. In fact, the need is not for additional empirical data but for deeply interpreting them. As Glass (1976) pointed out that the results of a vast amount of studies can no more information provide as much as one can grasp knowledge with the aid of organizing, depicting, and interpreting data with reviews. Especially in psychology and education, new studies show conflicting findings. Some of these studies find statistically significant relationships, but some have conflicting findings. Much research literature has proven that this split is approximately 50–50 (Sedlmeier & Gigerenzer, 1989). Hence, it is not so feasible to develop understanding, theories, and cumulative knowledge by looking at empirical studies one by one (Hunter & Schmidt, 1990). On the other hand, research synthesis focuses on contradictory findings from empirical studies to draw a comprehensive conclusion.

As a quantitative research synthesis method, meta-analysis is one of the most common and comprehensive ones that include well-defined statistical procedures. This term was firstly defined by Gene Glass (1976) as referring to the definition that "meta-analysis is the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding" (p. 3). It is evaluated as both a research method and research synthesis in the literature. Cooper and Hedges (2019) define it as research synthesis, but Smith and Glass (1976) define meta-analysis as a research method rather than a review since it has special procedures and steps like a method. Hunter and Schmidt (1990) defined meta-analysis as a quantitative expression and analysis of effect sizes by using descriptive statistics across studies. Onuoha (2007) also addresses the meta-analysis as a set of statistical procedures that uses quantitative primary studies to conclude results. Cooper (2017) also defines meta-analysis as a research synthesis that solely uses quantitative procedures to combine the research results. Glass defines meta-analysis as the more empirical and precise form of research synthesis. Rosenthal and DiMatteo (2001) predominantly address that meta-analysis is more than a quantitative procedure. It is a systematic method of testing a hypothesis by formulating a research question, setting criteria for exclusion/inclusion of studies, developing a coding protocol, synthesizing and combining effect sizes, and defining moderator and mediator variables to evaluate the immense number of studies to perceive the whole picture in a specific research area. As summary, meta-analysis is a quantitative technique that uses effect sizes revealed in the primary studies to provide a deep interpretation of those grasped data. In this sense, it can be thought of as a well-defined method to provide practical knowledge and sufficiently comprehensive information for policymakers by measuring the overall effectiveness of an intervention.

2.3.1 Meta-Analysis as an Effective Synthesis Process

The prominent advantages of meta-analysis make it more widespread among researchers. Literature is full of contradictory findings, even in a single area of science education. In order to conclude a final comprehension, we need a consensus about the research findings. On the other hand, any single study cannot reveal conflicts and differences. Thus it is impossible to conclude common results across studies just by examining primary study findings. But, one of the main purposes of the science education literature is to reach objective findings by revealing differences and contradictions. Therefore, the scientific process should be cumulative. Each study gives evidence, supports or verifies previous results, reduces statistical imprecision for new studies to make a generalization, and reveals the possible conflicts. This makes meta-analysis significant for providing valid results across quantitative statistical results.

The meta-analysis also enables to compare the processes followed in primary studies that cannot be investigated through a single study. In this sense, meta-analyses can also investigate relationships between studies (Arthur et al., 2001). One can easily detect the conflicts between studies by using meta-analytic findings. Therefore, it can be interpreted about the validity and reliability of processes used in studies in science education. In fact, any study is unique in its methodology, subject characteristics, and application procedures. But, contrasting the processes may inform us about the validity and reliability of findings. One of the main ideas of meta-analytic studies is based on this priority.

Although the distribution of the researched areas in the literature is not very homogeneous, some topics are intensely studied, and some have not been investigated. The insufficient or not investigated subjects are called gaps. In this sense, gap analysis is another effective dimension of meta-analysis (Üstün, 2012). Identifying areas that are not studied at all or enough in the literature is difficult. In order to create more operative literature, fields should be criticized both comprehensively and homogenous. Meta-analysis makes it easier to do gap analysis because it provides an overview of the literature by using a large number of studies in a specific field. Moreover, it is also possible through a meta-analysis to calculate how effective and important a new study field will be even before the study is not done. This is important in terms of giving direction to incoming studies.

Borenstein et al. (2009) underline the role of meta-analysis on the consistency of intervention effect. Studies generally report the p-values, which do not give pure information about the intervention effect size. Even though primary studies report the effect size value, they may not be consistent with each other since there are numerous effect size values reported in the literature. On the other hand, traditional reviews have no precise mechanism for assessing the consistency of effects, and this point is neglected in reviews. Since, among review studies, only meta-analysis allows for observing the overall magnitude of significance by operating effect size value, this makes findings more operative and functional for researchers.

From a different perspective, the meta-analysis also has a mechanism to deal with study artifacts. These artifacts may stem from sampling error, measurement error, study designs or extraneous factors. Hunter and Schmit (2004) stated that there is no study without artifacts, and these imperfections may also impact findings. Meta-analysis is crucial in describing the distribution of actual correlations between independent and dependent variables. Although it is not possible to eliminate all artifacts, researchers may describe inconsistencies to handle imperfections more effectively by disclosing actual correlations.

2.3.2 Criticism on Meta-Analysis

Research synthesis has some limitations and disadvantages across other types. Researchers also have voiced some critical issues and are concerned about meta-analysis. Some of these critics were shared by many researchers, but some were ignored. We may list prominent criticals to clarify them more comprehensively.

- One number, effect size, cannot summarize a research field
- Garbage in, garbage out problem
- Mixing apples and oranges problem
- The file drawer problem invalidates meta-analysis
- Sampling bias, essential studies are ignored
- Non-independent findings violate the assumptions
- Sampling error and data error problems
- Problems stemming from statistical issues

Studies (Bailar, 1995; Cooper, 2017; Valentine & Cooper, 2008) questioned whether it makes sense to reduce treatment effect into one number with prediction and confidence intervals may provide actual treatment effect for the whole literature or not. As a descriptive answer to this issue, Borenstein et al. (2009) underlined that meta-analysis does not simply report summary effect but rather integrates numerous findings to interpret treatment effect by using the dispersion of study findings. Therefore, the mean effect value is reported by considering different assumptions to reach a more valid interpretation for a specific area of literature (Cooper, 2017).

Another issue that is described in the literature is the garbage in garbage out issue (Borenstein et al., 2009; Cooper, 2017; Hunt, 1997). Some of the studies have powerful true designs and give the sense that they provide valid and reliable findings with objective judgments. On the other hand, some studies include poor designs with few sample sizes. The persuasiveness power of the reliability of findings for these two studies may not be identical for researchers. Borenstein et al. (2009) stated that researchers might believe that artifacts in the primary studies will be carried over to the meta-analysis which also causes fundamental errors in meta-analysis findings. If the low-quality studies also have a large sample size which causes weighted effect size of these studies becomes more than other true experimental designs. Such a problem is very common and results in challenging conclusions for meta-analysis studies. Cooper (2017) addresses a way to eliminate this problem by including only high-quality studies or moderator variables set to detect the extent to which poorly and well-designed studies differ from each other in terms of effect size measures. Borenstein et al. (2009) focus that inclusion and exclusion criteria are enough for a study to become eligible for meta-analysis.

The third issue is mixing apples and oranges. This implies the improper use of combining different studies with different purposes and implications in the same analysis. As a response to this issue, Rosenthal and Dimatteo (2001) argue that meta-analysis studies try to summarize specific literature by combining any study without regarding their dependent and independent variables or research problems. Cooper (2017) stated that meta-analysis could be too broad or narrow for the investigated research question. The findings can be so different and inconsistent with each other. Nevertheless, combining different study findings is one of the strong dimensions of a meta-analysis by using relevant moderator variables and assessing consistencies across findings. Therefore, combining different studies in the same analysis doesn't pose a problem in investigating the same question.

The fourth problem is publication bias (file drawer problem) which is about the publication of studies reporting mostly significant results or studies with significant results have more tendency to publish in articles. Borensthein et al. (2009) stated that it is more likely to publish studies with high treatment effects than studies finding lower treatment effects. Therefore, it is less possible to reach studies with statistically non-significant results. Rosenthal (1979) called this phenomenon as 'file drawer problem'. For Rosenthal, this is the most noteworthy threat to the validity of meta-analysis studies. This is because publication bias may lead to both the loss of studies and the problem of using biased studies in meta-analysis. Moreover, meta-analysis directly reflects this bias in the computed mean effect. This is also a missing subject problem. The failure to include just published studies may result in less information and less powerful tests (Borenstein et al., 2009). On the other hand, Borenstein et al. (2009) claim that publication bias is not a problem caused by meta-analysis. This is a common issue in the whole literature. As a solution, there is some practical advice put forth by researchers with a meta-analytic perspective. A series of methods have been developed to assess bias's impact and remediate findings. Cooper (2017) also argues that comprehensive search and reasonable assumptions can mitigate the publication bias problem.

About the meta-analysis, Rosenthal and DiMatteo (2001) defined a critical issue that may limit the effectiveness of meta-analysis studies. This issue is defined as sampling bias which implies the insufficient data collecting process. This is mainly related to the scope of sampling. Researchers should limit their study content with inclusion and exclusion criteria (Rosenthal, 1979). According to Cooper (2017), researchers may set eligibility criteria very subjectively, yielding bias in sampling. The garbage in, garbage out problem relates to the inclusion and exclusion criteria, but it doesn't ensure that important studies were left out. Additionally, researchers cannot reach every published or unpublished study by using any source. This makes it impractical and impossible to cover the whole literature for any subject. Therefore, Borenstein et al. (2009) stated that researchers might not be comfortable with a meta-analysis's findings. This leads to excluding lots of related studies. On the other hand, meta-analysis does not have such a claim to cover any study done in that field rather meta-analysis claim also to study with a sample that represents the universe best. Additionally, meta-analysis studies should set eligibility criteria before the study is implemented to grasp sufficiently similar studies (Borenstein et al., 2009). Therefore, this limitation may be neglected from a meta-analytic perspective by performing a well-defined and comprehensive searching process.

Another prominent critical that may influence the validity of meta-analytic findings is the existence of non-independent findings. In the literature, the ‘lumpiness (non-independent data or singularity) and overemphasis of individual effects are also defined as limitations in the meta-analysis (Rosenthal & Dimatteo, 2001). Some primary studies may implement more than one independent variable on the same field of subject and the same sample. For example, the researcher may investigate the effect of cognitive conflict strategy on the subject of Newton’s first, second, and third laws of motion for achievement in the same sample. Three non-independent effect size values should be yielded for the same dependent variable, which causes overemphasis on the sample. Rosenthal and Dimatteo (2001) propose that it is easy to deal with this problem by combining these effects or using moderators to measure this effect. Therefore, non-independent data is not such a serious problem for meta-analysis.

With respect to Schmidt and Le (2004), the main weaknesses of a meta-analysis are sampling error and data error problems. Each of these problems stems from the imperfection in primary studies. Imperfections may be due to computational error, transcriptional error, coding error, or data handling. Meta-analysis studies cannot affect the primary studies' confidence interval, sampling methods or design. It is also exactly unknown where the magnitude of sampling error and data error stems from. It is also impossible to distinguish between sampling and data errors. Therefore, meta-analysis is weak in eliminating the source of errors. In fact, meta-analysis tries to eliminate these problems by determining moderator variables to disclose the possible heterogeneity between findings. In this way, the effect of imperfections can be observed.

Meta-analysis is very useful for obtaining precise effect sizes by combining lots of primary studies. Even though meta-analysis increases statistical power, there are some problems stemming from statistical issues. Firstly, combining studies will inflate type I error since we also combine errors to raise statistical significance (Ueno & Fastrich, 2016). Secondly, unjustified or mis-justified use of fixed- or random-effects models. The two models concern the different inferences and assumptions. Therefore, deciding on the models to yield valid and reliable conclusions is critical. This problem also implies a poorly performed meta-analysis. The third statistical issue arises from the failure to weight effect sizes correctly within and between studies (Cooper et al., 2019). The proper weighting should consider the precision of studies which is inversely proportional to the sampling variance. But, researchers may neglect the weighting while including more than one data from a study. Therefore, this problem rarely occurs in meta-analysis.

2.3.3 Previous Meta-Analyses on Different Teaching Methods

Meta-analysis is a tool to grasp more functional knowledge through comprehensive statistical data from primary studies of the intended area. It is easy to perform on available meta-analysis programs and does not need to design experimental or control groups. Therefore, the meta-analytic perspective is free from problems stemming from studying individual samples, including ethical rights like privacy, personal rights, or moral issues. So that there is no need for additional effort to pay attention to ethical issues and small risk factors rather enable to focus more on findings. For example, studying in the medical field involves debates about experimental studies and common ethical problems. First, researchers must select available samples critically regarding ethical issues and perform experimental steps within the limits of personal rights. Moreover, they should manipulate and create study groups within these limits. In this sense, meta-analysis studies provide great opportunities to integrate existing primary studies and to use a large body of research. Haidich (2010) introduces a hierarchy of evidence for clinical studies. The place of meta-analysis in this hierarchy is significant (Figure 2.4). Evidence-based data is crucial for fields that study individuals like science education. Therefore, this hierarchy signifies the prominence of a meta-analytic perspective on advancing science education literature.

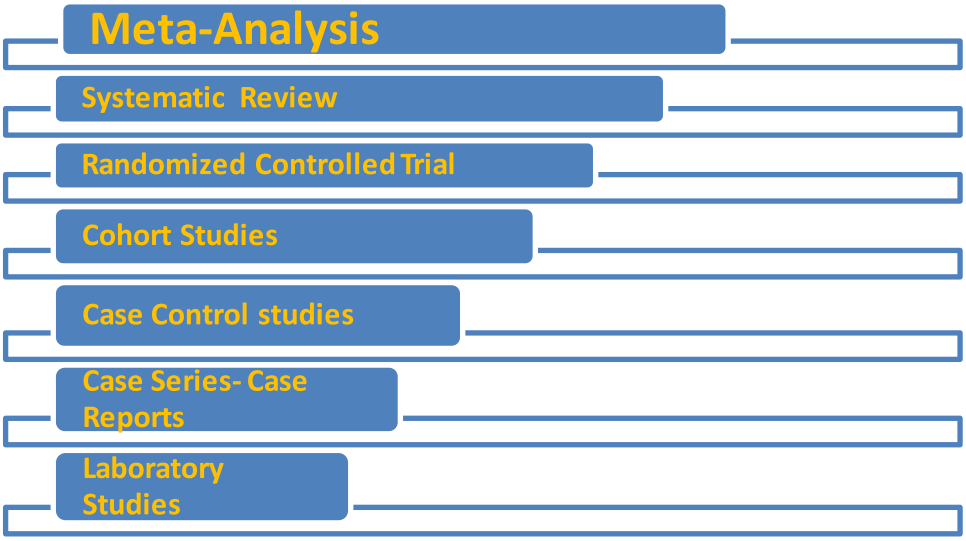


Figure 2.4 Hierarchy of evidence in medicine (Haidich, 2010).

Meta-analysis term was firstly used by Glass (1976) to define the statistical research synthesis. Since 1976, lots of meta-analyses have been published in very different fields of science education. They are mostly office-based, very cheap, and possible to publish in good journals with a high impact factor. Therefore, the number of meta-analyses dramatically increased (more than 10.000 per year) in many fields of science (Tebala, 2015).

In science education, the number of meta-analysis studies comparing different teaching methods is increasing dramatically (Figure 2.5). The increasing number of studies on education is responsible for this acceleration also. When we search titles with "meta-analysis" term in the "Education Source" database for each year it is easy to notice the increasing number of studies on meta-analysis. Below data are collected with searching "meta-analysis" phrase on the title between quotations at 1977-2021 years.

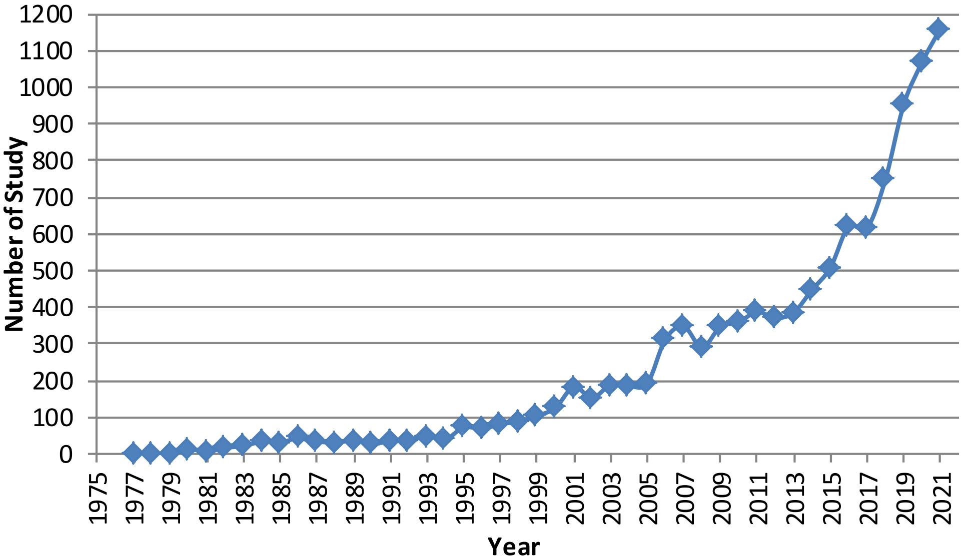


Figure 2.5 The number of published studies for years. "Education Source" database searched for "meta-analysis" phrase

Reviewing the previous meta-analyses comparing the effectiveness of different teaching methods is so informative to grasp more comprehensive, valid, and reliable findings for further studies. Additionally, pre-defined moderator variables that were handled by the previous meta-analysis give hints for further research about the effective moderators on intervention effect. The eligibility criteria for inclusion and exclusion of studies cause serious limitations on relevant sample size. These criteria should be set reasonably so as not to lead to validity and reliability concerns. Therefore, analyzing previous meta-analyses may inform researchers on the critical of new meta-analyses in more detail.

One of the earliest meta-analysis studies on education was done by Glass and Smith (1977) to determine the effect of class size on achievement. The meta-analysis included 80 studies on the relationship between class size and achievement of over 700 classes and over 900.000 students. These numbers show that there were many studies in a single field of education even in that year. As we pay attention to the above graph on acceleration for the number of studies, it can be derived that there is a huge number of studies in the field of science education. The above study revealed a definite inverse relationship between class size and achievement by integrating around 100 study findings on a curve as class-size increases achievement decreases.

The early studies on comparing educational methods were firstly done by Kulik et al. (1979) to compare the effectiveness of innovative instruction which is enhanced by audio-tutorial materials with conventional teaching on achievement. Researchers used 48 published and unpublished studies in the English language. Fourteen moderator variables including intervention length, subject, course level, publication status, and design are coded and investigated with regression analysis to observe the relationship with treatment effect. There are also eligibility criteria like a study had to take place in an actual college course. Secondly, the duration of the study had to be reasonably long--i.e., more than an hour or two of A-T in a one-semester course. Third, the study had to be free from obviously crippling methodological flaws. They investigated three main research questions are;

- How effectively does this innovative method prove to be in the typical comparative study?
- Is it especially effective for certain types of outcomes or certain types of students?
- Under which conditions does it appear to be most effective?

As a result of the analysis, a small overall effect size has been measured by integrating the findings. The Cohen’s d measure was used for standard effect size value during the evaluation process.

After the 1980s, studies on the effectiveness of new instructional tools like computers arise (Clark, 1983; Kulik et al., 1979, 1980; 1983; Wise & Okey, 1983). One of the comprehensive meta-analysis done by Kulik et al. (1980) was on the effectiveness of computer-based instruction versus traditional instruction. The 59 studies using computer-based instruction in the experimental group are located for use in this meta-analysis. Researchers also set 13 additional variables to describe different features of the application process for computers. For example, there are four major types of applications: tutoring, computer-managed teaching, simulation, and programming the computer to solve problems. Some researchers use computer as additionally but some use them during complete instruction. Or researchers maybe use computer both in control and experimental groups but some just use in experimental groups. Therefore, some moderators may affect the treatment effect. Wise and Okey (1983) also did a meta-analysis study on the effectiveness of CAI on achievement with twelve studies. The microcomputers used during instructions and the large effect size value regarding Cohen's d are measured as 0.82. Moderator variables do not code and analyzed during this analysis rather meta-analysts focus on the overall effectiveness of CAI.

After the 1990s, meta-analysis studies became more method oriented in education, especially to investigate instructional methodologies like problem-based method (Christmann et al., 1997; Kulik & Kulik, 1991; Liao, 1999), project-based instruction (Chen & Yang, 2019), concept maps as an instruction tool (Horton et al., 1993), computers assisted instruction (Chadwick, 1997; Flinn & Gravatt, 1995; Lee, 1999). Moreover, increasing the sample size in meta-analysis allows meta-analysts to analyze possible moderators that may influence treatment effect. In this way, more comprehensive knowledge is gathered about variables responsible for the actual treatment effect.

Chadwick (1997) performed a meta-analysis study on CAI as a dissertation in the secondary mathematics classroom. The study includes 41 relevant primary studies as the mean effect size is 0.33 which yields a small effect. This is one of the comprehensive studies that use 46 moderator variables, a well-designed coding book and predefined eligibility criteria.

The study quality is a very critical issue in meta-analysis literature. Some criticisms, like the mixing apples and oranges problem or sampling bias, are directly related to the primary study designs and qualities. Therefore it is accepted as one of the effective moderators but hard to code and analyze very objectively. Therefore, very few studies have been done on study quality. Chadwick (1997) has coded the study quality by making use of answers to the below questions;

Was there a random sample? Was there an equitable control group? Were the independent variables controlled? (i.e., one teacher for all groups) Was the instrument valid?

Overall Study Quality and mean effect size values
0-Poor study (.92)1-Marginal study (.39) 2-Good study (.42) 3-Excellent study (.08)
As a result of this analysis, the researcher shows that as study quality decreases, the treatment effect increase (Figure 2.6). This result is so dramatic that if researchers control more threads, the treatment effect becomes much lower. This result also implies how effective the moderators are except for treatment on achievement.

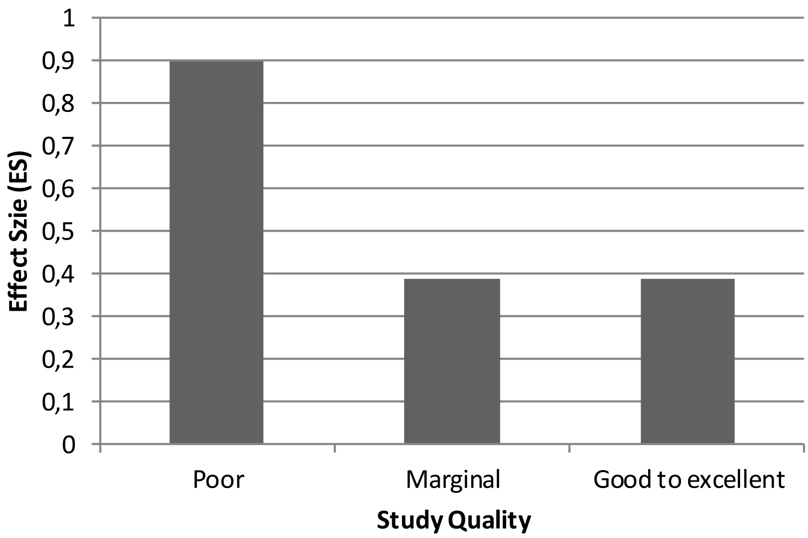


Figure 2.6 Bar graph comparing the effect size and study quality for poor, marginal, good and excellent (Chadwick, 1997)

The researcher also analyzed the time moderator which signifies that recent studies have larger effect size values than early studies, perhaps due to the more appropriate use of instructional technology. This result is also consistent with other meta-analysis findings (Armağan, 2011; Secondly, journals have more effect size value (0.81) than dissertations (0.22) that perhaps due to publication bias which implies that journals have more tendency to publish significant findings. This concern is also shared by many other researchers that conducted other meta-analysis studies (Bayraktar, 2000; Clark, 1983; Kulik et al., 1983).

The intervention length has also been discussed in meta-analyses for problem-based learning (Bayraktar, 2000) and computer-based instruction (Clark, 1983; Kulik et al., 1983). These studies impy that intervention lenght affects the treatment efficacy. Both a concise period of treatment and a very long treatment period have negative impact on efficacy. The result of some meta-analyses (Arık & Yılmaz, 2020; Üstün, 2012) is very identical to the results of Chadwick also (Table 2.1). Clark (1983) and Kulik et al. (1983) also imply that the treatment effect value increases when the intervention length increases. As stated later parts, these results are somehow parallel to this study's findings also.

Table 2.1 The results of a fully random-effects moderator analysis for the length of treatment (Üstün, 2012).

Subgroups	Effect Size and 95% confidence interval						Statistical test	
	Number of Studies	Point Estimate	Standard Error	Variance	Lower Limit	Upper Limit	Z-value	p-value
0-5 weeks	32	0.613	0.099	0.010	0.419	0.807	6.207	0.000
6-10 weeks	26	0.682	0.110	0.012	0.467	0.897	6.228	0.000
Over 10 weeks	19	0.480	0.127	0.016	0.232	0.729	3.792	0.000

The data collection process is so critical to inform about the quality of meta-analysis studies. More comprehensive and systematic reviews enable researchers to reach a more descriptive sample for the population. One of the most recent and systematic meta-analyses on the effect of the constructivist learning approach versus the traditional approach is done by Arık and Yılmaz (2020) in 57 primary studies. The Hedges'*g* is used as an effect size index and the eleven explanatory moderators are coded. The comprehensive data collection process has been done by researchers who report that 72,450 studies using lots of databases for a long time are reviewed. This is an extremely large number to reach a well-descriptive sample for the population. The year interval for sample is between 2000-2015. The moderator analysis reveals that "country", "sample size", "educational level", "type of publication", "type of measuring instrument", "developer of measuring instrument", "language of publication", "teacher effect", and "researcher effect" are effective moderators that change the treatment effect. But researchers do not perform a simultaneous regression analysis to observe the combined effects of moderators. The result of the study implies a very large mean effect (Hedges*'g*= 1.46).

In recent years, inquiry-oriented studies have become more popular, leading to an increase in a meta-analysis in this field. A recent meta-analysis is conducted by Kaçar et al. (2021) on the effectiveness of inquiry-based learning versus traditional instruction on achievement. The study is conducted on just Turkish publications between 2000-2020 years. The exclusion criteria are set as language, year, experimental studies with pre-test and post-test groups, articles or dissertations. There is no master thesis or papers due to study quality concerns of researchers. The random-effects model was used for analysis to estimate the overall mean effect. Hedges's *g* was used as the effects size index for the reason that some sample sizes used were below 20. The overall effect size value was observed as 1.181 for 30 primary studies which are very large effect size.

2.3.4 Previous Meta-Analyses on the Effectiveness of Conceptual Change Strategies

Three related studies that analyze conceptual change strategies in literature are discussed in this part. One of the most common meta-analysis studies on conceptual change was conducted by Guzzetti et al. (1993) on refutational texts, augmented activation activities and discussion web to achieve conceptual change. These strategies aim to eliminate misconceptions by providing dissatisfaction processes and introducing new scientific knowledge to achieve conceptual change. This study is limited to five years of time intervals and low number of studies because conceptual change strategies emerged after Posner et al. (1982) with a conflict perspective. Since the instructional implications were precisely defined after 1985, the number of experimental studies was so limited during these years. Until 1990, there were few methods performed by researchers to overcome misconceptions. This meta-analysis includes 23 related studies between 1986-1990 years. The eligibility criteria are described as studies should include misconception, naïve or preconceptions as a theoretical base, pretest-posttest results, and statistical necessity. Primary studies do not directly state that they follow the conceptual change instruction. Still researcher has inferred whether the strategy is CCI or not if the primary study satisfies conceptual change conditions. Almost all studies were conducted at the USA (%96). The subject area includes earth science, physics, and life science. Since there is not enough study to *analyze* moderators like subject, different strategies of CC or student grade, the researcher had just discussed the main research question about the effectiveness of conceptual change strategies.

Another study on conceptual change strategies is conducted by Armağan (2011) as a dissertation on the effectiveness of conceptual change texts versus traditional instruction. The researcher has reached 42 relevant studies among around 6000 studies after exposing eligibility criteria. The collected studies are limited to time interval 1998-2010, Turkish, pretest-posttest control experiment design and text-based strategy. Therefore, the size of the intended sample to include analysis is limited. Compared with early studies, the main effect size value is also a large effect (d=1.18). Researcher analyzes eleven moderator variables. One moderator variable that is special for this study is "technique used in primary studies". As a result of this moderator analysis, there is no statistical difference between techniques that use conceptual change texts or combining text and another strategy. The effective moderator on treatment effect is observed as just sample size. When the sample size increase, the treatment effect also increases. This is not very identical to other meta-analyses that investigate the effect of sample size (Arık & Yılmaz, 2020).

There was also a master thesis conducted by Gelen (2015), which investigated the effectiveness of conceptual change instructional tools frequently used like analogies, concept networks, conceptual change texts, concept maps, mind maps, and mind caricatures. The theoretical background for conceptual change instruction has based on Posner's cognitive conflict perspective. The tools are also coded as explanatory moderators for the scope of this study. The number of primary studies is 64. Most of the included studies are done in Turkey (%94). The collected studies are limited to the time interval 2002-2014. Pretest-postest control group design is also set as eligibility criteria to provide a standardized effect size value. The researcher used the Cohens'd index to calculate mean effect size. The overall effect size is calculated as 1.13 which is a very large effect. The moderators are instructional tools which are the special moderator for this study, subject domain, education level, intervention length, publication year, publication type, researcher effect, and computer effect. The results imply that country/region, subject domain, experiment design education level, intervention length, publication year, and researcher effect are effective explanatory variables for treatment effect. But, researchers do not conduct multiple regression to observe the simultaneous effect of these moderators on treatment.

The common point for previous meta-analyses is that they mainly consider the Posner et al. (1982) description of cognitive conflict for the scope of CCS. On the other hand, there are serious criticisms that researchers have already performed conceptual change strategies as instructional implications derived from different knowledge perspectives except for the dissatisfaction process (Figure 2.7). This is mainly due to the different epistemological and ontological approaches that make it more complex to clarify conceptual change as an instructional strategy in science education. In this sense, previous studies have small sample sizes and scope limitations to provide a comprehensive evaluation of conceptual change literature. Therefore, there is a need to integrate findings across related studies on the different conceptual change strategies by using a holistic meta-analytic perspective.

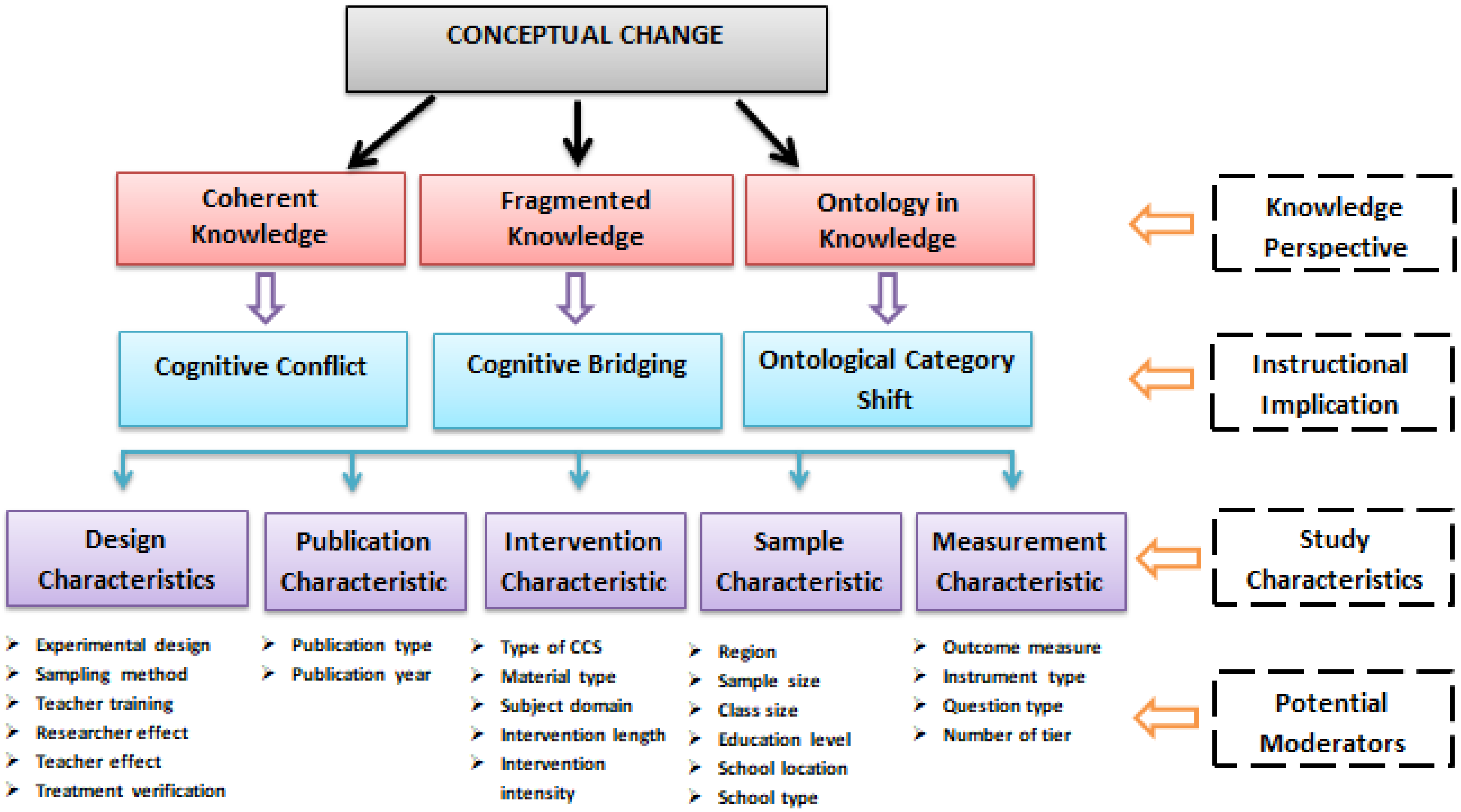


Figure 2.7 Study scheme for conceptual change strategies and related instructional implications

CHAPTER 3

METHODOLOGY

3.1 The Process of This Meta-analysis Study

Meta-analysis is a type of systematic research synthesis that follows a particular set of steps to integrate empirical data (Borenstein et al., 2009; Cooper et al., 2019). The main focus is integrating empirical findings obtained from primary studies in effect size values. Meta-analytic perspective uses statistical inferences to provide a more valid and comprehensive evaluation of previously defined research questions. In this sense, the specific attributions of primary studies were described with moderator variables that were derived from the literature and intended sample. The implementation process may seem so defined at first glance, but it is needed to use extra correction processes for publication bias, sampling and data errors, normalization of effect sizes, and other systematic artifacts. The steps of this meta-analysis study can roughly be stated below;

- Defining the research topic (Addressing the prevalence of conceptual change strategies in literature by performing a pilot search)
- Describing the independent and dependent variables of interest (Describing the conceptual change strategies by using literature focusing on theoretical knowledge)
- Specifying the eligibility criteria for a systematic search (setting the inclusion and exclusion criteria)
- Describing the scope of related databases and indexes (list the related science education databases and indexes, determine the databases that should be searched, and describe the order of searching for the databases)

- Reviewing, selecting, and recording the relevant primary studies by performing a comprehensive systematic search on databases, indexes, and journals
- Determining the final set of studies by using eligibility criteria
- Coding studies (Development of coding sheet and coding manual. Pilot coding for randomly selected studies, code all studies for each moderator and statistical values. Updating the coding sheet and coding manual with respect to pilot coding. Performing intra-rater and interrater reliability processes)
- Describing data analysis procedures and programs (Define the meta-analysis steps like main analysis and moderator analysis. Determine which meta-analysis program is more adequate to resolve research questions.)
- Determining the appropriate model for the analyses (fixed-effect, random-effects or mixed model)
- Investigating the publication bias by means of visual tools (funnel plot, forest plots) and statistical methods (regression tests, fail-safe N methods, trim and fill method)
- Investigating the heterogeneity among obtained effect sizes and further statistical analysis (by means of prediction interval and statistical values like tau-squared, I-squared, and *Q* statistic)
- Calculating the statistical power
- Performing main effect analysis (Investigating the main research questions, e.g., What is the effectiveness of CCS on science achievement when compared to traditional teaching methods?)
- Performing subgroup (simple meta-regression) analysis (Investigating the variation between and within subgroups to examine research questions)
- Performing simultaneous (multiple meta-regression) analysis (Defining the simultaneous effect of moderators on achievement and obtaining the general model)

- The results of the study will be summarized, reported, and conclusions will be drawn based on analyses

3.2 Meta-regression Process

Meta-analyses do not follow standard statistical procedures due to the nature of data and the existence of confounding variables simultaneously. It should probe the existence of relationships among confounding variables and dependent variables to observe the true effect size value (Pigott, 2012). The potential moderators are investigated through meta-regression analysis. These analyses provide to observe the heterogeneity within and between moderators as well. Borenstein et al. (2009) argue that the process is similar to regression analysis except that moderators are at the primary study level rather than the subject's level, and the dependent variable is effect estimate rather than sample data. Therefore, regression and meta-regression follow similar procedures during the analysis. The dependent variable is any effect size value (raw mean difference, standardized mean differences, risk ratio, odds ratio, frequencies, or correlations) and moderator variables are characteristics of primary studies that may affect treatment effect. Turner and Higgins (2019) define meta-regression as an extension of analog ANOVA analysis to investigate the properties of categorical or continuous variables by working with sets of covariates as single or simultaneously. In this sense, simultaneously investigating moderators' effects is also called multiple meta-regression, enabling the presence of the integrated effect of the moderator variables. It also allows controlling confounding variables to observe moderators' individual and simultaneous effects.

According to Borenstein et al. (2009), a meta-regression approach should satisfy two conditions before analysis. Firstly each subgroup should include enough primary studies. Secondly, the correlation between moderators should be reasonable that is to say, it should be free from multicollinearity. Later, the appropriate model required for weighting is determined (fixed-effect, random-effects, or mixed).

We performed meta-regression analysis in three stages. Firstly, we put into practice the main effect analysis, which is the investigation of the overall effectiveness of treatment on dependent variables. In this process, all studies were included in the analysis and synthesized according to the determined weighting model. In the second stage, the moderator variables that influence the treatment effect were investigated. These moderators were defined with respect to the characteristics of the included primary studies, which are also called subgroups. During the analyses, we examined whether there was any significant heterogeneity between subgroups in terms of effect size values. These analyses provided information about the reasons for the total variance between studies and the effect of each moderator on the treatment effect. Finally, we practiced multiple meta-regression analyses to investigate the correlation between moderators and treatment effects. In literature, as stated Table 3.1, multiple meta-regression is performed in different ways as simultaneously, sequentially, or stepwise methods to control confounding variables (Keith, 2019).

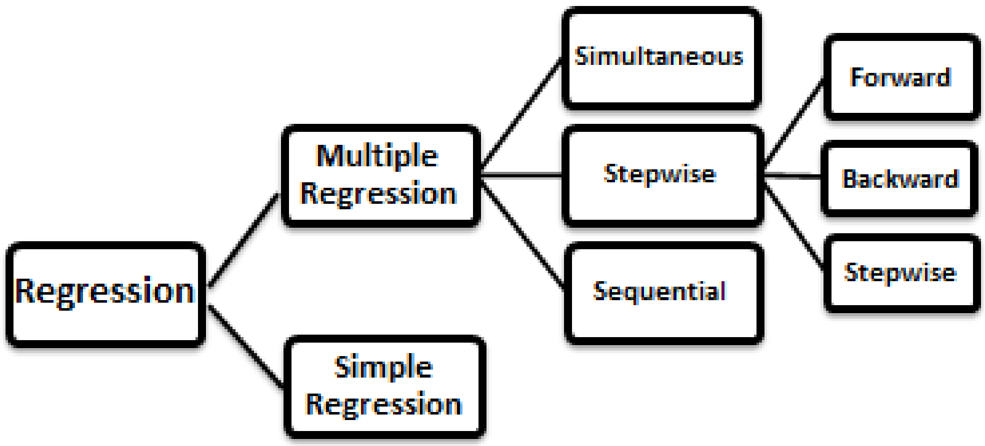


Figure 3.1 The regression analyses types

Simple Regression:

First of all, we assume the relationships that are of interest are linear because we use linear regression models in this meta-analysis study. The regression process that includes one independent and one dependent variable is called simple linear regression (Keith, 2019). The researcher tries to obtain a prediction function between two variables during this analysis. There is a classical data set below about the correlation between the students' high school GPA and their university entrance exam scores (Table 3.1).

Table 3.1 Students' high school GPA and university entrance exam scores

GPA	65	70	75	80	85	90	95
Exam Score	275	318	345	372	400	417	455

The regression equation for the data of GPA and exam score can be described as

$$Exam\ Score = 15 + 5.5\ GPA$$

As an example of educational methods, investigating the influence of subject on the effect of conceptual change instruction on science achievement includes one dependent (achievement score) and one independent variable (subject). Researcher may set a regression equation to predict the actual effect (Y) as;

$Y = a + bX$ which says that a person's score on the dependent variable (Science Achievement) is a result of a constant (a), plus a coefficient (b) times his or her value on the independent variable (subject). If we remove the effect of a subject, then the predicted value of Y will equal to the constant score a; Y=a. Therefore, researchers may investigate the influence of the subject on CCI for science achievement.

Multiple Regression:

It is extremely important to examine the correlation between the moderators, which is the strength of the meta-analysis. If there are more than one independent variable and one dependent variable, multiple regression analysis is required to investigate the effect of each moderator on the dependent variable at the same time (Yan & Su, 2009). Let’s return to the example that was used in simple regression. The researcher may want to investigate the influence of the subject on conceptual change instruction for achievement. On the other hand, sample size and intervention length can be effective moderators on the effectiveness of CCI. Therefore, it may not be possible to isolate the effects of other variables during analysis; rather the researcher should consider the intercorrelations among the four variables, which is also called multiple regression. Sample size and intervention length are also correlated with CCI and subject.

Researcher may set a regression equation to predict the predicted effect (Y) as;

$Y= a + bX_1+cX_2+dX_3$ where a is constant, b is the subject coefficient for CCI, c is the sample size coefficient for the effectiveness of CCI and d is the intervention length coefficient for CCI. Therefore, the researcher may investigate the influence of subject, sample size, and intervention length on the effectiveness of CCI for science achievement.

Similarly, the version of multiple regression on the meta-regression process is called multiple meta-regression. This analysis provides more comprehensive information on the total variance between and within studies. In this respect, the treatment effect is investigated more comprehensively employing multiple meta-regression analyses. Three major types of multiple regression have different purposes for performing, different interpretations, and different strengths and weaknesses (Keith, 2019).

Cleophas and Zwinderman (2017) stated that there are three main goals of multiple meta-regression. Firstly, it enables us to search for possible moderators associated with the treatment effect, which are also called explanatory variables. Higgins et al. (2019) argued that these variables are characteristics of primary studies that may influence the strength of the treatment effect. Secondly, in social sciences, it is accepted that moderators are not isolated. They have interactions among themselves as well as treatment effects. Therefore it is critical to take into account the interactions between moderators. That is to say, we may control the impact of each moderator one at a time to observe the individual influences of variables by using meta-regression. Finally, since multiple analog ANOVA analyses do not consider the interaction effect, analyzing the combined effect of moderators (subgroups) on the treatment effect is impossible. Therefore, meta-regression is a substantial analysis to investigate the research questions while working with multiple explanatory moderators.

Simultaneous Multiple Regression:

In simultaneous multiple regression, all independent variables enter the regression model at the same time (Keith, 2019). This analysis measures the combined effect of independent variables on a dependent variable (Figure 3.2). The primary aim is to determine the extent of the effect of independent variables on a dependent variable (Yan & Su, 2009). Additionally, this analysis gives information on each moderator's relative effects by controlling the other variables. This is so functional to analyze the strength of each variable on treatment effect. During the meta-analysis, researchers should investigate previously determined moderators that reflect the intended population's attributions. Therefore, simultaneous meta-regression is prominent for meta-analysts. During the analysis, meta-analysis programs provide standardized coefficients to predict the relative importance of each explanatory variable. During meta-analysis, the best explanatory models are determined by researchers to estimate the effects of the independent on the dependent variables theoretically before the analysis has been run. Sometimes, it may not be practical or possible to perform research by using a number of variables at the same time. For example, researchers may not collect data from different countries to observe the effect of different nationalities or regions on a treatment effect. In this sense, meta-analysts may easily gather primary studies to obtain and analyze data from different parts of the world. On the other hand, the main weakness of simultaneous regression is that new variables drastically change the coefficients for variables. Therefore, it may not be possible to obtain certain effects for each variable. Researchers should analyze each possible explanatory variable to increase the explained variance (R^2).

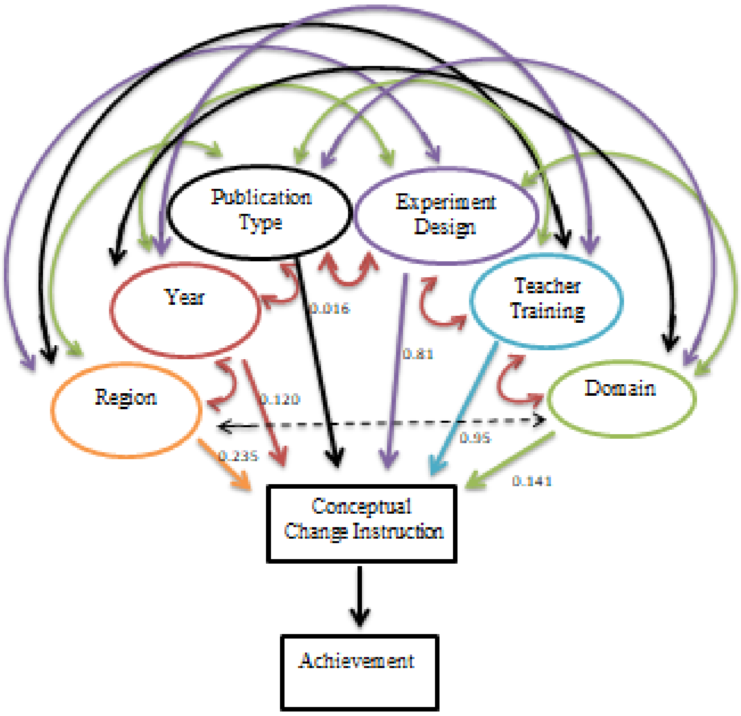


Figure 3.2 Path representation of the simultaneous regression on the effectiveness of CCI.

Sequential Multiple Regression:

The variables are entered into the regression equation one at a time (Keith, 2019). The order of entrance for variables is important and determined in advance by the researcher. The statistical significance and the magnitude of effect on the treatment effect of the independent variables depend on their order of entry into the analysis. This model aims to compare the different models that include different orders of entering variables into analysis to understand the proper order of entry. Therefore, a sequential model helps us to understand the strength of variables in the model individually.

In contrast, the simultaneous model does not provide this knowledge for researchers. For example, the researcher wants to investigate the effect of sample size, intervention length, study quality, and subject on the effect of conceptual change strategies on achievement. The simultaneous model informs us about the combined effect of four variables.

Stepwise Multiple Regression:

The variables are entered or eliminated from the model concerning their correlation to the dependent variable and significance. This is similar to the sequential model except that the stepwise model can be forward (firstly adding the most correlated variable with dependent), backward (non-significant variables eliminated) or stepwise search (changing variables through each step) by the software (Yan & Su, 2009). Keith (2019) argues that this stepwise regression does not give explanatory findings on variables. It may provide knowledge about which variables you should keep and exclude from the model. In this sense, researchers should decide on variables by using knowledge of theory, literature or investigate effective variables which they decide to add to the final model. Therefore, stepwise methods can be used to reduce the number of variables to perform simultaneous and sequential regression.

3.3 Fixed-Effect and Random-Effects Models

There are two broad statistical models to estimate the summary effect size in meta-analysis studies: fixed-effect and random-effects models. Both of them aim to summarize the quantitative results for a number of studies by weight with respect to the precision of studies. Both models assign more weight to more precise studies while computing the summary effect. The main assumption of the fixed-effect model is that it accepts a single population for all studies. The fixed-effect model's purpose is to predict this population's mean. Therefore, there is one true effect size for each sample, which is the same as the population mean and the variance between the observed effect and the true effect for sample only results from sampling error. If there is no sampling error, as shown in Figure 3.3, the observed effect for samples and the mean effect for the population should be same.

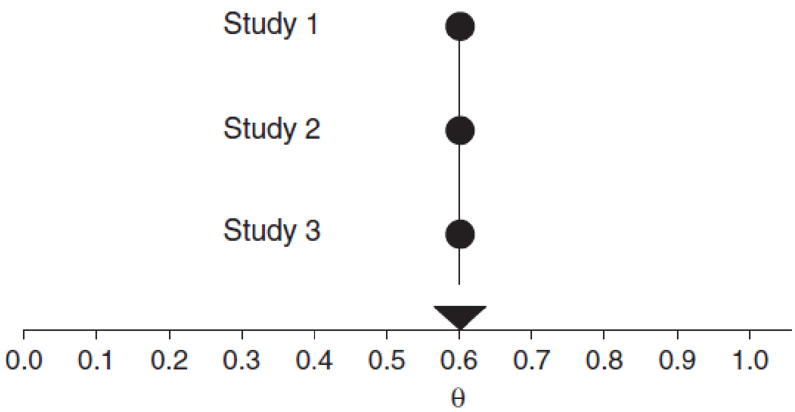


Figure 3.3 Fixed effect model: True effect (Borenstein et al., 2009, p.64)

If there is a sampling error, the observed effect size is distributed around the sample's true effect size (Figure 3.4). In this sense, the weighted mean for observed effects sizes, which is inversely proportional to variance, provides the mean effect size for the population. Therefore, variation of effect sizes stemming from sampling error is called within-study variance. This variation also informs us heterogeneity between primary study findings to evaluate observed effect sizes in the fixed-effect model.

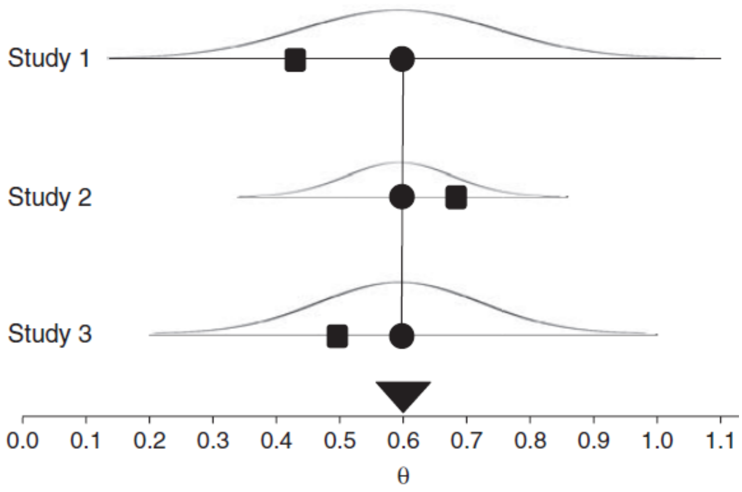


Figure 3.4 Distribution of effect sizes due to sampling error in the fixed-effect model (Borenstein et al., 2009, p.65)

But in the random-effects model, true effect sizes can vary from study to study (Hedges & Vevea, 1998). Each study may have a different true effect size since they represent different samples in the population (Figure 3.5). These different samples are the results of confounding variables. For example, aptitude level can affect student achievement scores on physics subjects in the same grade level. They represent a different sample in the same population, and they have different true effect sizes. Therefore, the observed effect size can change due to both sampling error and sample characteristics. Different moderators, such as aptitude level, sample size, age, gender, subject, and design, result in a distribution of mean effect sizes in the population. The main goal is to estimate the overall effect of distributions of mean effect sizes between samples. In the random-effects model, variation stems from both within-study variance and between-study variance.

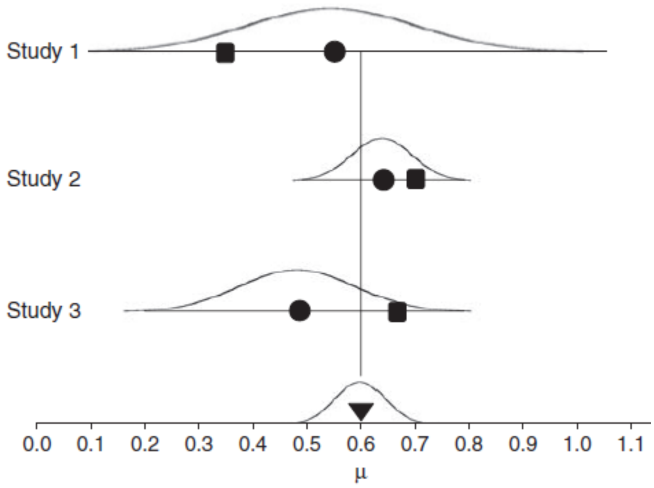


Figure 3.5 Distribution of sampling error in the random-effects model (Borenstein et al., 2009, p.72)

Borenstein et al. (2009) argue that there are two conditions for preferring the fixed-effect model. Firstly, studies included in the analysis should be functionally identical. That is, there should not be a between-study variance for different studies. Secondly, we should find a general effect size for the identified sample rather than generalize for the population. If we try to use a fixed-effect model to generalize findings to the population, we should accept that there is no effect of moderator variables. Nevertheless, it is impractical to set isolated independent-dependent relations without moderator variables. On the other hand, the fixed-effect model is common in literature because it is more simple to apply and not very complex to evaluate (Cooper, 1997; Hunter & Schmidt, 2004).

If unconditional inferences (i.e., inferences about population) are critical, the random-effects model should be used (Borenstein et al., 2009). It is more common in science education that there is uncertainty for attributions of population, and it is inevitable to infer population characteristics from the sample. Therefore the random-effects model is more functional and feasible due to its more critical view of true effect sizes. If rich data are available, then the random-effects model allows researchers to address a wider range of research questions. Therefore, the random-effects model is more recommended for meta-analysis in education (Borenstein et al., 2009; Hunter & Schmidt, 2004).

3.4 Validity Issues in Meta-Analysis

There are two broad critical issues on validity concerns of the meta-analytic perspective: publication bias and quality of studies. Publication bias results in missing studies, causing a biased sample for meta-analysts (Pigott, 2012). Quality of studies also causes discussions on meta-analyses as a source of validity concerns (Borenstein et al., 2009). Meta-analysis does not work on individual samples, rather it is interested in primary study findings. Therefore, it can be discussed that meta-analysis also reflects the errors of primary studies in statistically integrated quantitative results. Borenstein et al. (2009) stated that if the relevant studies use biased samples, meta-analysis findings also reflect this bias. In meta-analysis literature, there are parallel concerns stemming from the quality of primary studies also. In this part, these two validity issues were discussed.

3.4.1 Publication Bias

There are prominent criticisms that meta-analyses may cover a biased sample of all existing studies (Hunter & Schmith, 2004; Lipsey & Wilson, 2001; Pigott, 2012; Rosenthal, 1979). Firstly, this idea was voiced by the Journal of Abnormal Social Psychology editor in 1956, who stated that studies reporting negative findings were less likely to be published in his journal (Thornton & Lee, 1999). Long after this idea was introduced, publication bias was firstly identified by Rosenthal (1979) to inform meta-analysts about this bias, which may lead to missing studies and biased overall effect size in meta-analyses. Several recent meta-analyses also show that published studies have a larger overall effect size value than unpublished ones (Arık & Yılmaz, 2020; Karakuş & Öztürk, 2016). Borenstein et al. (2009) argue that this is due to the fact that studies reporting high effect sizes are more likely to get published than studies that report smaller effect sizes. Moreover, journals tend to reject studies with negative or non-significant results (Üstün, 2012). Therefore, studies reporting significant results are included more likely in meta-analysis. This idea implies that the results of meta-analyses can be inflated due to publication bias. Therefore, researchers should be aware of the possible strength of publication bias affecting their findings (Rendina-Gobioff, 2006). There are some arguments investigating the characteristics of unpublished studies and publication bias by Rendina-Gobioff (2006), Hunter and Schmith (2004), and Rosenthal (1979). First, there is a statistically significant difference in meta-analyses between published and unpublished study findings for mean effect sizes. They also claim that studies with small sample sizes are not frequently included in a meta-analysis. Rendina-Gobioff (2006) summarizes this tendency in her dissertation, as shown in Table 3.2.

Table 3.2 The impact of the relationship between the variance and effect size observed in a study for publication bias (Rendina-Gobioff, 2006)

		Effect Size	
		Small	Large
Variance	Small (N=large)	Published (Statistical Significance)	Published (Statistical Significance)
	Large (N=small)	Not Published (No Statistical Significance)	Published (Statistical Significance)

This tendency results in publishing small studies only if they have large effect sizes since the ones with small effect sizes do not reveal statistically significant results. In the literature, this phenomenon is called “small-study effect” (Borenstein et al., 2009; Sterne & Harbord, 2004). This effect implies the increasing chance of small studies with statistically significant results to publish in articles. In this sense, the weight of these studies in meta-analyses also increases. Borenstein et al. (2009) also state that this increases the weight of more studies showing statistically significant effect size values. Therefore, the influence of a biased sample becomes more striking due to publication bias and significantly distorts findings to estimate the effect under investigation. Today, this issue is accepted as a threat to the validity of a meta-analysis.

Hunter and Schmith (2004) argue that other factors affect the availability of studies like biased sampling and methodological imperfections in primary studies. These sources of bias have also aggravated this issue. But these imperfections cannot be observed clearly in meta-analytic findings and they are not systematic failures like publication bias. Borenstein et al. (2009) state that there are many sources for sampling biases, but they are not so systematic in influencing overall effect size. Therefore, they call this problem availability bias to reflect the other source of biases on this issue. The criticism can be described as the problem of the mis-representativeness of the sample due to some reasonable source of errors. These arguments make this issue more critical to take into account for meta-analysts. Therefore some statistical methods and instruments had developed to detect, correct for, and prevent publication bias more effectively.

Using Borenstein et al. (2009) and Thornton and Lee (1999) studies, we may precisely review the detecting methods for publication bias under three headings as a proportion of significant studies, plots, and statistical methods. The most simple way to observe the distortion of study findings is by comparing the significant and nonsignificant studies as proportion. But this method does not inform researchers actually to demonstrate the possible bias. Two common visual tools are used widely in meta-analysis to observe plots' distortion funnel plots and forest plots. They use visual symmetry on plots; therefore, they are also open to interpretation. The statistical methods can be listed as Egger's method, Rank correlation test, Begg's method, Truncated sampling, Weighted distribution theory, Maximum likelihood, Hackshaw's method, Sugita's method, Givens' method, and Fail-safe N methods (Rosenthal, 1979; Orwin,1983).

There is practical advice of literature to mitigate the publication bias for meta-analysis. For example, conducting a prospective meta-analysis to predict the unpublished study characteristics and following editorial policy to standardize the quality of studies (Thornton & Lee, 1999). But the most effective and functional way is a systematic and comprehensive review of the literature, including published and unpublished studies. This part will explain some of the methods to deal with publication bias.

3.4.1.1 Forest Plots

Forest plots are visual tools to investigate publication bias by depicting the combined effect of point estimates bounded in confidence intervals on a plot (Borenstein, 2005). Researchers may display the distributions of relative effect sizes, confidence intervals and overall effect on a single plot. Figure 3.6 depicts an example of a forest plot showing Hedges' *g* with 95% confidence intervals for 20 studies investigating the effect of conceptual change instruction on achievement. The location of dots on the horizontal axis represents the effect size of each primary study. The size of each dot is the sample size of the primary study, representing the weight in the analysis. In addition, the length of fine lines passing over dots represents the 95% confidence interval. The shorter line represents less variation and more precise study findings. The diamond below gives information about both the overall mean effect size and the confidence interval for the meta-analysis sample with the help of the width of a diamond.

Researchers may visually infer the variance for each primary study, confidence interval distribution, and overall effect size. At the same time, the relation between sample size and effect size can be observed by arranging findings on the plot with respect to precision (Borenstein et al., 2009). These visual tools allow readers to assess the characteristics of the findings across studies and their summary effect (Borenstein & Hedges, 2019). Therefore it is functional to address the forest plots to infer the publication bias at the beginning of the analyses during meta-analysis (Borenstein, 2005). On the other hand, just visual investigation of a plot without supporting statistical evidence may cause researchers to make subjective judgments

(Pigott, 2019). Therefore, different perspectives on detecting bias should be considered by researchers.

Study name	Statistics for each study			
	Hedges's g	Standard error	Z-Value	p-Value
Stavy, 1991	1.214	0.274	4.431	0.000
Woloshyn al,1992	0.756	0.264	2.864	0.004
Kucuk al, 2015	0.646	0.298	2.168	0.030
Yang al, 2012	0.373	0.458	0.814	0.415
Slotta al,2006	0.899	0.362	2.483	0.013
Launey, 1995	1.278	0.301	4.246	0.000
Chang, 1995	0.755	0.164	4.604	0.000
Jensen al,1994	0.546	0.157	3.478	0.001
Gokhale, 1996	0.459	0.294	1.561	0.118
Windschitl al,1995	0.042	0.132	0.318	0.750
Saigo, 1999	0.190	0.214	0.888	0.375
Seyedmonir, 2000	0.030	0.371	0.081	0.936
Sanger al.,2000	0.934	0.246	3.797	0.000
Diakidoy al.,2001	1.383	0.300	4.610	0.000
Erdmann, 2000	0.233	0.142	1.641	0.101
Sungur al.,2001	0.953	0.298	3.198	0.001
Cakir, 2002	0.734	0.224	3.277	0.001
Eryilmaz, 2002	0.169	0.174	0.971	0.331
Uzuntiyaki, 2001	2.457	0.405	6.067	0.000
Tekkaya, 2001	0.586	0.304	1.928	0.054
	0.540	0.050	10.745	0.000

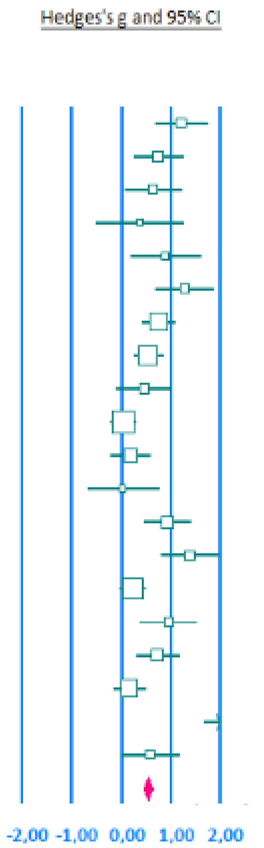


Figure 3.6 An example of a forest plot showing Hedges'*g* with 95% confidence intervals for 20 studies investigating the effect of CCS on achievement.

3.4.1.2 Funnel Plots

Funnel plots are another visual tool to identify publication bias by displaying the relationship between sample size and effect size. This method was firstly introduced by Draper et al. (1987) to detect availability bias. The effect size values are scattered, creating the shape of an inverted funnel diagram, implying that the low precision studies are scattered more than high precision studies. The effect size is reported in the x-axis, and sample size, variance, or study weight is displayed on the y-axis. If there is no publication bias, the plot resembles that the top of the shape associated with studies with high precision, including high sample size and low variance, the wider base of the funnel associated with studies of small sample sizes, small precision and large variances (Pigott, 2019). Therefore, the shape of the plot enables to interpret the possible publication bias by focusing asymmetry on plot. We can observe the symmetric funnel plot in Figure 3.7 and the asymmetric funnel plot in Figure 3.8.

On the other hand, there are concerns about the validity of this method. Hunter and Schmith (2004) argue that the application of funnel plot does not necessarily inform about the publication bias or vice versa because this method does not give reliable evidence for bias. Readers or researchers cannot deduce the number of missing studies or their strength in summary effect on plot (Choi & Lam, 2015). There can be unpredictable factors that lead to asymmetry on the plot except for publication bias, and this method does not necessarily inform researchers about other factors (Egger et al., 1997; Sterne & Harbord, 2004). Therefore, there should be other methods to detect publication bias more precisely. In this sense, there are correcting or preventing methods for publication bias in the literature.

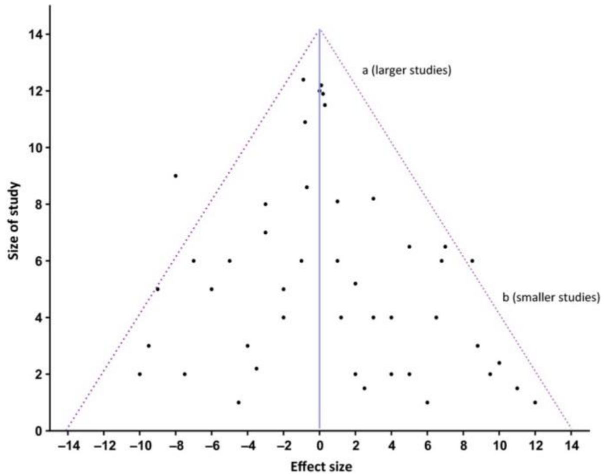


Figure 3.7 Symmetric funnel plot (Choi & Lam, 2015, p.338)

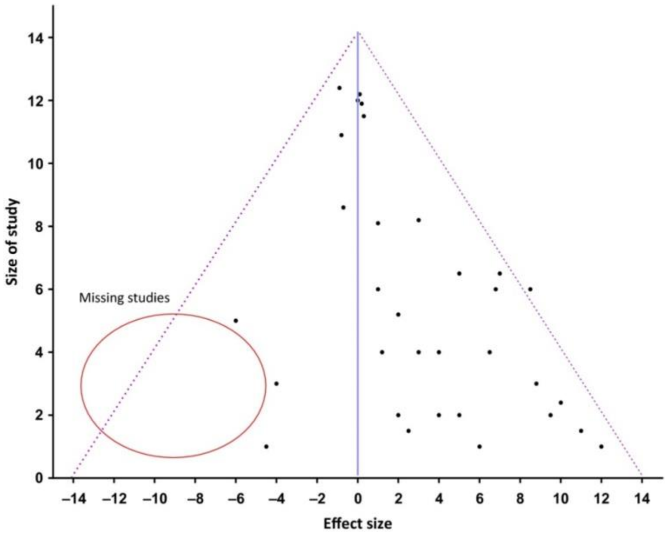


Figure 3.8 Asymmetric funnel plot (Choi & Lam, 2015, p.339)

3.4.1.3 Egger's Linear Regression Method

Insufficient evidence derived from funnel and forest plots makes it essential to develop statistical analysis. Egger et al. (1997) introduced a simple linear regression model to estimate the asymmetry in funnel plots by testing a statistical hypothesis of whether the intercept significantly differs (at $p < 0.1$) from zero. The effect size value for the primary study divided by its standard error gives the value y_i and the inverse of the standard error gives the x_i value. The regression model tests the regression of y_i on x_i value and looks at whether the intercept of these two values is different from zero. The regression equation can be expressed as;

$$\frac{Ti}{\sqrt{vi}} = \beta 0 + \beta 1 \frac{1}{\sqrt{vi}}$$

where Ti is the effect size for study i, and vi is the standard error for the effect size in study i. If there is no publication bias, the $\beta 0$ value becomes zero and yields Ti =1, indicating that the regression line goes through the origin (Pigott, 2019).

Researchers set the null hypothesis that "*there is no asymmetry in the funnel plot*". If the tested value for p< .05 is rejected, as stated below (Table 3.3, p= 0.23), we fail to reject null hypothesis. Thus, there is no asymmetry for the funnel plot that indicates nonsignificant publication bias. This analysis gives more valid evidence for researchers when used with visual detection tools like forest plots and funnel plots also.

Table 3.3 Egger's regression test results.

Intercept	1.99
Standard error	1.62
95% lower limit (2-tailed)	-1.36
95% lower limit (2-tailed)	5.33
t value	1.22
df	25
p value (2-tailed)	0.23

3.4.1.4 Rosenthal's Fail-Safe N

One of the most commonly used fail-safe N (FSN) methods suggested by Rosenthal (1979) is to observe the publication bias most easily and clearly by researchers. This method was derived from the Stouffer method by summing individual Z-scores and dividing by the square root of the number of scores (Orwin, 1983). This method aims to calculate the number of additional studies yielding average null results that make the null hypothesis reduce the combined significance to the desired level (usually 0.05). Although there are no firm guidelines for the critical FSN, Mullen et al. (2001) proposed that if the ratio of $N/(5k+10)$ (where k is the number of individual studies in the meta-analysis) exceeds 1, the effect of publication bias assumed to be ignored. Rosenthal gives an illustration to clarify the method as follows:

94 experiments examining the effects of interpersonal self-fulfilling prophecies were summarized (Rosenthal, 1979). The mean Z of these studies was 1.014, k was 94. How many new filed, or not retrieved studies *(X)* would be required to bring this very large Z down to a barely significant level (p < .05)?

$X = (94/2.706)\ [94(1.014)^2] - 2.706] = 3263$

To make the treatment effect trivial, one should find 3263 more studies with Z value 0 (Rosenthal, 1979, p. 640).

CMA directly reports the FSN results as shown in Table 3.4.

Table 3.4 Rosenthal's FSN for all studies included in meta-analysis

Z-value for observed studies	57.94
p-value for observed studies	0,000
Alpha	0.05
Tails	2.00
Z for alpha	1,96
Number of observed studies	204
Fail safe N	8055

There are two concerns about Rosenthal's approach regarding the validity of the method to detect publication bias. Vevea et al. (2019) address that the most critical point of Rosenthal FSN is that there is no clear-cut and justifiable criterion for fail-safe *N* value. This makes the method nonfunctional for researchers. Secondly, the *z*-scores do not directly account for the sample sizes or precision of the studies, which means fewer studies might be required to overturn the meta-analytic results. In the light of these criticisms, some other methods put forth to detect the publication bias by Orwin (1983) and Gleser and Olkin (1996).

3.4.1.5 Orwin's Fail-Safe N

The applicability of the Rosenthal FSN method is limited to probability levels (usually, 0.01, or 0.05) (Orwin, 1983) because Rosenthal's approach was based on finding a significant p-value (Borenstein, 2005). On the other hand, p-value does not so clear for researchers to evaluate the practical significance of the treatment effect. This assumption may be violated if researchers use different p-values (result may be significant for 0.05 but not for 0.01). Therefore, Orwin (1983) proposed the effect size value to check the null hypothesis by controlling the number of effect sizes that need to reduce an observed overall mean effect size to a particular criterion level.

There are two definite distinctions between Orwin's approach. Firstly, there is no need to define the critical effect size value to zero value rather researchers may set this value particularly. Orwin proposes flexible criteria for cut-off point of effect size value to mitigate cut-off problem. Secondly, Orwin uses the standardized effect size value to set the critical number for missing studies. CMA directly gives the FSN result as below output (Table 3.5).

Table 3.5 Orwin's FSN for studies included in meta-analysis

Hedges'g in observed studies	0.98
Criterian for trivial Hedges' g	0.1
Mean Hedges' g in missing studies	0.00
Z for alpha	1.96
Number of observed studies	204
Fail-safe N	1790

This output (Table 3.5) implies that there should be at least 1790 more studies with mean overall effect size zero to decrease the observed mean effect equal or below the 0.1. The observed effect size is 0.98 and the number of studies included in this analysis is 204. The cut-off effect size value for trivial Hedges'*g* is hypothetically set as 0.1 by the researcher.

3.4.1.6 Duval and Tweedie's Trim and Fill Method

It is common for researchers to work with biased samples during meta-analysis studies. There are several methods to determine the publication bias in literature, but these methods have notable limitations and do not propose adjustment processes for biased samples. Therefore, it is worthwhile to adjust the observed findings hypothetically to estimate an adjusted overall effect size value. The existing methods include complex mathematics and procedures to estimate the missing studies (Pigott, 2019). In this sense, Duval and Tweedie (2000a, 2000b) present a new method that is computationally easy to perform and practical for estimating the missing findings that result from publication bias (Duval & Tweedie, 2000a; 2000b). This method uses the asymmetric funnel diagram to estimate the missing studies by yielding a final symmetric funnel plot. They use an iterative and simple algorithm.

Duval and Tweedie (2000b) summarize the process below;

> "We trim off the asymmetric outlying part of the funnel after estimating how many studies are in the asymmetric part. We then use the symmetric remainder to estimate the true center of the funnel and then replace the trimmed studies and their missing counterparts around the center. The final estimate of the true mean, and also its variance, are then based on the filled funnel plot" (p. 456-457).

The basic aim of this method is not to provide an "adjusted effect size" rather it intends to interpret the degree of publication bias and inform researchers about the meta-analyses that need to be evaluated much more carefully (Duval & Tweedie, 2000b). Here is an example of the trim and fill method (Figure 3.9). The filled dots (right of the plot) represent the theoretical missing effect sizes that were added by method to ensure symmetry. We may see the observed mean effect with a blank diamond and the adjusted mean effect with a filled diamond. The funnel plot slightly shifted to the right due to the theoretical missing effect size values.

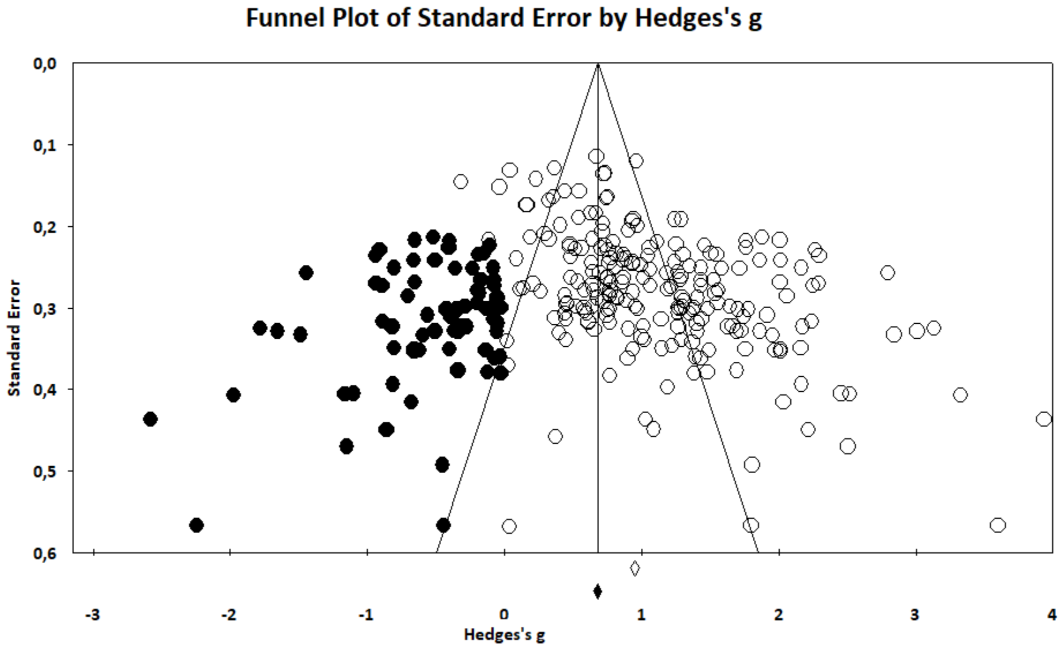


Figure 3.9 The funnel plot of adjusted mean effect due to theoretically missing studies for the random-effects model

3.4.1.7 PET and PEESE Models for Publication Bias

One of the effective ways of detecting and adjusting publication bias is PET-PEESE (Precision Effect Test & Precision Effect Estimate with Standard Error) models that we stated the output at Table 3.6. The PET model estimates publication bias and PEESE enables the report of adjusted measurement in JASP or R programs. Statistically, the PEESE model tries to correct the correlation between effect size and standard errors. The background of this model infers that effect sizes and standard errors should yield uncorrelated findings if there is no publication bias (Bartos et al., 2021). In that, if there is publication bias, studies with low standard errors are more likely to be published. Therefore, the degree of correlation between standard error and effect size gives evidence of publication bias.

The limitations of this model are that it cannot quantify the number of missing studies and is strictly related to the p-values, which just provide the existence or absence of the publication bias. Secondly, it is highly sensitive to sample size (Stanley, 2017). Moreover, results may not provide valid findings in high heterogeneity. Therefore, this model is generally used in simulations and hypothetic findings with large sample sizes (Carter et al., 2019).

Table 3.6 The output of PET- PEESE model results

	Estimate	SE	*t*	*df*	*p*	95% Confidence Interval	
						Lower	Upper
PET	-0.05	0.14	-0.38	216	0.705	-0.32	0.22
PEESE	0.48	0.08	6.24	216	< .001	0.33	0.63

SE: Standard error

3.4.1.8 Selection Models in JASP

This is a correction method by using the likelihood measurement of missing studies due to publication bias (Bartos et al., 2021). It provides an effective estimation process by accounting for heterogeneity within effect sizes in different weight functions like fixed effect, random effects models or mixed model estimates (Figure 3.10). A researcher may also set different sides as one or two tail selections with different p-values. Bartos et al. (2021) identify the selection models as “selection models typically use maximum likelihood to obtain a publication bias-adjusted pooled effect size estimate by accounting for the relative publication probabilities in each interval (called weights) and using the weighted likelihood function” (p.8). This method enables us to specify the expected direction of missing published studies and quantifies the adjusted overall mean value like the trim and fill method. Therefore, it is very efficient and easy to evaluate by researchers. The primary limitation of estimation models is that using a p-value informs about the existence or absence of publication bias like PET and PEESE models. Therefore, a more sensitive analysis needs to be applied.

Fixed Effects Estimates

Mean Estimates (μ)

	Estimate	Standard Error	z	p	95% Confidence Interval	
					Lower	Upper
Unadjusted	0.970	0.017	56.828	< .001	0.937	1.004
Adjusted	1.000	0.019	53.818	< .001	0.964	1.037

Note: 1 observation was removed due to missing values.
Note: Only the following one-sided p-value cutoffs were used: 0.025, 0.05.

Random Effects Estimates

Mean Estimates (μ)

	Estimate	Standard Error	z	p	95% Confidence Interval	
					Lower	Upper
Unadjusted	1.113	0.046	24.282	< .001	1.023	1.202
Adjusted	0.920	0.095	9.731	< .001	0.735	1.106

Note: 1 observation was removed due to missing values.
Note: Only the following one-sided p-value cutoffs were used: 0.025, 0.05.

Figure 3.10 The output of the selection models for one sided p-value cut-off s of .025 and .05.

3.4.1.9 Publication Bias in Robust Bayesian Meta-Analysis

This is the most comprehensive and informative method for publication bias. The trim and fill methods, PET and PEESE, and selection models have serious limitations in terms of precision and the distribution of effect size values. In order to obtain a more comprehensive finding, robust Bayesian meta-analysis (RoBMA) promotes the necessity of model combinations in which a number of models are simultaneously averaged and able to summarize the results (Maier et al., 2020). This analysis comprehends 36 different types of publication bias analysis, including selection models and PET-PEESE models. The final state of knowledge enables researchers to quantify the previous findings by averaging and reporting. Additionally, there are some limitations in selecting the random or fixed-effect models for the trim and fill method. RoBMA is free from p-value or selection model problems by means of averaging all these models (Gronau, 2021). In this way, more robust findings on publication bias are yielded.

3.4.2 Primary Study Concerns

Primary study quality is a prominent factor during the investigation of research questions for research synthesis (Valentine, 2019). However, subjective and unclear definitions of study quality make it difficult to address a common consensus on this issue. Therefore, setting the quality of studies as an inclusion or exclusion criteria is a discussion concern in meta-analyses. In the meta-analysis literature, the problem of including poor-quality studies in analyses is described as “Garbage In Garbage Out” issue. (Lipsey & Wilson, 2001; Sharpe, 1997). Some researchers also use the study quality as an explanatory variable to inform researchers and readers on the validity issues (Chadwick, 1997; Lipsey & Wilson, 2001; Shadish & Haddock, 1994). On the other hand, it is not worthwhile to remove studies due to their quality concerns (Borenstein et al., 2009). So researchers should set eligibility criteria to ensure whether the included studies are similar characteristics or not.

Eligibility criteria were set for this meta-analysis to include and exclude studies into analyses. Because of the above concerns and as suggested by Glass (2006) and Borenstein et al. (2009), study quality was not set as an eligibility criterion. But, the possible explanatory moderators related to validity concerns that might influence the treatment effect were investigated. For example, experimental design, researcher effect, teacher effect, teacher training, treatment verification, type of assessment instrument, confusion method and medium, intervention length, intervention intensity, level of control internal validity threads, and outcome measure type were investigated by simple and multiple meta-regressions.

Cooper et al. (2019) state that good studies use research methods that are well aligned to investigate research questions. Therefore, the research design is critical to examining research questions. In this study, the internal and external validity issues are the main concerns of study quality. Therefore, investigating the validity issues is critical to define study quality. Internal validity has two important characteristics for study quality: high measurement validity and randomization (Borenstein et al., 2009; Cooper et al., 2019). Some primary studies and meta-analyses also imply that there is a relation between study quality and treatment effect (Chadwick, 1997; Dechartres et al., 2016; Shadish et al., 2008). In fact, there is no certain direction of findings for low or high-quality studies. Therefore, researchers do not have an exact idea about the influence of study quality.

Experimental design is one of the possible explanatory variables that have an influence on the validity of studies for treatment effect. According to Fraenkel et al. (2012), randomization is the most effective way of controlling internal validity threads, which enables identifying study quality (Figure 3.11). There are different groupings for designs in the literature, but Fraenkel et al. (2012) stated the most commonly used definition as poor experimental, quasi-experimental, true experimental and factorial design studies. For the scope of this study, we included just control group designs and if there is no randomization or extra design for internal validity threat control while setting experimental groups, this type of design is accepted as the most sensitive type for validity threads and called poor experimental design. If the researcher randomly assigned the control and experimental groups rather than individuals or intend to control internal validity threats by factorial designs, we called quasi-experimental. If the researcher randomly select the samples for the group members to the control and experimental groups, we called true experimental design.

Threat	Technique: Standardize Conditions	Obtain More Information on Subjects	Obtain More Information on Details	Choose an Appropriate Design
Subject characteristics		X		X
Mortality		X		X
Location	X		X	X
Instrumentation	X		X	
Testing				X
History			X	X
Maturation		X		X
Subject attitude	X		X	X
Regression		X		X
Implementation	X		X	X

Figure 3.11 General techniques for controlling threats to internal validity (Fraenkel et al., 2012, p.180)

The potential effect of experimenter traits on findings has been discussed for years, especially in educational psychology literature to ensure internal validity and study quality (Dusek,1971; Kennedy, 1976; Kintz et al., 1965; Rosenthal, 1979). These traits can be listed as gender, academic performance, teacher training, and the same teacher effect. Dusek (1971) stated that student achievement performance is strongly related to the academic performance levels of teachers. Since the teacher's capacity for student orientation may significantly affect the treatment effect, it is substantial to investigate the teacher's abilities. Another effective dimension of experimenter traits is that trained teachers are more likely to apply methods more efficiently. Experimenters mostly integrate teacher training for a more effective instruction process. But the critical issue is that this is not standardized or not all studies give importance to the training process. In this sense, the effect of training will be discussed more comprehensively in this study. Teacher effect is also discussed in the literature as the teachers are the same or different for control and experimental groups. It is commonly stated that the same teacher for both groups is more favorable to allowing controlling experimenter effect. But, it is also common that different groups have been instructed by different teachers. In this study, experimenter effect is investigated under the explanatory variables under three titles researcher effect (researcher is the only teacher, researcher is one of the teacher or researcher is not teacher), teacher effect (same teacher or different teachers for control and experimental groups) and teacher training (stated as trained or not stated).

Controlling the implementation process also enables to control of internal validity threads and enhances the study quality (Figure 3.11). Different aspects of implementation should be considered to handle validity issues in literature comprehensively. We investigated intervention length, treatment verification, and intensity through meta-regression analyses. Especially, meta-analysts have intensely handled intervention length as an explanatory variable in science education (Arık & Yılmaz, 2020; Armağan, 2011; Chadwick, 1997).

3.5 Data Collection Process

The main theme of this synthesis is evaluating and understanding the composition of a vast amount of studies in conceptual change strategies (CCS). Mainly, meta-analysis aims to provide quantitative evidence and use theoretical arguments to evaluate study findings (Borenstein, 2009). There is no necessity to collect a certain number of samples to conduct a meta-analysis, but the sample should reflect the population attributions. In addition to the number of studies, the scope of the studies is very important for meta-analysis. From a deeper perspective, the distribution of results in primary studies is a key component of successful meta-analytical work that discloses the heterogeneity in the intended population (Maksimovic, 2011).

The essence of methodology in meta-analysis is that reviewed samples should include a certain number of properties of the population to test researchers' predefined hypotheses more comprehensively. Therefore, meta-analysts need more primary studies to obtain more attributions from a population. Moreover, more precise evidence can be obtained from a large sample size in which meta-analysis studies ensure an adequate sample size to estimate the essence of the literature. From this perspective, the greater number of samples provides a more comprehensive analysis of the literature. In this sense, the collecting study is one of the most prominent parts of this study. In order to conduct a comprehensive and well-detailed meta-analysis, we systematically reviewed the available studies in the field. Therefore it is critical to make a systematic and comprehensive literature search in databases and journals.

3.5.1 Inclusion Criteria for Studies

Primary studies should have the following criteria to be included in this meta-analysis:

1. The study's research design should be experimental.
2. The study should focus on a between-subjects comparison of CCS and traditional instruction (The researcher must conduct traditional instruction for control group).
3. The dependent variable of the study should be achievement in a science domain.
4. The study should be published/reported after 1983 since the theoretical framework for instructional implications of conceptual change was constructed by Posner et al. (1982).
5. The study should be reported in English or Turkish.
6. The study should provide enough data to calculate an effect size measure.
7. Study sample should be student.
8. Duplications should be removed

Problems that may not be noticed during the review stem from duplication, irrelevant strategy or methodology, insufficient statistical data, and irrelevant control groups are resolved during the coding process also.

3.5.2 Literature Search Steps

3.5.2.1 Defining Keywords

There is a vast number of studies that are impossible to review, even in a single topic of science education. Therefore it is essential to define the most relevant keywords and keyword sets to reach relevant studies by conducting a well-established review process. The most reliable keyword description is a review of the field and related studies that include intended keywords for the CCS. Therefore, this study included two stages of literature search: pilot search and systematic search. The purpose of the pilot is to review the related studies to describe the most comprehensive keyword set. In this sense, initially, the below keywords and combination of these keywords were searched for databases to reach the final set of keywords;

"conceptual change" or "misconception in science" or "cognitive conflict" or "refutation texts" or "refutation map" or "conceptual change texts" or "bridging analogy" are searched in seven broad databases as (ERIC, Academic Search Complete, Toylor& Francis Online Journals, Proquest, Education Sources, Wiley Online Library, Google Scholar) reached by METU library (September 2017).
Some practical applications were provided by office programs to make easy and systematic the searching process;

- "a b" the a and b keywords between quotation marks enable to search a and b together. That is, for "conceptual change" search engines review the 'conceptual change' keyword set simultaneously and consecutively.
- *"a*b" * sign between a and b keywords and between* quotation marks *enables to* search a and b together *even if other phrases are inserted between a and b keywords. e.g.,* for "conceptual *change", search engine review the 'conceptual understanding and change in application'. This method is functional to search titles or abstracts for intended keywords.
- *'AND' conjunction enables to search texts including both a and b keywords.*

e.g "refutation" AND "text" search for both 'refutation AND text' keywords simultaneously, but they don't necessarily become consecutively.

- *'OR' conjunction enables one to search texts including a or b keywords.*

  e.g. "refutation" OR "text" search for both keywords one by one or together.

3.5.2.2 Defining Relevant Databases

A database is an organized collection of structured information or data usually stored electronically in a computer system. They may include lots of indexes, including a vast number of journals and articles. A typical database like Scopus includes more than 82 million documents and 234 thousand books, which records 1788 from 17 million authors. There are many databases, including educational studies to search and each database review gives a huge number of related articles about any topic. Therefore, it is impractical to search every database for any subject or keyword. But, they are not isolated from each other so their scope is not independent. On the other hand, the review process has a serious time limitation. That means researchers should know the scope of intended databases to become conscious of the comprehensiveness of their review. In this sense, some critical questions should be answered before the review process. For example, What is the comprehensiveness and scope of Google Scholar? If you search Google Scholar, do you need to search ERIC also? If you search the EBSCOhost databases, do you need to search Scopus also? What is the scope of ProQuest? Or if you searched the ULAKBIM Turkish National Databases, which Turkish articles should you search more? Is EBSCOhost searching the Turkish Databases, Academia or ResearchGate? Researchers may increase the number of such questions. The below process enables us to answer to a certain extent the above questions to guide this study to pursue the most valid way for the searching process.

There are a number of science education databases and indexes that cover totally or partially each other and none of the databases is independent of the others. Therefore, it is practical to begin with the most comprehensive searching tool. In this sense, the university library search engines simultaneously provide a number of databases. The Middle East Technical University (METU) Library serves as an integrated searching tool for researchers who study at METU. This search engine provides a simultaneous search for more than 200 databases and indexes like EBSCOhost, Google Scholar, Scopus, Web of Science, ERIC, Science Direct, Dissertations and Theses, ULAKBIM Turkish National Databases, ProQuest, Education Index, Social Science Index, SSCI, etc. At the same time, it is possible to limit the subject, publication type, language, location, time duration, source type, content provider, and index type. Researchers also use off-campus for 24 hours as cost-free and full access easily. Therefore it is functional to search any field very comprehensively like science education. But, this source also has some limitations for researchers. Some of these databases are completely searched but some are not. Many journals, indexes, and databases are partially included in METU Integrated Search like Google Scholar, Turkish National Thesis Center (TNTC), Turkish Education Index, and ProQuest (Figure 3.12). For example, METU Library search engine searches the Google Scholar and EBSCOhost databases. But it does not search the Academia or ResearchGate which Google Scholar searches. Both Google Scholar and METU Library integrated search tools search lots of common databases like Wiley Online Library, Web of Science, Scopus, Taylor & Francis, SpringerLink, ERIC, Science Direct, etc. At the same time, each database also uses the same indexes and journals. Therefore, it is impossible to draw a scope of each database independent from each other. In this sense, there should be a comprehensive database investigation before the searching process to describe the review scope.

Some visuals are described to clarify the scope of the databases used in this study (Figure 3.13). In order to disclose the relation between databases, randomly selected articles among about 20,000 studies are used by triangulation checking. In this sense, the same articles are reciprocally checked by investigating each database one by one. For example, randomly selected 40 articles among the 20,000 articles without looking at their content obtained by searching Web of Science had written in

METU Unique Search engine to search again. The same articles also check in Scopus to determine the scope of Web of Science on Scopus. The same process was also conducted with the METU search engine and Scopus with 40 different articles which were randomly selected.

It has been seen that all of the articles also exist in the METU library search engine review. Moreover, I contacted METU Librarian responsible for databases to confirm that the METU search engine could search the whole Web of Science documentations. This process also followed other databases providing 100% searches by the METU search engine. Therefore, these databases (like WoS, Scopus, ERIC, EBSCOhost, etc.) are not searched individually since METU Search Engine searches them. It is critical here that researchers may know which databases should also be searched to reach a more comprehensive review. In this sense, further searching is necessary for TNTC (Turkish National Thesis Center), Turkish Education Index, ProQuest, and Google Scholar databases. Their scope is also depicted as below figures (Figure 3.13, Figure 3.14).

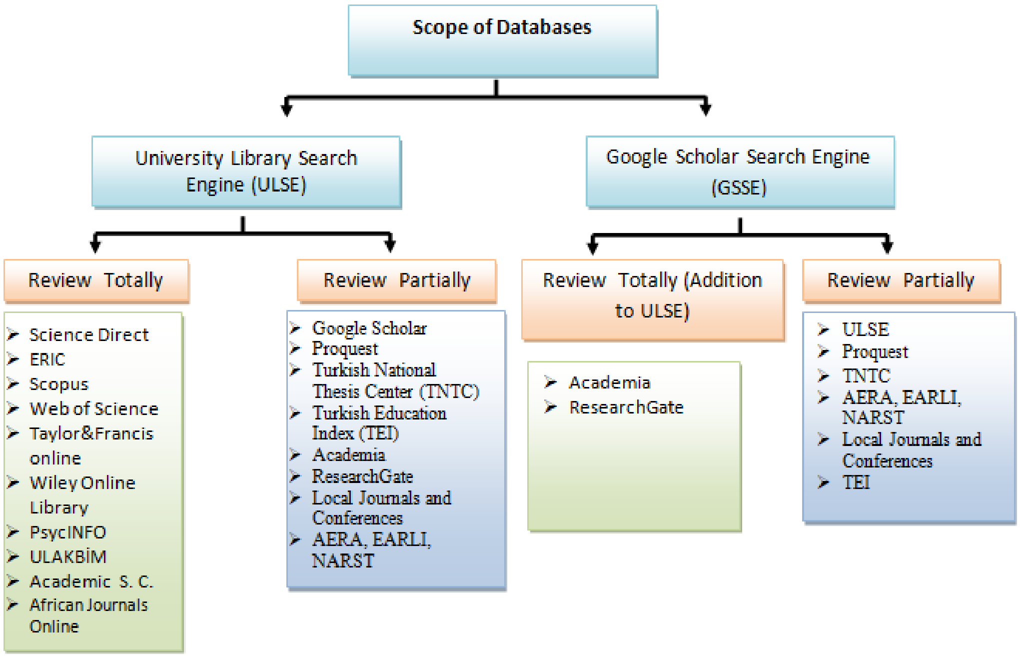


Figure 3.12 Triangulation check on the scope of databases scanned in the study.

The scope of databases was individually depicted at above Figure 3.12. According to findings and information obtained from the METU library database, the METU Search engine covers Scopus, WoS, and EBSCOhost completely. Using mutual comparison, the randomly selected 40 out of 15,500 studies in the WoS database were also searched for Scopus. The 36 out of 40 studies also existed in Scopus which implied that Scopus covered the WoS for 90%. As vice versa, the randomly selected 22 out of 21,250 studies in the Scopus database were also searched for WoS. The 22 out of 40 studies also existed in WoS, which implied that WoS covered the Scopus for 55%. Therefore, it could be observed that the scope of Scopus was larger than WoS. At the same time, there was a significant number of studies that the two databases also share.

A similar investigating process was also done for the EBSCOhost and Scopus databases. The mutual comparison informs us that the randomly selected 40 out of 15,521 studies in the EBSCO database are also searched for Scopus. 27 studies out of 40 studies in the EBSCO database also exist in Scopus, which implies that Scopus covers the EBSCO host for 68%. As vice versa, the randomly selected 40 studies out of 17,255 studies in the Scopus database were also searched for EBSCOhost. 21 out of 40 studies also exist in the EBSCOhost, which implies that the EBSCOhost covers the Scopus for 53%. Therefore, it can be observed that the scope of Scopus is larger than the EBSCO host. However, randomly selected 40 studies in Scopus and EBSCOhost databases were also searched using the METU search engine. All of the studies also exist in the METU search engine, which implies that the METU search engine fully covers both Scopus and EBSCO host databases. Therefore, it is unnecessary to review the databases that METU search engine scans fully. Some of these databases were listed in Figure 3.12 and all databases were also listed in Appendix B.

As depicted in Figure 3.13, the relationship between the scope of five databases tries to be summarized in a given figure. The randomly selected 142 out of 157 studies in the METU Search Engine database were also included by Google Scholar, implying that Google Scholar covers the METU Search Engine for 90%. As vice versa, randomly selected 126 studies out of 150 studies in the Google Scholar database were also included by METU Search Engine, which implies that METU Search Engine covers Google Scholar for 90%. But these two databases do not fully cover TNTC and ProQuest. At the same time, TNTC, ResearchGate, and ProQuest do not cover each other. If you searched for Google Scholar, it is unnecessary to review for ResearchGate.

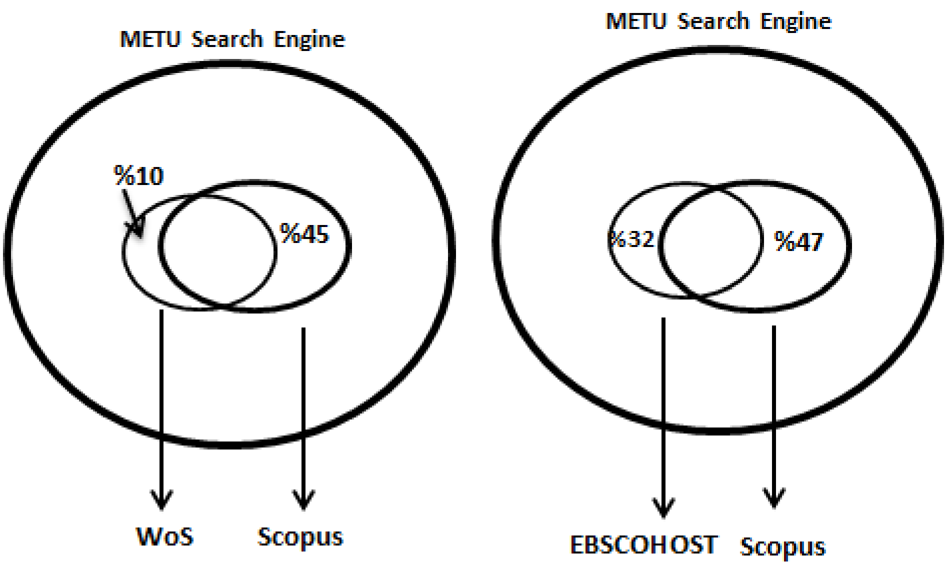


Figure 3.13 The visual demonstration of the relationship between METU library search engine, Scopus and WoS databases.

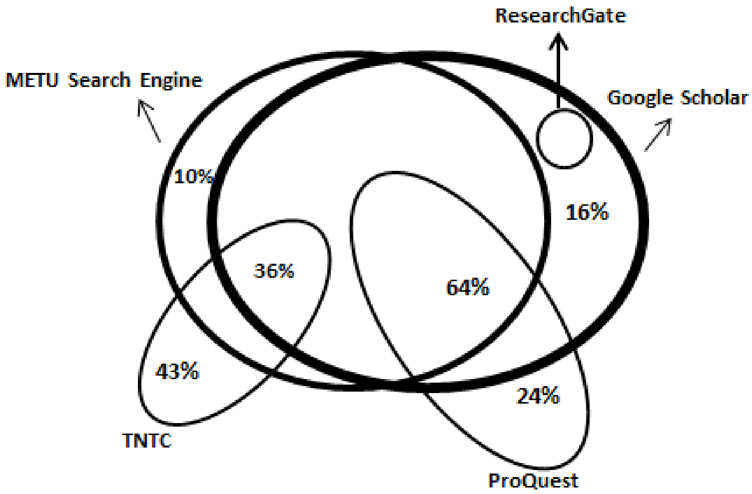


Figure3.14 The visual demonstration of the relationship between METU library search engine, Google Scholar, Turkish National Thesis Database (TNTC), ProQuest and ResearchGate databases.

These databases were reviewed in detail and progressively to reach the most effective and comprehensive review process for articles, conference papers, thesis, and dissertations. I investigated each database carefully to determine relevant review process for CCS. The period of defining keywords and defining databases for a comprehensive literature review took about two years for this study.

3.5.2.3 Pilot Search

CCS is not very defined in literature as it is stated in previous meta-analyses (Armağan, 2011; Gelen, 2015; Guzzetti et al., 1993; Mufit et al., 2020). The nature of conceptual change makes it necessary to describe clearly this field as different from other instructional methods. For example, Armağan (2011) mainly focused on conceptual change texts as the only source of CCS, Mufit et al. (2020) and Rahim et al. (2015) stressed the effectiveness of cognitive conflict, Guzzetti et al. (1993) investigated both text structures and discussion methods. Gelen (2015) reviewed assisted conceptual change materials like concept maps, cartoons, and conceptual change texts as CCS. The common point for previous meta-analyses is that they had mainly focused on the Posner et al. (1982) description of cognitive conflict for the scope of CCS. On the other hand, there were serious criticisms that existing conceptual change strategies were derived from different knowledge perspectives except for the dissatisfaction process as previously described in the literature review part. Therefore, the comprehensive meta-analysis on CCS should cover existing CCS derived from different knowledge perspectives and the intended sample should reflect each perspective in the population. In this sense, the pilot search process had three prominent aims (1) obtaining a number of sources about whether there was enough study to conduct a comprehensive meta-analysis or not (2) describing the possible keyword sets to reach a more representative sample of the population and (3) clarifying and covering the field of CCS that derived from different knowledge perspectives.

We revised the initial keywords based on the pilot search to create clusters of keywords. The first cluster was *“conceptual change” OR “cognitive conflict” OR "conflict map" OR “refutation text” OR "refutation map" OR “bridging analogy” OR “anchoring analogy” OR“ontological category” OR “ontological shift” OR “conceptual shift” OR “cognitive shift”.*

It was observed that the cognitive conflict perspective was the most popular strategy in conceptual change (73%). The rest of the studies were about bridging and ontology perspectives. This search also implied that there was a significant number of studies except cognitive conflict perspective. The total number of studies obtained by relevant keywords is about 7,191 studies on their title and abstracts which is a severe number for any field for meta-analysis. The result of this pilot study informed us there was enough study for a comprehensive meta-analysis.

Table 3.7 Total number of sources reviewed in pilot search for CCSs

Strategy	Keywords	Databases						
		ERIC	Academic Search Complete	Scopus	Toylor&Francis	ProQuest	WoS	Total
Cognitive Conflict	“cognitive conflict”	137	523	1087	1687	1002	837	5273
Ontological Category Shift	“ontology” AND “conceptual change”	7	12	36	71	139	30	285
	“ontological category”	10	42	129	524	91	77	873
Cognitive Bridging	“analogy” AND “conceptual change”	23	23	97	108	315	97	663
	“bridging analogy”	4	3	17	24	9	8	65
	“anchoring analogy”	1	1	3	3	0	2	10
Total		182	606	1379	2417	1556	1051	7191

Secondly, defining possible keywords that we were able to search both comprehensive and relevant studies is crucial. The pilot study improved the understanding of critical keyword sets to reach a more representative sample of the population. As a result of the above findings and applying eligibility criteria, the relevant study in the field was about 68 for the pilot study (Figure 3.15). Most of the sources were journal articles (46), secondly dissertations (15), the least one is master thesis (7), there is no paper yet. The investigation of the 68 studies that derived from the pilot study enabled us to describe the below keyword sets for systematic search. The period of pilot search was about 2 years for this study.

Final set of keywords is described as below;

Keyword Phrases

“conceptual change” OR “cognitive conflict” OR "conflict map" OR“refutation text” OR"refutation map" OR“bridging analogy“ OR “anchoring analogy” OR “ontological category” OR “ontological shift” OR “conceptual shift” OR “cognitive shift”

Keyword sets

"refutation AND conflict" OR "dissatisfaction AND misconception" OR "conflict AND misconception" OR "ontology AND misconception" OR "analogy AND misconception" . and Turkish versions of these words for Turkish sources.

Thirdly, clarifying conceptual change strategies and their scope is one of the main issues of this study. The pilot study can disclose the most studied fields of CCS (Table 3.7). According to the pilot review, cognitive conflict is the most common but not only strategy for conceptual change. Therefore, this review gives evidence for further investigation on different conceptual change strategies like analogy-based and ontology-based approaches rather than conflict perspectives.

The keywords that describe conceptual change process have also been examined. These keywords are not used in the pilot review process but used to describe conceptual change strategies in the eligibility process. Therefore, they are so critical to describing the final set of studies.

The critical keywords to describe the grasped primary studies on whether they are compatible with conceptual change strategies are or not;

conceptual change, misconception, alternative conception, naïve conception, mis-ideas, naïve ideas, naïve knowledge, primary conception, weak conception, belief-like structures, intuitive conceptions.

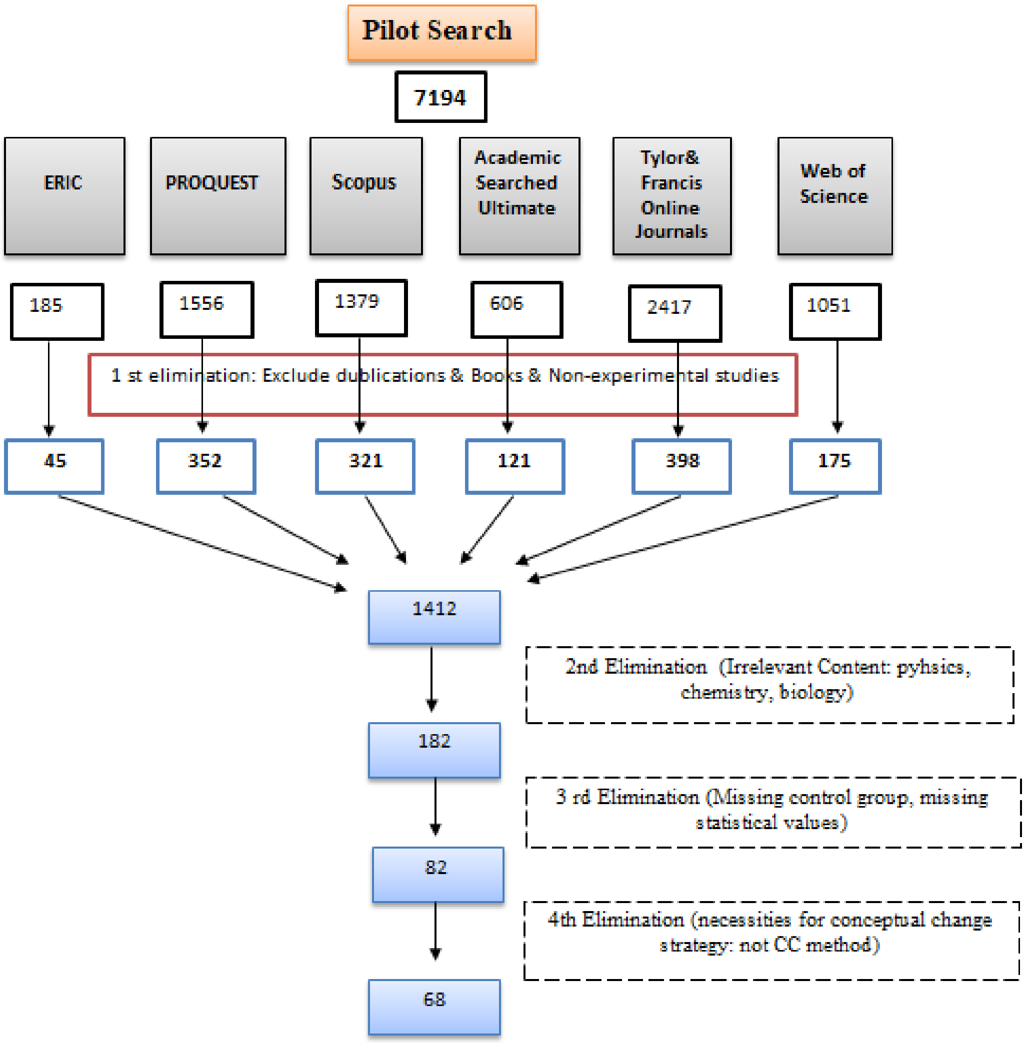


Figure 3.15 Application of eligibility process for pilot study findings.

3.5.2.3 Systematic Literature Search

After the pilot study, it was described the feasibility of study, critical keywords and possible conceptual change strategies that would be reached. The databases were defined very systematically and their scope was described during this study. In this way, the necessary knowledge was gathered to conduct a comprehensive search process for conceptual change strategies. As a result of the systematic search, ı gathered 161 relevant studies meeting the eligibility criteria (Figure 3.16).

The easiest and most comprehensive way is to use the METU library on off-campus status as a METU student and complete the searching with further scanning on different databases and indexes that are not fully covered by the METU Library search engine. This engine also decreases the possibility of duplication problems for a number of studies that may be published in different databases simultaneously. Some of the databases were addressed below to illustrate the comprehensiveness of the reviewing process. The databases that I fully covered by using the METU Library search engine were;
ERIC, Science Direct, Academic Search Ultimate, Scopus, Web of Science, EBSCOhost, Tylor, and Francis Online Journals, ULAKBİM, Dergipark, African Journals Online, SSI, SSCI, Teacher Reference Center, Wiley Online Library.

Except the METU Library Search Engine, I searched Google Scholar, Turkish National Thesis Center, Turkish Education Index, ProQuest, Journal of Turkish Science Education, Journal of Research in Science Teaching, International Journal of Science Education, MDPI Education Sciences, AERA (American Educational Research Association), EARLI European Association for Research on Learning and Instruction), and NARST (National Association of Research in Science Teaching).

The journals that I searched for further review to collect Turkish studies were Eğitimde Kuram ve Uygulama, Kastamonu Eğitim Dergisi, Milli Eğitim Dergisi, Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, Hacettepe Üniversitesi Eğitim

Fakültesi Dergisi, Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi, Ankara University, Journal of Faculty of Educational Sciences, Elektronik Sosyal Bilimler Dergisi, Eğitim bilimleri indeksi, Sobiad, Kara Harp Okulu Bilim Dergisi, Bayburt Eğitim Dergisi and TR dizin. By using smart search property, related articles were limited to subject, year, content provider, and publication type.

The third step for searching process was the secondary search which included reviewing references and authors's CV. The reference list, obtained from the studies we had, helped us to reach further studies that may be overlooked during searching. At the same time, it is possible to notice further databases or journals that should be searched. Furthermore, personal contacts was required to reach studies with no full internet access. Some studies need a membership for databases or university library sources. Therefore, some studies need extra effort and financial sources to reach during the literature search. For the scope of this study, about 14.000 references from 345 related primary studies were searched. 267 related studies was reached for the further eligibility process. After reviewing related study references, it is worthwhile to collect information about researchers who study on conceptual change strategies. In this way, some of the studies which exist in their bibliography that might not be published in articles or taken down from articles were reached. In this sense, 5300 studies from 278 researcher bibliographies were reviewed. At the end of these reviewing, I collected 42 related studies. Conference papers, special reports and unpublished studies were also researched by using authors's CV. During this study, it was contacted more than 200 authors from very different parts of the world about their studies on conceptual change. The literature search began in 2014 and lasted until October 2021.

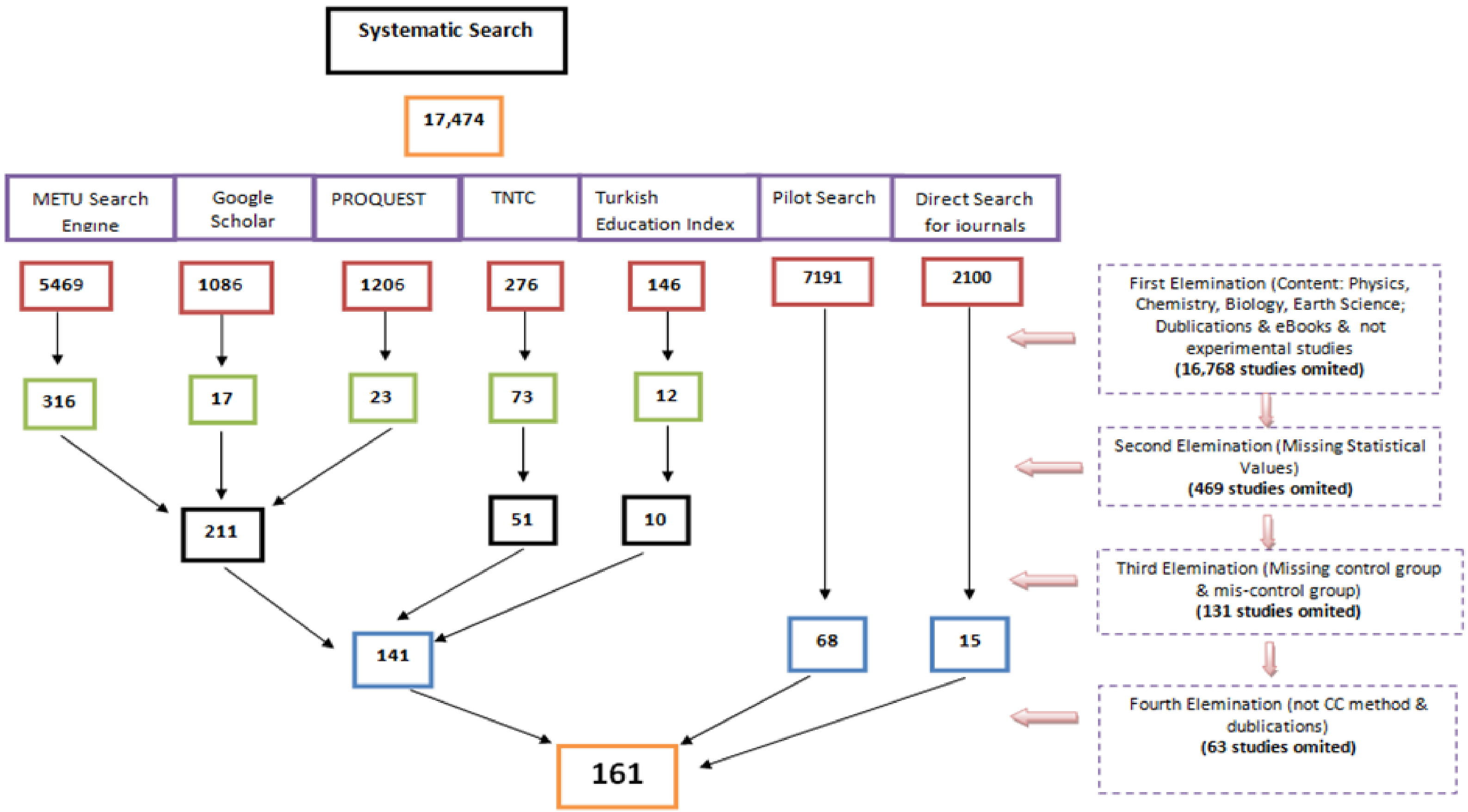


Figure 3.16 Systematic searching process steps and eligibility criteria demonstration.

The eligibility criteria are a big deal for conceptual change literature due to the comprehensive nature of methodology for researchers. Therefore, most of the studies are excluded due to the methodological limitations on strategy. The properties of conceptual change strategies and critical keywords that describe the CCS were also defined in the pilot search process. Studies firstly eliminated with respect to their content as physics, chemistry, and biology. There were lots of studies performed on the medical field and maths subjects. Secondly, the study design should be experimental which includes the control group as exposed to traditional instruction. The dependent variable for this study is achievement. In this sense, primary studies should report achievement scores rather than affective attributions. Moreover, we investigated 23 moderator variables through meta-regression analyses. Thirdly, missing statistical value is critical to measuring effect size value. A significant number of studies did not give statistical data to measure effect size. Fourthly, the necessities for conceptual change strategy may not be ensured by researchers. For example, researchers expose refutational texts to disclose misconceptions rather than eliminate them. Therefore, some studies do not satisfy conceptual change instruction processes but focuse on misconception detection.

We reported the number of relevant studies and the whole seraching process by using PRISMA flow chart diagram that depicted at Figure 3.17. This diagram includes three rows and two colomns. The first row represents the identification fields for studies in different databases, indexes and journals. The second row represents the screening and inclusion processes for studies by using pre-defined keyword sets. The third row repsresents the included studies as a result of searching processes. The number of the final included studies is 218.

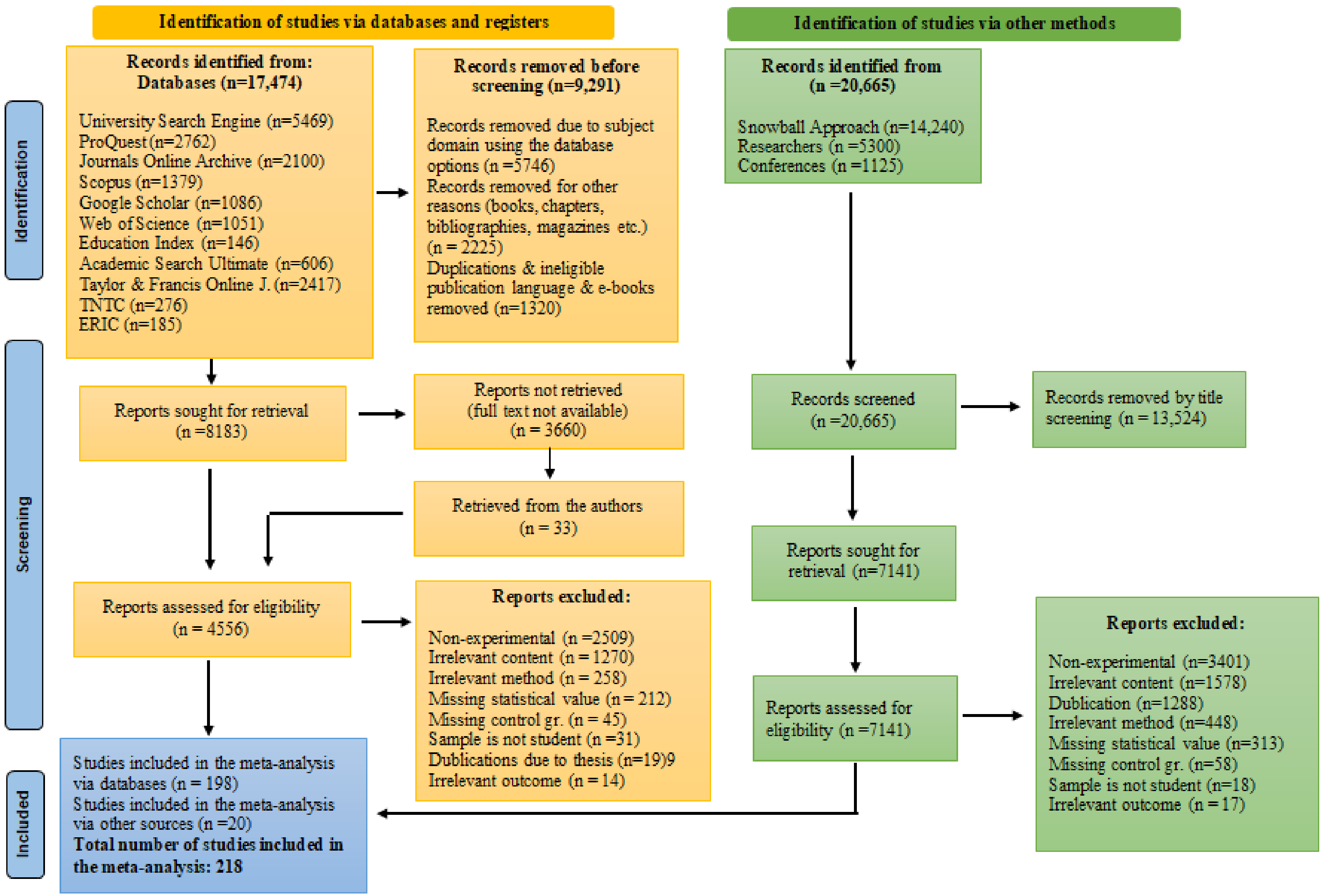


Figure 3.17 Flow diagram depicting the acquisition process of primary studies (PRISMA; Page et al. 2021).

Studies requested and obtained by email: 10 (Çiğdem Şahin (1), Refik Dilber (2), Ruth Stavy (1), Sibel Er Nas (2), Yavuz Akbaş (1), Gonca Çakmak (1), Yang, et al. 2012 (2)

Studies Requested by ResearchGate: 36 articles

Obtained Studies: 23 (Hedges et al., 2010; Magliaocca et al., 2014; She & Lee, 2008; Bahar, 2003 (3); Luara et al., 2005; Chatila et al., 2009; Weller, 1995; Suits, 2000; Hynd & Alverman, 1986; Coetzee & Imenda, 2012; Dilber, 2010; Akbulut et al., 2007; Çepni et al., 2006; Abuhillail, 2019; Palmer, 2003; Balcı et al., 2006; Chinn & Brewster,1993; Creswell, 2009; Başer, 2006 (2); Yürük & Geban, 2001). Literature search and study selection steps stated at Figure 3.18.

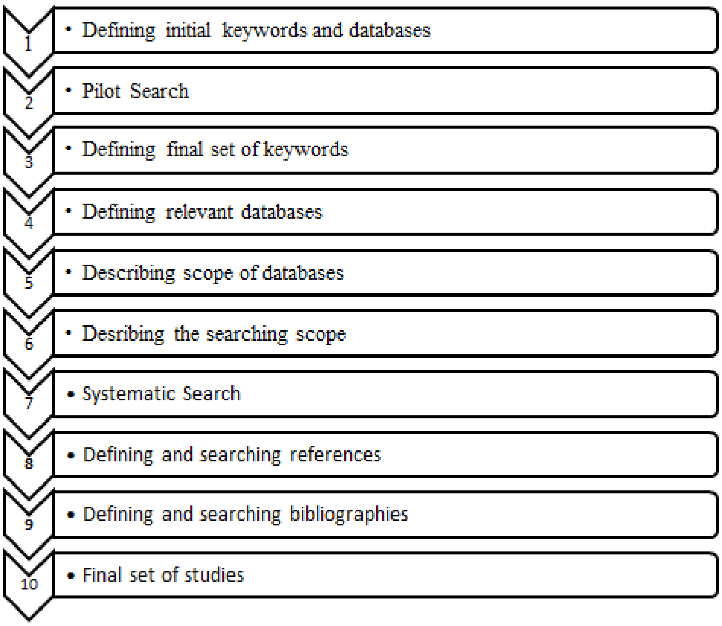


Figure 3.18 Literature search and study selection steps

3.6 Study Coding Process

3.6.1 Coding Sheet and Coding Manual Development Process

Coding process takes serious time for researchers to complete. For this study, developing the process for the coding sheet and coding manual to form the final version took about three and a half years. We reviewed several primary studies and meta-analyses in detail. Three experts in the field of education provided lots of elaborative feedbacks. The feedback process was continuously provided by two advisors throughout the study process also. Seven different coders also coded each item, and changes were made by using their feedback also. Therefore, many changes had been made for each item during the development process to reach the most reasonable and coder-friendly version of the coding sheet.

The primary study coding process is one of the crucial steps of meta-analysis to ensure analysis reliability. Wilson (2019) informs that novice meta-analysts make crucial mistakes on three points. Firstly, recording the one effect size value for a single study leads to missing the composition of multiple effect sizes or using dependent effect sizes problems. Secondly, researchers underestimate the time consumption for the coding process. Especially dissertations and thesis take hours to code for even a single study. Therefore, novice researchers mostly follow a less comprehensive coding process to take more time for analysis. There are 218 primary studies for this study, which is a very large number to code for a meta-analysis. That is why time management through the well-designed coding process is crucial. Finally, detailed coding is prominent in conducting a comprehensive meta-analysis. Novice meta-analysts generally prefer to code ultimately used moderators which reduces the utility of meta-analysis. Therefore, it is necessary to develop a study-specific coding sheet and coding manual that describes coding items and statistical data. The coding sheet used in this study includes 38 items for moderators and detailed parts to record statistical values. The development process of the coding sheet and manual can be summarized as;

Firstly, the previous comprehensive meta-analyses were examined to disclose the possible explanatory variables (Arık & Yılmaz, 2020; Armağan, 2011; Bayraktar, 2000; Chadwick, 1997; Kulik et al., 1980; Üstün, 2012). These studies inform the common explanatory variables for meta-analyses. The included items grasped from these studies were publication year, treatment duration, publication type, country/region, experimental design, teacher and researcher effect, sampling method, subject domain, education level, sample size, class size, treatment characteristics, instruments, and statistical data part.

Secondly, the relevant meta-analyses on CCS were investigated to reach more specific variables (Armağan, 2011; Gelen, 2015; Guzzetti et al., 1993; Mufit et al., 2020). These studies also include study-specific variables due to the different nature of studies. For example, Guzzetti et al. (1993) focused on text-based and discussion-based CCS studies for moderator analysis. Armağan (2011) informed researchers about the effect of different conceptual change techniques. On the other hand, Gelen (2015) focused on cognitive conflict tools like concept maps and cartoons. Therefore, it is crucial to investigate related primary studies on CCS as special for this meta-analysis. The included items from these studies were integrated instructional material and computer effect.

Thirdly, the study-special attributions of CCS inform us about possible explanatory variables. At the same time, relevant primary studies included in this study enable the coding sheet to advance. The study-specific items are a type of conceptual change strategy, material, type of outcome measuring, and intervention intensity. One of the base assumptions of this study is testing the relation between each conceptual change strategy and evaluating the subgroups for the effectiveness as three types of conceptual change strategies.

The first draft of the coding sheet was developed on Microsoft Word with 36 items. The items were listed on a word sheet that enables coders to select. But increasing primary studies lead to an increase in the moderator variables which should be investigated for treatment effect. Therefore, increasing moderators and the statistical data forms (add Chi-square, nonparametric scores, and z-value) made it unpractical to use in Word format. Moreover, as a result of pilot coding for 40 studies, treatment fidelity, age, and grade level were omitted.

The second draft of the coding sheet was developed on a Microsoft Excel sheet which is based on the feedback provided from experts in the field of education one of them has previously prepared a well-advanced coding sheet and manual. This excel sheet made the selection and working process more effective and coder friendly. A number of codes can be copied and pasted from excel to meta-analysis programs easily in this way. At the same time, practical calculations for effect size values like weighting more than one effect size value for the same study or changing Cohen's *d* value to Hedges*'g* were done very practically by integrating the formula into an excel sheet. The number of primary studies is 218 for this study, at the same time, 274 effect size values were used for calculations. In this sense, it is so time-consuming to work with such a large number of studies in a limited period of time. Therefore, excel is a very effective and necessary tool for this study. Finally, we added measuring outcome (achievement test or conceptual change test), a number of tiers (one tier, two tiers, more), and treatment interval items. Pilot search, medium effect, level of internal validity items, and gender were omitted as a result of feedback and coding problems. The publication year was divided as received year and implementation year to obtain more detailed knowledge. Medium and level of controlling internal validity added to internal validity threats items, teaching method changed as a type of conceptual change method, teaching method medium changed as an instructional medium.

The third draft of the coding sheet was also developed in Microsoft Excel format with respect to feedback getting from six coders and four experts (Figure 3.19). These coders are also experts in educational studies. In the final draft, treatment intensity and class size items were added. The country item was coded as region during the analysis, year was induced to one variable for analysis. School-level was changed as educational level, group design was changed as experimental design, and instructional medium was changed as material.

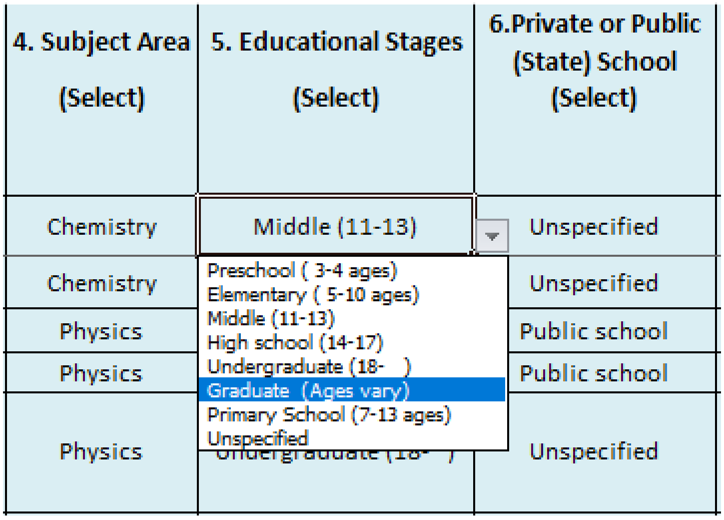


Figure 3.19 Third draft of coding sheet on Microsoft Excel

The coding manual was also advanced during the development process for the coding sheet. We stated an example for the year item (Figure 3.20). The manual includes very detailed descriptions of items for coders and researchers to ensure the reliability of the coding process. Coders can get information about how to code, the information about subitems, and the purpose of coding the items for study. Most importantly, coders should infer some information for items that are not reported in primary studies like publication type, implementation year, country, subject, experimental design, or outcome measure type. Coders should infer the information from reported information. For example, the researcher might not report the experimental design but give information on the randomization process or the country where the study was done can be inferred from the researchers' university. Therefore, the coding manual should be so systematic and detailed that coders could code easily for each item without any help except from the manual. Each item was also depicted next to the item information in the manual.

2. Year

2.1 Received Year (Select)

In this item, you will select an option between an open list within three choices. Later you will explicitly write just the year (not month or day) numerically to the item you selected from open list. The choices are listed below with order.

- **Received Year:** The year that the article was initially submitted by the author.
- **Accepted Year:** The year that the article was initially accepted by the article.
- **Published Year:** The year that the article was published by the article.

Figure 2. Received Year

For the articles, write the received year of the study, which is the year the article was first submitted by the author. If the received year is not indicated, record the year that article was accepted to be published on the journal. If the accepted year is not indicated either then write the publication year of the study. (Please follow the order).

For the master or dissertation theses, directly record the defense year as the publication year.

Figure 3.20 Coding manual prepared for item 2.

3.6.2 Moderating Factors in Conceptual Change Instruction

3.6.2.1 Publication Characteristics

The effect of publication characteristics is commonly criticized in meta-analytic literature. For example, dissertations are more detailed and controlled studies than master theses. So, moderators may be controlled more comprehensively to provide more valid evaluations in dissertations across other types. This tendency can change the overall effect of meta-analyses. That is why researchers should consider the moderating effects in different publication characteristics to investigate the moderator's combined effect and independent effect on effect value.

3.6.2.2 Sample Characteristics

Meta-analysis includes studies with very different sample characteristics like region, school type, school location, education level, sample size, class size, etc. The distribution of effect value may also stem from these characteristics. For example, some meta-analytic studies focus on educational interventions working better in lower grades like preschool and elementary (Arık & Yılmaz, 2020; Kim et al., 2021). Additionally, class size (Chadwick, 1997) or region (Yu, 2021) sample characteristics can be explanatory moderators for overall effect distribution which should be the researchers' focus.

3.6.2.3 Design Characteristics

As a prominent criterion, the quality of studies is closely related to experimental design in the assignment process, which implies that randomization interacts significantly with the treatment effect. "Random assignment is intended to eliminate the threat of extraneous, or additional, variables—not only those of which researchers are aware but also those of which they are not aware—that might affect the outcome of the study" (Fraenkel et al., 2011, p.267). That is why it is valuable to control the possible roles of experiment design in the treatment process.

Additionally, low internal validity may be responsible for implausible alternative explanations of the results. Marczyk et al. (2005) propose that researchers should be aware of the potential threats stemming from the teacher effect, researcher effect, assessment instruments, implementation, and verification processes to avoid validity problems. Otherwise, researchers cannot draw valid inferences and causality for the study findings. In this sense, some moderators responsible for internal validity issues were included in meta-analyses like teacher effect (same or different teachers for control and experimental groups), treatment verification, and teacher training process.

3.6.2.4 Intervention Characteristics

Different instructional implications of CCS suggest different conceptual change mechanisms while defining educational interventions and yield very divergent effectiveness on science achievement (Slotta & Chi, 2006; Smith et al., 1993; Tsai, 2003; Zohar & Kravetsky, 2005). Additionally, subject domain and type of material have similar impacts on achievement scores. Therefore, intervention characteristics may be responsible for significant variation in effect size values.

Additionally, the effect of intervention length was frequently criticized in literature (Chadwick, 1997; Clark, 1983; White, 1988). One important hypothesis is that short intervention length causes a lack of deeper knowledge comprehension. That is why increasing intervention length enables most likely to boost effect value (White, 1988). The novelty effect on the intervention process may also be effective in effect value. For the scope of this study, intervention length and intervention intensity were also investigated to clarify the confounding effects.

3.6.2.5 Measurement Characteristics

Measuring the process is critical because some primary studies use general achievement tests to measure the effect of CCS. On the other hand, conceptual change strategy impacts student misconceptions rather than factual knowledge. In this sense, it is expected that the studies that use a misconception test for assessment should provide a higher effect value than studies that use general achievement tests.

Assessment instrument type (adapted test, pre-existing test, or researcher-developed test), question type for assessment process (open-ended, objective type, or mix) are also associated with the variation in effect size. For example, the validated standardized tests are more robust against random errors and researcher bias (Bayraktar, 2000). Therefore, it is expected to have less effect on treatment for systematic errors. Similarly, open-ended questions aim to disclose higher learning outcomes rather than knowledge levels. That is why the intended learning outcome may change with measurement characteristics. This situation may be responsible for variation in effect sizes. We reported the intended study characteristics, variables, and subgroups that investigated the scope of this study in Table 3.8.

Table 3.8 Study characteristics on the efficacy of CCS for science achievement.

Study Characteristics	Variables	Subgroups
Publication Characteristics	Publication Type	Journal articles, Doctoral dissertations, Master theses, and Proceedings
	Publication Year	1989-2000 ,2001-2005, 2006-2010, 2011-2015, 2016-2020
Sample Characteristics	Region	Africa, America, Europe, Asia, Turkey
	Sample Size	16-46, 47-56, 57-72, 73-100, 102-396
	Class size Interval	8-22, 23-26, 27-30, 30-38, 38-87
	Sampling Method	Random Sampling, Nonrandom sampling
	Education Level	Elementary, Middle School, High School, University
	School Location	Urban, Suburban, Rural
	School Type	Public, Private
Design Characteristics	Experimental Design	Poor experimental, Quasi experimental, True Experimental
	Researcher Effect	Not teacher, One of the teachers, Only teacher
	Teacher Effect	Same teacher, Different teachers
	Treatment Verification	Stated, Unstated
	Teacher Training	Stated, Unstated
Intervention Characteristics	Type of CCS	Cognitive conflict, Cognitive bridging, Ontological category shift
	Material	Computer-based, text-based, hands-on
	Subject Domain	Biology, Chemistry, Physics
	Intervention Length	1-48 course hour (continuous)
	Intervention Intensity	1-8 course hours per week(continuous)
Measurement Characteristics	Outcome Measure	Misconception assessment test, general achievement assessment test
	Type of Assessment Instrument	Pre-existing test, adapted test, researcher-developed test
	Type of Questions	Open-ended, Objective, mix for both
	Number of Tiers	One, Two, three, or more, mix

3.6.3 Coding Reliability

The reliability of coding is critical to ensure the quality of meta-analyses. On the other hand, objective coding is a difficult issue and should be considered very professionally. Wilson (2019) states that coding is a subjective, challenging, time-consuming, and tedious process for meta-analysts. Moreover, some factors significantly affect coding reliability. Four prominent factors distorting coding reliability are stated in literature as deficient reporting in primary studies, ambiguities in the judgment process, coder bias, and coder mistakes (Stock, 1994; Wilson, 2019). Stock (1994) recommends that coding should be repeated until apparent consensus is achieved. Moreover, the coding manual that enables the standardized coding process should be detailed, continuously revised, and provide coders with clear directions.

There are two coding reliability measurement processes: intra-rater and inter-rater reliability. The intra-rater reliability is the consistency of different codes for the same coder. The researchers should increase this reliability for meta-analyses by preparing a detailed coding sheet and manual. At the same time, the coder should recode each item number of times.

Interrater reliability is the degree of coding consistency between different coders (Wilson, 2019). At least two different coders should code a random sample of studies. Different statistical indices describe the reliability assessment like agreement rate (AR), Cohen's Kappa and Weighted Kappa, Andrés and Marzo's Delta, Krippendorff's Alpha, intercoder correlation, and intra-class correlation. The *AR* is the most common for meta-analyses because of its simplicity and widespread use (Wilson, 2019), even if there are some limitations. The AR can simply be defined as the percentage of agreement across the items.

There should be at least two coders who have experience with coding and are knowledgeable about science education. For inter-coder reliability, coders need to code the articles twice to control missing points and errors. Another process was the control of time intervals. After a month or more, I also recorded the studies to forget the first draft. At least one coder should also recode some articles randomly.

Then, the ‘agreement rate’ (AR) was calculated for the studies to reach an average AR. The coder reliability should be more than 0.80; nevertheless, we need to recode the articles. The AR is calculated by the following formula (Orwin & Vevea, 2009):

$$AR=\frac{number\ of\ observations\ egreed\ upon}{total\ number\ of\ observations}$$

The main research question for this study is measuring the different conceptual change strategies (CCS) on achievement and they are also study-specific explanatory variables. Therefore, the CCS coding is crucial for this study's reliability. In this sense, there were four parts of the coding reliability process for this study (1) intra-rater reliability process, (2) pilot coding reliability for CCS, (3) comprehensive coding reliability for CCS and finally (4) coding for full codes except for CCS.

For the intra-rater reliability process, there is one coder for this study: the researcher. Wilson (2019) proposes that an intra-rater coder should code many times until the highest reliability is assured. Therefore, the 207 studies are revised three times to increase their reliability. The AR for the final version was measured as 96%. The inconsistency for reliability mainly stems from inferring values that were not directly stated in studies, like the existence of training, verification level, class size, and outcome measure type. The final version was also revised during the inter-rater reliability process. Therefore, reaching a full AR for intra-rater reliability is not meaningful. During this process, the coding sheet and manual were revised many times. This process aims to provide more feedback to revise the sheet and manual to increase inter-rater reliability.

For the pilot study, 22 randomly selected studies out of 207 were coded for the type of conceptual change strategies by three coders (including the researcher) who were knowledgeable on CCS as cognitive conflict, cognitive bridging, ontological category shift, or none of their options. We provided the manual and sheet. Coders read the manual and informed the researcher about unclear directions about strategies. The average AR before the revised process is 81%. On the other hand, after the revision of inconsistencies, the coders coded the items again and the average reliability for AR was measured as 91% which was significantly high.

In the second part, it was requested that authors code their own studies to provide more valid results. Vevea et al. (2019) state that contacting original investigators of studies is one of the main strategies to reduce coding errors. Totally 179 studies were coded into three groups cognitive conflict, cognitive bridging, and ontological category shift, out of 218 studies. The rest 29 studies could not be coded due to some reasons like insufficient reporting or unclear description of the strategy. These 29 studies did not fit the defined strategies, but they were also conceptual change strategies which were analyzed in the first main question. On the other hand, authors should be alive and located in the working place. They should be the original data collectors, they need to keep the knowledge and be willing and able to provide the intended knowledge also (Vevea et al., 2019). Therefore, the provided feedback percentage may not be so high.

For the reliability process, we contacted about 100 authors for 128 studies by using their official mail addresses and bibliographies. It provided feedback for 83 studies by 65 authors. The feedback rate is quite high (65%) for email contact, which implies that this study may also provide significant feedback to authors. The 22 authors are from different parts of the world including Asia, Africa, North America, and Europe. The rest 43 researchers are from 35 different universities in Turkey. The full list of authors who provide feedback was reported in the appendix. AR for CCS reliability is 96% (80 out of 83). Almost all original investigators agreed with grouping for their studies. 2 researchers out of 3 who disagreed with coding provided options rather than directly agree with grouping. We gave the mail content below. We tried to get objective feedback without directing the authors.

“Dear ………………,

I am currently working on my dissertation focused on conceptual change instructions, and at this point I need your help regarding your study entitled "…………………………………………………………………"

As part of my dissertation I categorized the experimental studies on conceptual change instructions as “cognitive conflict”, “cognitive bridging”, and “ontological category change” according to the theoretical approaches taken by the researchers.

Based on this categorization I located your study under the category of “…………………”

To increase the trustworthiness of the categorization process, I would like to know your agreement on locating your instructional approach that you tested in your study under the category of “……………………………………”.

If you kindly let me know your agreement or disagreement about this categorization, I would really pleased it. I briefly described the categories below. If you need further information please let me know. Your study added here.

Best wishes,

Cognitive conflict approach: Focuses on students’ misconception. The major aim of the instruction is to locate students’ misconceptions, falsification of these misconceptions and then helping students’ gain the scientific conception.

Cognitive bridging approach: Focuses on students’ productive conceptions. The major aim of the instruction is to locate students’ productive pre-conceptions and then using them to help students gain the scientific conception.

Ontological category change: Focuses on the ontological nature of students’ conceptions. The major aim of the instruction is to identify the ontological category of students’ conceptions and then helping students shift their conceptions into an appropriate ontological category.

Note: If there is more than one category for single study, please state each separately.

Çağatay PAÇACI
Advisor: Assoc. Prof. Ömer Faruk Özdemir
Co-Advisor: Assist. Prof. Ulaş ÜSTÜN
Department of Mathematics and Science Education
Middle East Technical University
Ankara- Turkey“

In the final coding process, six researchers with educational science experience coded randomly selected 40 primary studies out of 207 studies (Table 3.9). Three of them had Ph.D. degrees in educational science. The other three are Ph.D. candidates in education.

Table 3.9 The interrater reliability measurement for AR for each coder

Coders	Number of Coded studies	First Review (%)	Second Review (%)
1st Coder	9	84	93
2nd Coder	12	89	96
3rd Coder	4	83	92
4th Coder	3	86	96
5th Coder	8	83	-
6th Coder	4	100	-
Total	40	88	94

Before the revision process, the AR for the final coding process was 88%. The inconsistency rate for final coding was 12%. In this sense, the source of inconsistence for this study can be listed as follows;

1- The percentage of wrong coding by the second coder due to not fully scanning articles (4.2%)

2- The percentage of wrong coding by the second coder to not making inferences from articles (3.3%)

3- The percentage of wrong coding by the researcher (2.5%)

4- The percentage of wrong coding by the second coder due to misunderstanding some explanations (length of treatment, researcher effect, instrument type…) (2.1%)

The final revised AR is 94% which is significantly high to pursue further analyses. You can observe the changes in primary and final drafts in Figure 3.21.

The following steps enable us to describe the eligibility of a primary study for a particular method more practically for researchers.

1- Firstly, scan the info that can be obtained from the study title and abstract: Is the study related to a particular method, is the study design quantitative or qualitative, and is the subject of the study relevant?

2- Scan info obtained from methodology: Is the experimental group exposed to conceptual change strategy and the control group traditional strategy.

3- Get from results: Is there statistical data? is it about the achievement of conceptual change?

4- Literature review: Decide whether or not the conceptual change is conducted as a treatment method.

5- Procedure: Define the type of CCS.

6- Look for duplications (compare researchers' early thesis and papers)

7- Code the moderators

8- Code statistical values

9- Compute the effect size in Hedges' *g* index.

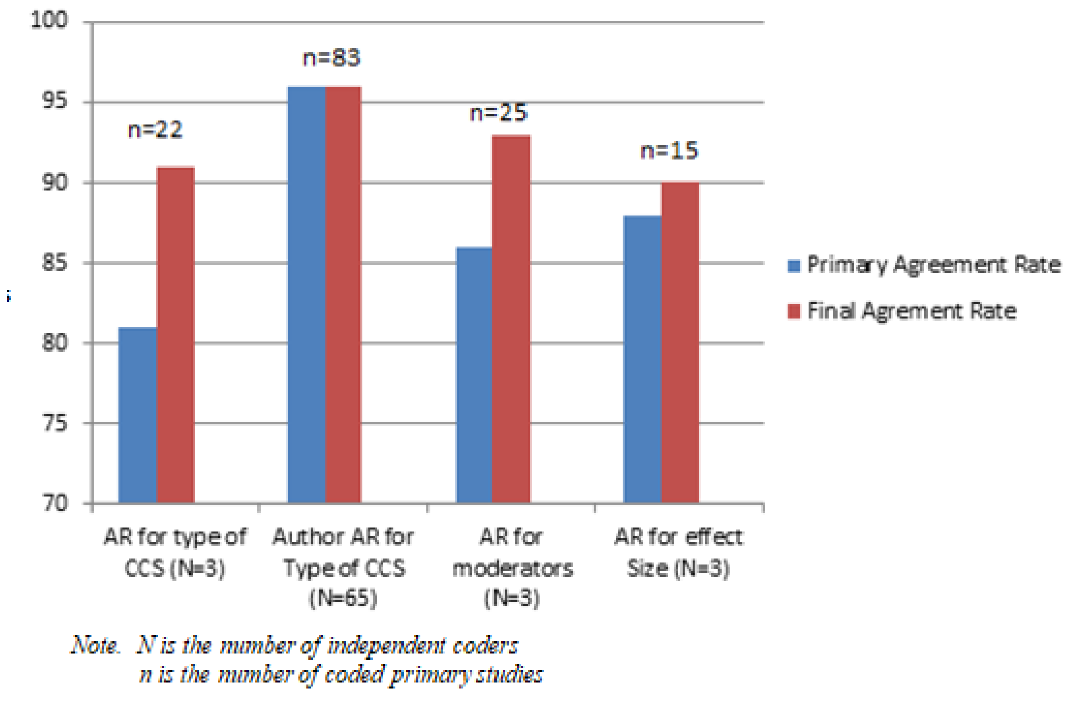


Figure 3.21 Agreement rates

3.7 Statistical Issues in Meta-Analysis

Meta-analysis is an effective tool to reach reliable evidence for science education by statistical interpretation of primary study findings. There is further analysis for effect size measurements in meta-analyses to comprehend more detailed findings like publication bias analysis, heterogeneity analysis, power analysis, main effect analyses, and moderator analyses.

3.7.1 Test of Heterogeneity

A central theme of meta-analysis is to compute the summary effect and disclose the patterns of effect sizes (Borenstein et al., 2009). If the primary study findings display a consistent pattern, it should be investigated why it is so, on the other hand, if the findings are very divergent, the reasons behind this pattern should also be known. One of the most informative and objective procedures is heterogeneity analysis. Heterogeneity refers to the variation in study findings. It is common that primary study findings also include moderator variables influencing treatment effects.

Moreover, there are different sources of variations in findings due to population characteristics or random errors (Borenstein et al., 2009). In this sense, it is difficult to disclose the heterogeneity between true effect sizes. Therefore, meta-analysis benefits from different heterogeneity analyses to measure the consistency or variation of treatment effects. Borenstein (2019) informs that heterogeneity is not a problem for the quality of evidence rather, it is the strength of meta-analyses. If differences between results spread out very divergently, findings may not reflect the true effect. Rather researcher needs to understand the effect of explanatory variables on the main effect. In this way, it is evaluated how much the null hypothesis tests the true effects that investigate the treatment effect. When findings of including studies have a genuine difference, meta-analysis reports the significant heterogeneity; or whether the variation in findings is compatible, meta-analysis reports homogeneity.

There are some issues in heterogeneity analyses like prediction interval, standard deviation, Q statistic, I^2 and τ^2. Each of these statistics has advantages and limitations.

3.7.1.1 Confidence Interval and Prediction Interval

Minus or plus 1.96 standard error from the mean effect size yields a confidence interval (Borenstein et al., 2009). The confidence interval implies that the true mean is located in this interval at 95%. Therefore, the confidence interval gives evidence for the true mean location but does not give evidence about the variation of effect size values across populations (Borenstein, 2019). However, the mean effect size minus or plus 1.96 standard deviations provides a prediction interval (Borenstein, 2019). A prediction interval implies that 95% of the true effect size values are in this interval for 1.96 standard deviations.

The sample's standard deviation (SD) can be derived from sample standard error by Sample SD = Sample standard error x sqrt of the degree of freedom.
Then the prediction interval for normal distribution is;

Prediction interval= Mean +/- (1.96 x Sample SD)

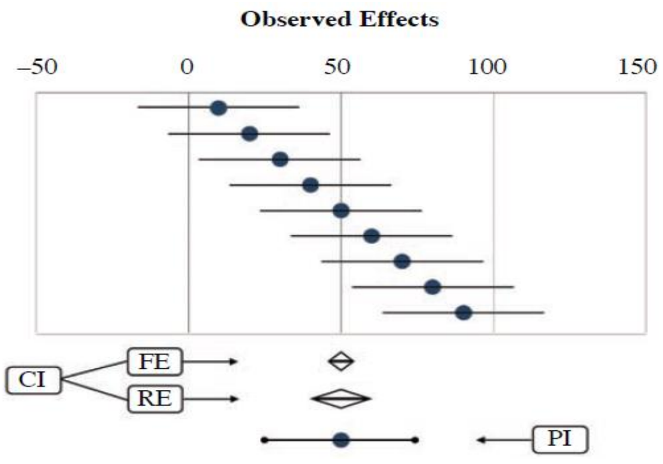


Figure 3.22 Confidence intervals and prediction interval for a fictional meta-analysis (Borenstein, 2019)

The meta-analysis studies are mainly concerned with the true effect size variations so that the prediction interval provides a more descriptive statistical interval for the dispersion of true effect size values. Therefore, reporting the prediction interval provides to inform for heterogeneity.

As depicted in Figure 3.22, CI gives knowledge about the population mean by using the sample standard error. The prediction interval (PI) informs about the samples' dispersion regardless of FE or RE models. Ninety-five percent of the population is expected to vary in the range of prediction intervals.

3.7.1.2 *Q* Statistic and Chi-squared Significance Test

Cochran's Q is one of the most common heterogeneity tests. It is the standardized sum of squares of deviation for observed effects from the mean effect (Borenstein, 2019). The deviations from the mean effect size are weighted by the inverse variance.

$$Q = \sum_{i=1}^{k} W_i (Y_i - M)^2$$

where Wi is the study weight (1/*Vi*), *Yi* is the observed effect size, M is the summary effect and k is the number of studies. True heterogeneity is estimated by excluding df, which is k-1 (where k is the number of studies). This test uses chi-squared distribution to test the null hypothesis that all studies share a common true effect size, implying that all the variance in observed effects is due to sampling error (Borenstein et al., 2009). It is used to obtain a p-value to test that there is no variation in true effects.

This test has the limitation that it is not convenient when the sample size is small. Higgins et al. (2019) state that when the sample size is small (below 20), the Q test works poorly to detect true heterogeneity, and Q statistics are not given information about the extent of heterogeneity but rather inform about whether it exists or not (Huedo-Medina et al., 2006). This is an important problem for research synthesis.

Still, this test can be used to increase the reliability of other heterogeneity tests for studies with small sample size.

3.7.1.3 The I^2 Statistic

Another heterogeneity measuring statistic on the degree of inconsistency in the effect sizes is I^2 which describes the ratio of study variance to the total variance. As a more practical definition, the I^2 value is the proportion of true variance to observed variance (Borenstein et al., 2009). The observed variance includes both true and random errors, yielding total variance. Therefore, this ratio tries to answer whether there is any unexplained variance in the true effect sizes due to random errors. The I value also represent the ratio of true and observed effects standard deviations. Meta-analyses generally take into consideration I^2 value rather than I value. CMA program uses the "Goodness of fit" term in statistical results for this value. The general formula is;

$$I^2 = (V_{True} / V_{Obs}) X\ 100$$

It can also be calculated with Q values. It is computed with the equation as;

$$I^2 = 100 \times (Q - df) / Q$$

where Q is Cochran's heterogeneity statistic and df is the degrees of freedom. The degree of I^2 values for low values is less than 25%, moderate value is about 50% and high values are more than 75% (Higgins et al., 2003). There are also different intervals in the literature for this value (Pigott, 2012). But, these intervals are also reasonable to use for this study. For example, I^2 is 0.63 means that 63% of the observed variance reflects true variance in effects, and 37% of the variance is due to random errors (Borenstein et al., 2017).

I^2 statistics have some practical applications for meta-analysis. The high I^2 means that variability stems from true heterogeneity rather than random error. The main advantage is that it is free from the sample size, study type, and outcome data

(Higgins et al., 2003). On the other hand, I^2 statistics may not be so sensitive for true variance between subgroups rather, we should give attention to the total variance (Borenstein et al., 2009). In this sense, I^2 statistics should be evaluated with Q and τ^2 statistics to observe the true variance between subgroups and between variables.

3.7.1.4 The *T* (Tau) and τ^2 Estimation Statistics

The *T (*Tau) statistics is the standard deviation of true effects. The mean plus or minus 1.96 standard deviation provides the prediction interval for 95% of true effects for the normal distribution of the sample (Borenstein, 2019**).** The square of *T (*τ^2) is the variance of the true effect also. This is also called between studies variance. The weighting process is proportional to variance, so τ^2 is a functional statistic for meta-analysis. It provides to calculate treatment effect by measuring the extent of variance on true effects which informs about heterogeneity in true effects. The random-effects model assumes that there are two variation sources between-study variance and sampling error. τ^2 represents the between-study variance rather than sampling error. In this sense, the null hypothesis is tested that $\tau^2=0$ yields only a sampling error that exists with the alternative hypothesis of Ha: $\tau^2 \neq 0$ (Pigott, 2012).

Although the Q statistics can be used to check the heterogeneity, it does not give information about the degree of heterogeneity. τ^2 gives the degree of heterogeneity even for small sample sizes different from Q statistics. The values for τ^2 are equal to (1/3)v, v, and 3v, representing respectively low, moderate, and large degrees of heterogeneity (Pigott, 2012).

On the other hand, the limitation of τ^2 is that it is not common for all effect sizes. For example, $\tau^2= 0.3$ may indicate different interpretations for different effect size indices. But, τ^2 can also be used to drive I^2 statistics by the relation; $I^2= \tau^2/(\tau^2+v)$ where v is the within-study sampling variance (Pigott, 2012).

3.7.1.5 Mistakes in Heterogeneity Interpretations

Interpreting heterogeneity is a big deal for meta-analyses because it evaluates moderators' effect on the main effect. But there are some crucial mistakes made by researchers during interpretations. Firstly, describing a statistical model to interpret analysis is vital. There are two distinct models: the fixed-effect model, which accepts the common mean effect size for each population, and the random-effects model, which accepts different mean effect sizes for each population due to their different characteristics. Some meta-analysts decide on models in terms of heterogeneity tests. They prefer the random-effects model to interpret the variation if there is significant heterogeneity. This approach is improper from the meta-analytic perspective (Borenstein, 2009). This issue is discussed frequently in the literature (Borenstein et al., 2009; Borenstein, 2019; Cooper et al., 2019).

Secondly, there are misinterpretations of statistical values of heterogeneity. For example, using the I^2 value as the index of heterogeneity is not a proper evaluation of I^2. A low I^2 represents low heterogeneity, the high value represents high heterogeneity is not a correct interpretation of I^2 value (Borenstein, 2019). This is the proportion of true variance in observed variance. Therefore, this value does not give evidence for heterogeneity alone. In this sense, reporting one index can lead to misinterpretations.

Thirdly, the confidence interval does not give evidence for heterogeneity rather prediction interval should be preferred to report to inform about the dispersion in effect sizes (Borenstein, 2019). The confidence interval is just the dispersion of the population mean, giving evidence just about the population means rather than effect sizes. Researchers may confuse these two intervals but they are pretty different.

3.7.2 Moderator Analysis

In any experimental setting, it is ordinary to remember that there are variables that affect control and experimental group scores. The actual variation in effect sizes across groups means signs of potential explanatory moderators. They are one of the most prominent factors affecting the relationship between dependent and independent variables. For practical reasons, researchers may not include the effect of moderator variables or couldn't detect the presence of explanatory moderators. Moreover, it is difficult to detect variances across studies due to artifacts (such as sampling error or measurement errors) or variance across studies due to moderator variables (Hunter & Schmidt, 2004). That is, researchers underestimate the effect of moderator variables in primary study findings.

Primary studies or traditional reviews cannot investigate moderators without further analysis. On the other hand, meta-analyses investigate the dispersion among the effect sizes obtained from primary studies. Effect sizes within subgroups are compared statistically. Using heterogeneity tests, one would infer that some moderator variables must account for this difference. This conclusion also stated that more research is needed to identify the undefined interactions (moderators) that have caused the diverging findings. Then, the researcher set different hypotheses to test the differences between true and observed variances. Analysis of variance (ANOVA) is not practical to use in meta-analysis rather an analog ANOVA based on Q-test is conducted in meta-analyses as statistical test to compare subgroups. There are two common models in meta-analysis: the fixed effect model and the random-effects model. The differences are stated in the previous parts. In this study, the random-effects model is preferred to compare moderator analysis.
In the scope of this study, there are two levels of analysis: main effect and moderator analyses (Figure 3.23). The purpose of main effect analyses is that the researcher investigates the main research questions by hypothesis testing. All studies related to intended CCS are included and synthesized in these analyses according to the determined model. Statistical evidence is provided about how effective the investigated strategy is on achievement.

In the second level, the meta-analysis also compares the mean effect for different subgroups of studies. The moderators that reflect the characteristics of subgroups are determined from the literature and the attributions of the included primary studies. The defined moderators are coded and the effect sizes are compared between subgroups with respect to these moderators. The moderator analyses are conducted with simple meta-regression and multiple meta-regression during this study. Simple meta-regression analyses include one independent (moderator) variable and one dependent variable (effect size value). During this analysis, the researcher aims to obtain a prediction function between two variables. The effect of an independent variable on a dependent variable (achievement) was investigated through the random-effects model. Each explanatory moderator allows us to explain a certain amount of variance in the total variance which is R^2.

$$R^2 = 1 - \frac{T^2_{within}}{T^2_{total}}$$

The increasing R^2 informs researchers to the more explained variance by the moderator on the overall effect. As an example for this analysis, one can investigate the effect size values for CCS in terms of different publication types. Other moderators are not included in the analysis. This provides evidence for the effectiveness of each subgroup on the main effect. However, the moderators are not isolated from each other; rather, they interact with lots of variables simultaneously. So that the simple meta-regression does not give sufficient information because it underestimates the influence of other moderators. Therefore, there is a need to examine the simultaneous models as well.

In the simultaneous analyses, we examined how much the moderators simultaneously affect the overall effect size and the correlation of moderators with each other. The ultimate aim for simultaneous analyses was to describe the combined effect of moderators that provides the best explanatory model to estimate the results of the intervention process on the dependent variable. During the simultaneous analyses, all moderators are included in analyses simultaneously to control the unique impact of each moderator. These analyses provide information on the reasons for the total variance of studies. The better the total variance for studies is explained, the more information we yielded about the distribution of effect sizes.

This approach is not a problem when we have enough studies for each subgroup to code. However, as the number of subgroups increases, the number of samples decreases, so coding more variables causes fewer studies to remain in the subgroups. For this reason, the literature and the number of samples in the subgroups should be considered during the coding process.

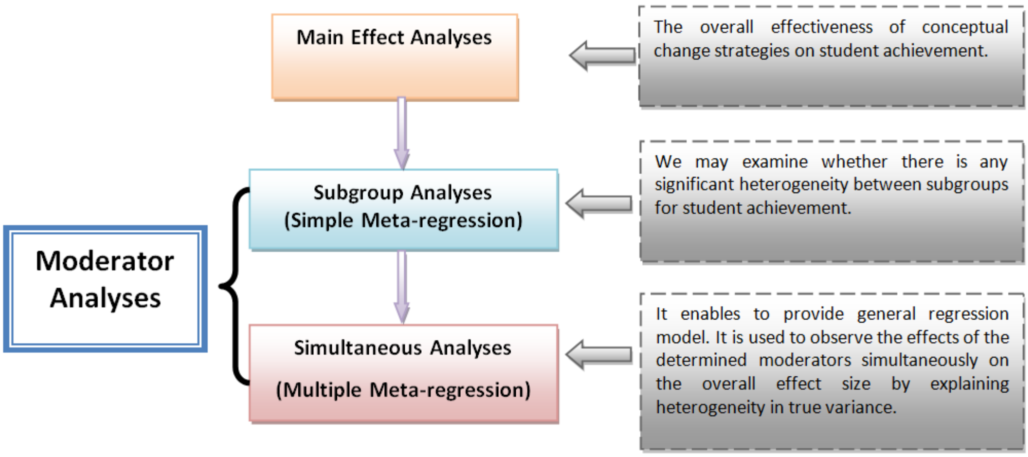


Figure 3.23 The analysis process for the main effect and moderator analyses.

3.7.3 Power Analysis

Ellis (2010) describes statistical power as “the probability that a test will correctly identify a genuine effect” p.52. Borenstein (2009) stated that power is primarily a function of precision. The precision of findings is also related to the different factors. Matt and Cook (2019) stated that total sample size, subgroup sample sizes, the type of experimental design, and study variance might be responsible for statistical power

(Figure 3.24). From a meta-analytic perspective, precision also changes with the analysis model. Under the fixed-effect model, precision is expected more for any included primary studies due to the large total sample size (Borenstein, 2009). At the same time, the confidence interval is also very narrow for primary study intervals.

$$SE_M = \sqrt{\frac{\sigma^2}{k\ x\ n}}$$

In this sense, even if the effect size is small, the power in the fixed-effect model is close to one because the power is just related to the k (the number of studies) and n (the sample size in each study) (Borenstein, 2009). The random-effects model's source of variation is more than the fixed-effect model. The between-study variance is added to the within-study variance, which decreases the precision. But power still becomes by the cumulative sample size. Thus, the power is still large with respect to primary studies.

The formulas in meta-analysis to measure power is identical to primary studies.
We will use a parameter lambda (λ) to represent an alternative true value of Z, defined as;

$$\lambda = \frac{\delta}{\sqrt{V_\delta}}$$

where δ is the true effect size and $V\delta$ is the corresponding variance. Then, power is given by:

$$\text{Power} = 1 - \Phi\ (c\alpha - \lambda) + \Phi\ (-c\alpha - \lambda)$$

where $c\alpha$ is the critical value of Z associated with significance level α, which is 1.96 for α of 0.05. Alpha should be set to take into consideration the potential impact of Type I error. Φ (x) can be calculated in EXCEL by using the NORMSDIST function (Borenstein, 2009).

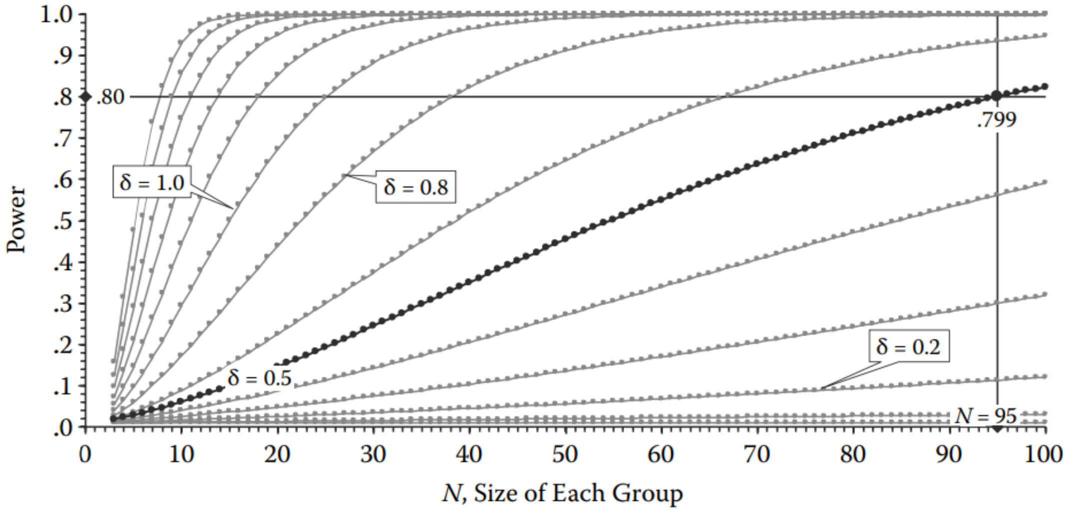


Figure 3.24 For α = .01. a horizontal power cursor has been added at power = .8. The N cursor has been moved to where the power cursor cuts the δ = 0.5 power curve and shows that N = 95 for that power and population ES (Cumming 2012, p. 336).

3.7.4 Effect Size in Meta-Analysis

Effect size is the quantitative measurement of the difference between the control group mean and treatment group means. It also informs about the magnitude of the relationship between dependent and independent variables (Borenstein and Hedges, 2019). Primary studies report the effect sizes to light on analysis, ensure more valid interpretations, adequately answer research questions, demonstrate the statistical significance, and lead researchers to come to a more clear conclusion for study results (Ferguson, 2009). Borenstein and Hedges (2019) stated the three properties of effect size value should be satisfied as “(1) it should be comparable to one another in the sense that they measure (at least approximately) the same thing, (2) researchers in the substantive area of the work represented in the synthesis should find the effect size meaningful, (3) it should be computable from the information that is likely to be reported in published research reports.” (p. 209).

Effect sizes can be used for differences between means, correlations, estimations, variations, and ratios. Therefore, many effect size estimators are used in literature to interpret the primary studies (Table 3.10). It can be grouped into three general subtypes of effect sizes raw (unstandardized), standardized, and transformed. It can be used raw mean difference if all studies employed the same scale to measure treatment effect, if there are different scales for different measurements, then we must use the standardized mean difference (Borenstein & Hedges, 2019). The nature of educational science makes it reasonable to use standardized mean differences. At the same time, Cohen's d, Glass, Δ, and Hedges g are t common standardized mean differences used in meta-analysis. In fact, Cohen's d is the most common index in primary studies. The raw mean differences are divided by pooled standard deviation of the control and treatment groups. But it has some bias for small sample sizes which yield a higher effect size value (Borenstein, 2009). This bias is removed by a correction process in Hedges' g. The Glass Δ represents the same process except that it uses the just control group SD. Meta-analysts mostly propose the Hedges' g to perform in meta-analyses due to its correction factor (Borenstein et al., 2009; Ellis, 2010; Pigott, 2012; Borestein and Hedges, 2019). Therefore, Hedges' g value was used in this study during investigations.

$$\text{Cohen's } d = \frac{M_1 - M_2}{SD_{pooled}}$$

$$\text{Glass's } \Delta = \frac{M_1 - M_2}{SD_{control}}$$

$$\text{Hedges's } g = \frac{M_1 - M_2}{SD^*_{pooled}}$$

Standardized mean differences are more practical and appropriate effect size indices than correlations and ration values. Meta-analyses also mostly report the standardized mean differences for an effect size value. For this purpose, the meta-analytic perspective proposes the mean difference indices (such as Hedges's g, Glass's Δ, BES-based RD, and Cohen's d) in primary studies to make a more valid investigation in literature. On the other hand, squared indices of r-like quantities such as r^2, ω^2, ε^2, and η^2 are not practical to use since they lose the direction of effect size (negative or positive) due to squared power (Rosnow & Rosenthal, 2003).

Cohen's *d* is one of the most common effect size indices to compare groups with continuous outcomes. The below equation enables us to compute Cohen's d for empirical studies where X_1 is the experiment group mean, X_2 is the control group's mean, S_{within} is the pooled standard deviation for independent groups.

$$d=\overline{Y_1} - \overline{Y_2} / S_{within}$$

$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 - (n_2 - 1)\, S_2^2}{n_1 + n_1 - 2}}$$

Table 3.10 Three families of effect size estimators (Rosnow & Rosenthal, 2003, p.222).

		Subtype	
Family	Raw	Standardized	Transformed
Difference	M_1-M_2 (raw difference)	Hedges' *g*	Probit d^2
	Cohen's g	Cohen's d	Logit $d^{'2}$
	Π	Glass's Δ	Cohen's b
	d', Risk difference (RD)	BESD-based RD	Cohen's q
Correlation	r_{φ} $r_{equvalence}$ $r_{contrast}$ $r_{alerting}$ $r_{effect\ size}$ r_{BESD} $r_{counternull}$	Fisher z_r	
Ratio	Relative risk (RR) Odd ratio (OR)	BESD-based RD BESD-based OD	

Note. Numbers in parentheses refer to equation numbers that define these estimators.

Another common effect size indices are Hedges's g which is calculated with the same process for cohen's d multiplied by correction factor J where;

$$j=1-\frac{3}{4df-1}$$

$$g= J\,x\,d$$

As seen from the equation, J is always below one. Therefore, the g value is always smaller than the d value due to the correction factor except for the infinite sample size. That is why the g and d values become closer when the sample size increases.

3.7.4.1 Transformation Between Standardized Effect Sizes and Conversion Effect Sizes for Other Indices

In primary studies, quantitative measurements are reported with different indices for results like odd ratios, correlations, F-test scores, t-test scores, and z-scores. Some of these studies give standardized effect sizes, and some do not. Moreover, they may not report the same standardized indices like Hedges' g, Glass Delta, or raw differences. Therefore, transformation and conversion are needed to conduct a meta-analysis with effect sizes. The Figure 3.25 expresses the possible paths for different transformations. We also reported the conversion equations for different effect size indexes in Table 3.11.

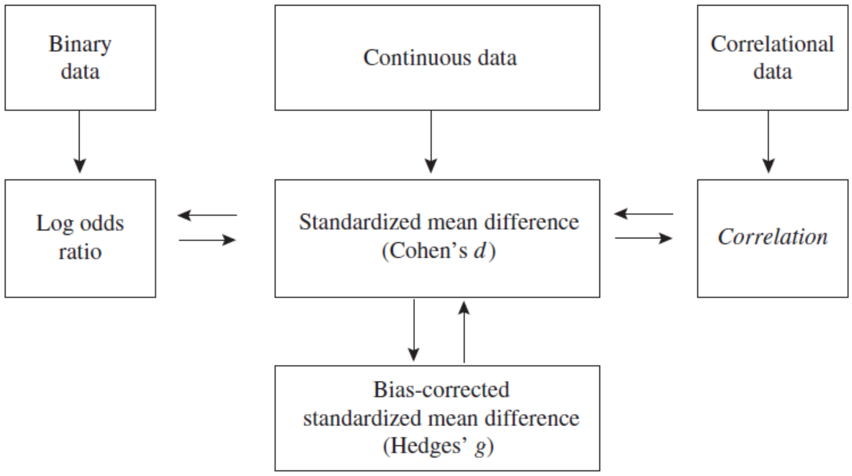


Figure 3.25 Effect size transformations between different indices (Borenstein & Hedges, 2019; p.233)

3.7 The Follow-up Process to Compute Effect Size Values

Meta-analyses use just the effect size values. Therefore, primary studies should report enough statistical value to measure ES. However, measuring ES without an adequate measurement process is not clear. In this sense, meta-analysts should inform readers about the decisions during computing ES. It is normal to face computational problems during the measurement of effect sizes. Some primary studies report multiple effect size scores within a single study (separate data is presented for moderator variables like group size, gender, or design) that should be weight or prefer adequate data. Borenstein et al. (2009) propose using descriptive statistics to calculate ES values since primary studies report the number of different ES indices. For example, the F value for ANOVA and ANCOVA yield different ES values. Thus, descriptive data provide to measure more standardized effect size values. Later, if there is no descriptive data, inferential statistics can be used. Another common issue is that some empirical studies report incomplete statistical data to compute effect size. In this situation, Lipsey and Wilson (2001) propose converting formulas to calculate mean and standard deviations in studies provided with both pre-test and post-test scores. If pretest-posttest scores are not available (only post-test scores available), the means and standard deviations of the post-test scores are used to calculate the effect size (Lazonder & Harmsen, 2016). Another problem is the small sample size bias in meta-analysis. There is a correction procedure reported by Hedges and Olkin (1985) in the literature, which proposes to use Hedges' *g* indices. For this study, the calculations in the CMA version 3 program were conducted in the below order;

1- Firstly, descriptive data was used, if it was given, inferential statistics were used

2- Studies with control groups were used, if there was no control group, it was omitted.

3- For the studies given pretest-posttest, these two were included in the effect size calculation, if only post was given, only the post-test was included. (CMA analyzes over post data std error)

4- For inferential data, firstly, scores without covariate were used like z-score, t-score, ANOVA, or MANOVA. if only ANCOVA, MANCOVA, etc., values that include covariate scores were reported, those results were used.

5- If effect size was given but no descriptive or inferential data, the reported effect size result was used.

6- In studies that give more than one effect size for the same group, the groups were considered independent and combined with the pooling process. For pooling, the effect size formula in the 11th item was used.

7- If there were two or more CC methods in the same study, the method that does not change the subgroups dramatically was removed to ensure that the study was evaluated as a unit of analysis.

8- When the misconception test and the general achievement test were given independently, only the misconception test was used to observe the CCS effect.

9- In three-tier tests and two-tiered tests, if the relevant effect size was given in the 2nd or 3rd tier, only that effect size was used.

10- When the conceptual change test and achievement test scores were given simultaneously, only the conceptual change test score was recorded since primary studies were accepted as a unit of analysis.

$$Pooled\ Effect\ Size = (W_1 * ES_1) + (W_2 * ES_2) + (W_3 * ES_3) \ / \ W_1 + W_2 + W_3$$

The calculations for ES in the CMA program were done by following the below steps, if;

a) Pre-post data is given: CMA program: continuous (means)- unmatched group pre and post data- Means, SD pre and post, N, Pre/post correlation.

b) Only post data is given: CMA program: continuous (means)- unmatched group post data- Means, SD, the sample size in each group.

c) Only Cohen d is given: CMA program: continuous (means)- unmatched group post data- Cohen's d and sample size

d) Only F is given: CMA program: continuous (means)- unmatched group pre and post-data- F for the difference between changes, N

e) Only t value is given: CMA program: continuous (means)- unmatched group post data only, sample size and t value

f) For U score: https://www.psychometrica.de/effect_size.html ES converting program used. Hedges *g* and std error values are obtained by entering the obtained Cohens'd data, the values given in item c in the CMA program.

For Eta –square:

https://www.psychometrica.de/effect_size.html ES converting program used. Hedges *g* and std error values are obtained by entering the obtained Cohen d data, the values given in item c in the CMA program.

11-The equation used for the pooling process to compute combined effect sizes:

$$Weight = 1 / (SE)^2$$

$$Pooled\ Effect\ Size = (W_1 * ES_1) + (W_2 * ES_2) + (W_3 * ES_3) \ / \ W_1 + W_2 + W_3$$

Table 3.11 The conversion equations for different effect size indexes table

Indices	Equations	Additional Info	References
d for a single group	$d = (M - \mu)/s$	s: sample SD	Cumming (2012) p.286
d for Independent groups	$d = \bar{Y}_1 - \bar{Y}_2 / S_{within}$	$S = \sqrt{\frac{(n_1 - 1)S_1^2 - (n_2 - 1) S_2^2}{n_1 + n_1 - 2}}$	Borenstein and Hedges (2019) p.211
d for matched groups or Pre-Post Scores	$d = \frac{M_{diff}}{S_{av}}$	$S_{av} = \sqrt{\frac{s_{pre}^2 - s_{post}^2}{2}}$	Cumming (2012) p.291
Converting from d to g	$g = J\,x\,d$	$j = 1 - \frac{3}{4df - 1}$	Borenstein and Hedges (2019) p.213
Converting from r to d	$d = \frac{2r}{\sqrt{1 - r^2}}$	$V_d = \frac{4V_r}{(1 - r^2)^3}$	Ellis (2010) p.16
Converting from r to Hedges' g	$g = \frac{r/\sqrt{1 - r^2}}{\sqrt{df(n_1 + n_1)/n_1 n_1}}$	n_1: Treatment group size n_2: Control group size	Durlak, (2009) p.928
Converting from χ^2 to r	$r = \sqrt{\frac{\chi^2}{n}}$	n is the sample size	Durlak, (2009) p.928
Converting from t-test to d	$d = \frac{t}{\sqrt{N}}$	Independent Samples T Test	Cumming (2012) p.287
Converting from t-test to η^2	$\eta^2 = \frac{t^2}{(t^2 + N - 1)}$	Independent Samples T Test	Ellis (2010) p.15
Converting from Fisher r to z	$z_r = 0.5 x ln(\frac{1+r}{1-r})$		Cumming (2012) p.388
Converting from z to Fisher r	$r = \frac{e^{2z_r - 1}}{e^{2z_r + 1}}$		Cumming (2012) p.388

3.7.5 Softwares Used During Statistical Analyses

Meta-analyses include some specific statistical values, visuals, and calculations to analyze findings like funnel and forest plots, FSN methods, regression methods, prediction intervals for heterogeneity analysis, and meta-regression analyses for moderator variables. The unit of analysis is primary studies, and the data used are effect size values which are different from regression analysis. The variances are also effective in providing a weighting process for ES. The defined model also measures variances like fixed or random-effects models. In this sense, some programs carry out these processes to conduct meta-analyse,s such as R, JASP, RevMan, Jamovi, Meta-Essentials, OpenMEE, OpenMeta, MetaStat, MetaGenyo, and CMA. These programs are free to use except for CMA. On the other hand, their main weakness is that there is no user-friendly process for multiple meta-regression analyses except CMA v3.0. This analysis is critical because it investigates both the combined effect of moderators on treatment effect and analyzes the individual effects of moderators by controlling the other moderators in the general model. Since the explanatory moderators are not isolated, without meta-regression analyses, it is not reliable and practiced to talk about the actual effect of moderators. So CMA 3.0 was selected to analyze research questions. Additionally, Excel, SPSS, and JASP were used for further analysis.

This software provides many useful tools. It can be purchased from the official CMA software for about $ 195 (professional version for student copy) for six months. We used this program in two parts initial analyses in the first six months plus final analyses in the last six months. Therefore, it seems to need to use at least 1-year account to finalize analyses.

With CMA 3.0 software;

- You can work on a functional spreadsheet interface
- One can figure out the treatment effect automatically
- Perform the meta-analysis easily, quickly, and accurately

- Enable to analyze graphs and plots with a single click
- Use the cumulative meta-analysis function
- Use a “Remove-One” analysis to gauge each study’s impact
- Work with multiple subgroups and outcomes within and between studies
- 8- Assess the potential impact of publication bias
- One can perform multiple regression analysis
- One can obtain VIF values
- It can be obtained that correlation matrixes for subgroups (Comprehensive Meta-Analysis website. https://www.meta- analysis.com 18.10.2021)

CHAPTER 4

RESULTS

The initial aim of this meta-analysis was to determine the overall effectiveness of CCS versus traditional strategies. At the same time, the effectiveness of three types of CCS cognitive conflict, cognitive bridging, and ontological category shift that yield from the theoretical knowledge and practical data were analyzed by conducting regression analyses. Secondly, the relationship between the effectiveness of the CCS and moderator variables that reflect the attributions of the population were investigated. In order to accomplish the determined goals, a systematic and comprehensive literature search was undertaken for a number of databases, indexes, and journals. As a result of the searching process, 218 primary studies were gathered and investigated under the determined analysis processes.

It was tried to define the scope of this study as inclusive as possible. The language is limited to Turkish and English studies. The time limitation was not applied except for the limitation stemming from the nature of conceptual change literature. More importantly, the conceptual change approach was tried to be drawn with a broad but objective perspective. The literature search began in 2014 and lasted until October 2021. As a result of a comprehensive literature search for more than 29,000 primary studies, 218 convenient studies were included in analyses. The time interval for studies is 1989-2021. The sample distribution for CCS is 150 cognitive conflict, 30 cognitive bridging, and 9 ontological category shift.

In this sense, this chapter includes descriptive statistics, main effect analyses, simple meta-regression analysis, and simultaneous meta-regressions to analyze moderator effects on treatment effects. In main effect analyzes, publication bias and heterogeneity issues were also investigated through funnel plots, forest plots, prediction intervals and statistical methods like Rosenthal FSN, Orwin FSN, and Egger's regression test. The power analyzes were also conducted to measure the possible effect of Type I error. As a result of the analysis, it was aimed to yield a general regression model that best fit to explain the dispersion in effect size values from the overall mean effect in the population.

4.1 Descriptive Statistics

This meta-analysis included 218 mean effect size values revealed from 218 primary studies. The unit of analysis is primary studies. Only one effect size value was yielded from each primary study. If the intended study reported more than one effect size value to address the effectiveness of CCS, we combined the ES values as inversely proportional to the variance to reach a common effect value.

Table 4.1 Descriptive statistics for effects sizes in meta-analysis

Variable	*k*	Mean	Variance	Std. deviation	Minimum	Median	Maximum
Hedges' *g*	218	1.13	0.50	0.71	-0.32	0.98	3.94

The arithmetic average of all effect sizes is 1.13, ranging from -0.32 to 3.94. Table 4.1 illustrates the distribution of effect size values. The histogram (Figure 4.1) shows the theoretical normal curve for included effect size values in this study. The data in this histogram included 272 effect size values ranging from -0.51 to 4.73.

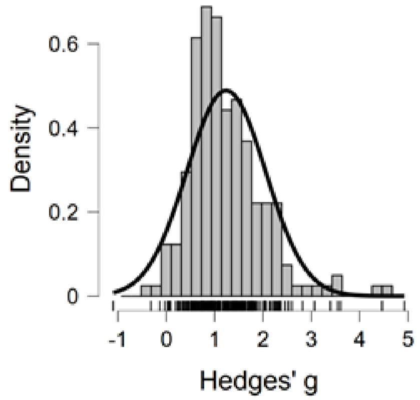


Figure 4.1 Histogram and theoretical asses fit for 272 effect sizes included in the meta-analysis.

10 of 272 effect sizes included in the meta-analysis are negative, while 262 effect sizes are positive values. Most studies' effect size values were distributed between 0.0 and 2.0. The distribution of the studies is very close to the normal distribution.

As a result of theoretical knowledge and practical data obtained from the sample gathering process, 33 moderator variables were defined. Some of these moderators were not practical to code, like treatment fidelity and study quality; some of the moderators were impossible to code due to insufficient knowledge like confusion method and medium; some moderators were not reported by researchers like gender distribution and level of internal validity threats. Therefore, we coded 23 moderators and analyzed in the scope of this study. As s result of the coding process and linearity analyses, we defined 21 categorical and 2 continuous moderators (intervention length and intervention intensity). We analyzed nine research questions under meta-regression processes with simple and simultaneous analyses. The simultaneous model that best fit to define effects size distribution was yielded by using theoretical and practical knowledge. The ultimate aim is to provide a simultaneous model that best explains the source of true heterogeneity between findings.

4.2 Main Effect Analysis

4.2.1 Overall effectiveness of Conceptual Change Strategies

What is the overall effectiveness of the conceptual change strategies on science achievement compared to traditional teaching methods?

4.2.1.1 Unit of Analysis

We accepted each primary study as unit of analysis. Every primary study has only one mean effect size value for calculations. We also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We investigated 218 primary studies under five comprehensive study characteristics (Table 4.2). Each primary study population is unique in social sciences. So that, the literature proposes that the random-effects model is more reasonable for analyzing moderator variables for educational study contexts. That is why the random-effects model was used for further analyses.

Table 4.2 Descriptive summary of the primary studies under each categorical variable for the random-effects model.

Study Characteristics	Variables /Subgroups	*k*	%	*g*	95% CI
Publication Characteristics	Publication Type				
	Journal Article	133	61	1.06	[0.95, 1.17]
	Doctoral Dissertation	40	18	1.35	[1.12, 1.58]
	Master Thesis	35	16	1.01	[0.82, 1.19]
	Proceeding	10	5	0.98	[0.68, 1.29]
	Publication Year *				
	1989-2000	14	6	0.56	[0.35, 0.78]
	2001-2005	48	22	1.08	[0.88, 1.28]
	2006-2010	73	33	1.30	[1.15, 1.45]
	2011-2015	69	32	1.09	[0.95, 1.24]
	2016-2020	14	6	0.75	[0.49, 1.00]
Sample Characteristics	Region*				
	Africa	8	4	0.88	[0.48, 1.28]
	America	28	13	0.68	[0.50, 0.86]
	Asia	23	11	0.79	[0.56, 1.02]
	Europe	12	6	0.66	[0.31, 1.01]
	Turkey	147	67	1.28	[1.18, 1.38]
	Sample Size				
	16-46	45	21	1.09	[0.92, 1,26]
	47-56	45	21	1.07	[0.90, 1,24]
	57-72	45	21	1.23	[1.04, 1,42]
	73-100	41	19	1.18	[0.97, 1,38]
	102-396	42	19	0.93	[0.73, 1,12]
	Class Size				
	8-22	50	23	0.95	[0.76, 1.14]
	23-26	45	21	1.05	[0.88, 1.21]
	27-30	41	19	1.26	[1.08, 1.45]
	31-38	40	18	1.14	[0.97, 1.31]
	39-87	42	19	1.13	[0.91, 1.35]
	Education Level*				
	Elementary	13	6	0.96	[0.64, 1.29]
	Middle	50	23	1.03	[0.90, 1.16]
	High school	101	46	1.24	[1.10, 1.39]
	Undergraduate	54	25	0.93	[0.76, 1.12]
	School Location				
	Rural	22	10	0.96	[0.72, 1.20]
	Urban	157	72	1.13	[1.03, 1.24]
	Unspecified **	39	18	1.08	[0.90, 1.36]
	School Type				
	Private	11	5	1.03	[0.73, 1.33]
	Public	165	76	1.11	[1.01, 1.21]
	Unspecified**	42	19	1.10	[0.89, 1.31]

k: Number of studies; g: Hedges' g value; CI: Confidence Interval
**The moderators that report significant heterogeneity*

*** These subgroups have not been included in heterogeneity analyses.*

Table 4.2 (Continued)

Study Characteristics	Variables /Subgroups	*k*	%	*g*	95% CI
Design Characteristics	Experimental Design*				
	Poor	23	11	0.84	[0.63, 1.06]
	Quasi	162	74	1.23	[1.12, 1.33]
	True	33	15	0.64	[0.49, 0.79]
	Sampling Method				
	Nonrandom Sampling	196	90	1.09	[1.00, 1.18]
	Random Sampling	22	10	1.24	[0.95, 1.52]
	Researcher Effect				
	Not teacher	108	50	1.20	[1.07, 1.33]
	One of the teachers	12	6	0.99	[0.65, 1.32]
	Only teacher	48	22	1.14	[0.97, 1.30]
	Unspecified**	50	23	0.88	[0.72, 1.05]
	Teacher Effect				
	Different teachers	44	20	1.12	[0.94, 1.30]
	Same teacher	134	61	1.18	[1.07, 1.30]
	Unspecified**	40	18	0.81	[0.64, 0.97]
	Treatment Verification				
	Unstated	107	49	1.06	[0.96, 1.21]
	Stated	111	51	1.13	[1.00, 1.25]
	Teacher Training*				
	Unstated	111	51	0.92	[0.81, 1.03]
	Stated	107	49	1.29	[1.16, 1.42]
Intervention Characteristics	Type of CCS				
	Cognitive Bridging	30	14	1.06	[0.84, 1.28]
	Cognitive Conflict	150	69	1.10	[1.10, 1.21]
	OntologicalCategory Shift	9	4	0.88	[0.50, 1.26]
	Unspecified**	29	13	1.21	[0.95, 1.47]
	Material*				
	Computer-based	32	15	0.87	[0.70, 1.12]
	Hands-on	69	32	1.23	[1.07, 1.39]
	Text-based	117	54	1.09	[0.97, 1.21]
	Subject Domain*				
	Biology	42	19	0.82	[0.64, 0.99]
	Chemistry	85	39	1.37	[1.23, 1.51]
	Physics	91	42	0.98	[0.86, 1.10]
	Intervention length*	144	66	-	-
	Intervention Intensity	142	65	-	-

k: Number of studies; g: Hedges' g value; CI: Confidence Interval
**The moderators that report significant heterogeneity*
*** These subgroups have not been included in heterogeneity analyses.*

Table 4.2 (Continued)

Study Characteristics	Variables /Subgroups	*k*	%	*g*	95% CI
Measurement Characteristics	Instrument Type*				
	Adapted test	23	11	1.01	[0.76,1.26]
	Preexisting test	40	18	0.88	[0.68, 1.09]
	Researcher developed test	155	71	1.17	[1.07, 1.28]
	Question Type*				
	Mix**	73	33	1.09	[0.94, 1.23]
	Objective	113	52	1.21	[1.09, 1.33]
	Open-ended	32	15	0.75	[0.55, 0.94]
	Number of Tiers				
	1	164	75	1.13	[1.03, 1.23]
	2	31	14	0.91	[0.69, 1.12]
	3	15	7	1.13	[0.84, 1.42]
	Mix**	8	4	1.18	[0.58, 1.78]
	Type of Outcome Measuring				
	Conceptual Change	192	88	1.12	[1.03, 1.22]
	General Achievement	16	7	0.87	[0.60, 1.14]
	Mix**	10	5	1.04	[0.70, 1.37]
	Overall	218	100	1.10	[1.01, 1.19]

k: Number of studies; g: Hedges' g value; CI: Confidence Interval
**The moderators that report significant heterogeneity*
*** These subgroups have not been included in heterogeneity analyses.*

4.2.1.2 Publication Bias

The publication types such as journal articles, doctoral dissertations, master theses and conference papers stated in Table 4.3 inform us about the publication distribution.

Table 4.3 The publication types used in this meta-analysis study for research question one.

Study Characteristics	Variables /Subgroups	*k*	*g*	95% CI	Heterogeneity			
					p	T^2	I^2	R^2
Publication Characteristics	Publication Type				0.09	0.34	84.57	0.02
	Journal Article	133	1.06	[0.95, 1.17]				
	Doctoral Dissertation	40	1.35	[1.12, 1.58]				
	Master Thesis	35	1.01	[0.82, 1.19]				
	Proceeding	10	0.98	[0.68, 1.29]				

k:Number of studies; *g:* Hedges' g; CI:Confidence interval

Publication bias (file drawer problem) mainly stems from over-tending to publish studies with significant results. Therefore, it is expected to reach journal articles that have a higher effect size value than unpublished doctoral dissertations, master thesis, and conference proceedings. Results of this meta-analysis resting on heterogeneity analysis revealed that (Table 4.3), there is no significant evidence to report publication bias in the scope of 218 studies. We observed that subgroups' heterogeneity is non-significant ($p > .05$). This value is derived from the random-effects model, and each of the primary studies was accepted as the unit of analysis (yielded just one observed effect size value for each study). On the other hand, doctoral dissertations have a significantly higher effect size value than other types when compared to other types.

There are some visual and statistical methods like funnel plot, trim and fill method, Rosenthal and Orwin's FSN methods, regression and rank tests, and to check the publication bias. In Figure 4.2, we constructed a funnel plot for the random-effects model to observe the distribution of primary studies and the scope. The plot's shape implies that the studies' precision increases with decreasing standard error. In this sense, the upper part of the funnel plot should include more precise studies, and the bottom part should include less precise studies. The empty diamond indicates the mean effect size value. For this study, the shape of the inverted funnel represented an asymmetric distribution of studies. This visual tool implied missing studies at the left part of the funnel. This critique can be more evidence-based by the trim and fill method. This method provides an adjustment to remedy the overall mean effect by estimating potentially missing studies. The fundamental assumption is that the mean of the extreme effect values should equal the mean value. That is, these extreme values should be placed both on the left and right sides of the plot. There are three different models in the trim and fill method to estimate the missing studies: fixed-fixed, fixed-random, and random-random models (Duval and Tweede, 2000a).

Among these methods, the related simulations imply that the fixed-random yield the most conservative results for meta-analysis (Peters et al., 2007; Vevea et al., 2019). That is why simulation studies for three models propose researchers use the fixed-random model in addition to a random-random model to estimate the missing studies more precisely (Peters et al., 2007; Schwarzer et al., 2009; Shi & Lin, 2019; Vevea et al., 2019).

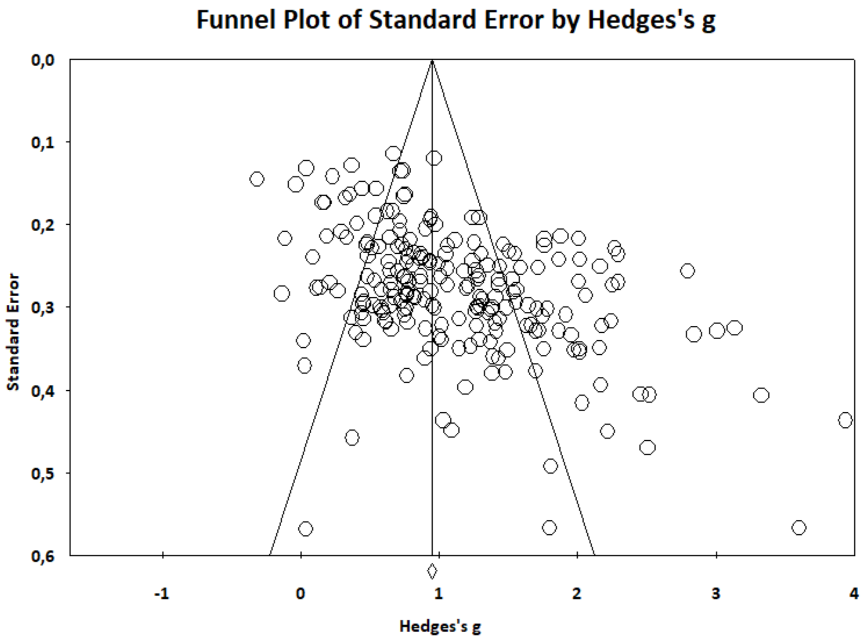


Figure 4.2 Funnel plot of all studies included in the meta-analysis based on the random-effects model without any adjustment.

For the random-random model (Figure 4.3), 24 missing studies were estimated at the right side of the mean to adjust the overall effect. The adjusted mean increased from 1.10 to 1.23 (Table 4.4).

Table 4.4 Adjusted mean values after Duval and Tweedie's trim and fill method for random-random and fixed-random models

	Model	*k*	Studies Trimmed	Direction of Mean	Hedges' *g*	Lower Limit	Upper Limit
Observed	Random- Random	218	-	Right	1.10	1.01	1.19
Adjusted	Random- Random	242	24	Right	1.23	1.13	1.33
Observed	Random- Random	218	-	Left	1.10	1.01	1.19
Adjusted	Random- Random	218	-	Left	1.10	1.01	1.19
Observed	Fixed-Random	218	-	Right	1.10	1.01	1.19
Adjusted	Fixed-Random	218	-	Right	1.10	1.01	1.19
Observed	Fixed-Random	218	-	Left	1.10	1.01	1.19
Adjusted	Fixed-Random	298	70	Left	0.71	0.61	0.81

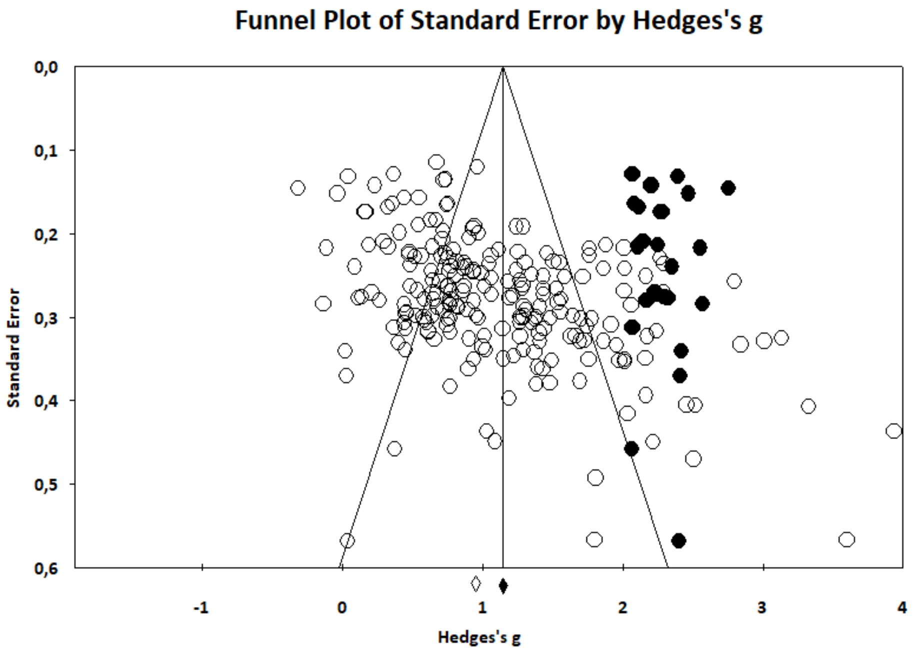


Figure 4.3 Adjusted funnel plot of all studies included in the meta-analysis based on random-random model.

This model tries to provide symmetric distribution of all studies around the mean by adding mirror values. It yields a more conservative result for the fixed-fixed model but yields less conservative for the fixed-random model. Therefore, the literature proposes to report the fixed-random model additionally. In this model, researcher creates a mirror image around the funnel plot center. For the scope of this study, 70 missing studies at the left of the mean were estimated. This is a severe number for this sample. As a limitation of this model, Vevea et al. (2019) stated that the fixed-random model gives over-conservative results in some situations. For example, if the studies with small effect sizes get together, the SD of the mean decrease. This led to an increase in the number of extreme studies which depend on the SD of the sample. Therefore, the missing study number increase dramatically. The sample for this meta-analysis also reflects this sensitivity, as can be observed in Figure 4.4. Therefore, the fixed-random model yields an over-conservative result.

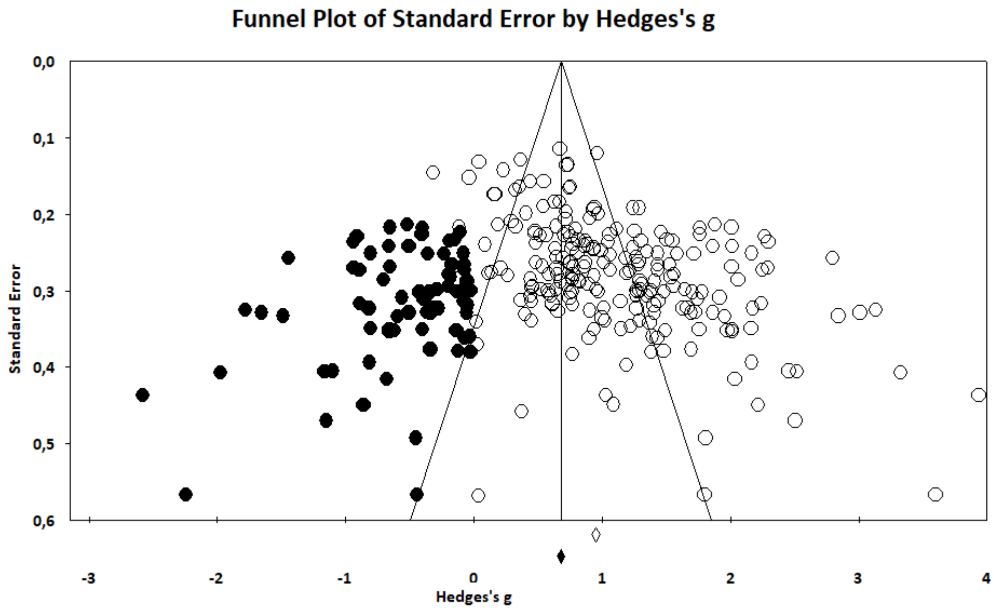


Figure 4.4 Adjusted funnel plot of all studies included in the meta-analysis based on the fixed-random model.

Alternatively, RoBMA provides a comprehensive publication bias analysis. This analysis compares and averages 36 models, including PEESE, selection models, and tail and cut-off points to inform readers more comprehensively. This analysis took about 60 minutes for 218 studies. This quantitative correction analysis reported that the adjusted mean value should be 0.93 which is smaller than the unadjusted value (1.10) but still a large effect size (Table 4.5).

Table 4.5 Output of the RoBMA for publication bias.

			95% CI	
	Mean	Median	Lower	Upper
Effect size (δ)	0.93	0.94	0.68	1.09
Heterogeneity (τ)	0.57	0.57	0.47	0.72

The above analysis informs that the unadjusted overall effect value is 1.10 and the adjusted value is 0.93 which informs around 15% of change in the overall mean. Therefore, we may infer that there is a publication bias effect in a small effect direction but publication bias does not change the overall mean dramatically.

In addition to the trim and fill method and RoBMA, we need to examine other methods that inform about the study's publication bias sensitivity like Rosenthal's FSN. This method checks the impact of missing studies by using a critical p-value and questioning the number of missing studies by averaging the null value to yield a nonsignificant result. Mullen et al. (2001) proposed the cut-off number (5k+10) should exceed the FSN value to evaluate the results more precisely, where N is the number of studies obtained from Rosenthal's, or Orwin's FSN and k is the number of studies, including meta-analysis.

The below table (Table 4.6) was generated for a .05 alpha level for two tails. According to Rosenthal's Fail-Safe N method, there should be 7380 more studies with a mean of 0.000 to bring a p-value more than .05, which is a trivial result. This is a relatively significant number for 218 studies, which implies that this result implies robustness for the publication bias.

Table 4.6 Rosenthal's FSN for all studies

Z-value for observed studies	59.01
p-value for observed studies	0.000
Alpha	0.05
Tails	2.00
Z for alpha	1.96
Number of observed studies	218
Fail-safe N	7380

Another sensitivity method is Orwin's FSN which is based on practical significance for effect size values. According to Orwin's Fail-Safe N method, there should be 1853 more studies with a mean of 0.000 to bring Hedges' *g* value to less than 0.100, which is a trivial result (Table 4.7). It can be said that it can be ignored for publication bias in this meta-analysis for Orwin's FSN.

Table 4.7 Orwin's FSN for all studies

Hedges'g in observed studies	0.95
Criteria for trivial Hedges' g	0.1
Mean Hedges' g in missing studies	0.00
Z for alpha	1.96
Number of observed studies	218
Fail-safe N	1853

Another method is Egger's Regression Test which is the most sensitive method for publication bias and is based on the symmetry of the funnel plot (Table 4.8). This test implies an asymmetric plot for this meta-analysis, implying a publication bias. We rejected the null hypothesis that " *there is no funnel plot asymmetry (β0 = 0)".* In this sense, we may state that this study has a publication bias, but we can ignore it.

Table 4.8 Egger's regression test results for all studies

Intercept	4.33
Standard error	0.53
95% lower limit (2-tailed)	3.29
95% lower limit (2-tailed)	5.37
t value	8.21
Df	216
p-value (2-tailed)	<.001

4.2.1.3 Overall Mean Effect Size and Corresponding Statistical Test

Null Hypothesis: Ho: $\delta1 = \delta2$

The mean of all observed effect sizes for studies investigating the effect of CCS on science achievement is equal to the mean effect sizes which are obtained from studies using traditional methods.

This meta-analysis used a random-effects model to observe the effect of CCS compared to traditional methods. The overall effect size is 1.10 with a confidence interval of 95% of 1.01 and 1.19, which is large effect size. The null hypothesis was rejected at the alpha level of 0.05 ($p<.001$), indicating that the mean of all true effect sizes for CCS is significantly different from the mean of traditional methods (Table 4.9).

Table 4.9 Overall effect size results and corresponding statistical test for research question one.

	Effect Size				95% CI		Statistical test	
Model	*k*	*g*	SE	Variance	Low. Limit	Up. Limit	*z*	*p*
Fixed	218	0.95	0.02	0.00	0.92	0.98	56.07	<.001
Random	218	1.10	0.04	0.00	1.01	1.19	24.78	<.001

k:Number of studies; *g:* Hedges' g; SE: Standard error; CI:Confidence interval

4.2.1.4 Power Analysis

One of the advantages of meta-analysis is yielding high power due to the large number of sample sizes used in primary studies. This power is calculated by collecting each study power. Therefore, it is expected to yield very high power for meta-analysis studies. For this meta-analysis; the variance for the effect size (Hedges'g) of 1.102 is 0.002 based on the random-effects model, so the parameter λ is:

$$\lambda = \frac{\delta}{\sqrt{V\delta}} = \frac{1.102}{\sqrt{0.002}} = 24.64$$

where δ is the true effect size, and V_δ its variance.
Then, power is calculated with an alpha level of .05 as:

$$\text{Power} = 1 - \Phi\,(c_\alpha - \lambda) + \Phi\,(-c_\alpha - \lambda)$$

In Excel,

Power = 1-NORMSDIST(1.96- 24.64)+ NORMSDIST (-1.96 - 24.64) ≈1.000

where c is the critical value of Z associated with a significance level (thus, for α=0.05, c_α= 1.96) and Φ is the normal distribution function which returns a one-tailed probability from the standardized normal distribution. Due to the large number of sample sizes, the power of the test is very high with respect to single studies. The power of 1.000 implies that the Type II error probability is almost zero. It means " fail to detect a real treatment effect."

$$B = 1 - Power = 0$$

In other words, the power to detect an effect size of 1.102 is almost 1.

4.2.1.5 Heterogeneity Analysis

The heterogeneity in this chapter is the heterogeneity between true effect sizes rather than observed effect sizes. Since observed variation includes both true variation and random error, it is not practical to measure heterogeneity between observed effect sizes. There are some statistical procedures to answer the question about heterogeneity as;

Is there evidence of heterogeneity in true effect sizes?
What is the variance of the true effects?
What are the substantive implications of this heterogeneity?
What proportion of the observed heterogeneity is real?

The above questions were investigated under the statistical values sequentially as the Q statistic (a measure of weighted squared deviations), the between-studies variance (T^2), and the ratio of true variance to the total variance (I^2). The first statistical method is Q statistics, based on the chi-squared distribution. We investigate the question "Is there evidence of heterogeneity in true effect sizes?". The hypothesis is that " All studies share a common effect size " is tested so it is reported a p-value for any observed value of Q for degrees of freedom $k - 1$. Shortly, alpha is set at .05, with a p-value less than alpha indicating to reject of the null hypothesis (Table 4.10, $p<.05$). It was concluded that the studies do not share a common effect size. There is significant heterogeneity between true effect size values.

Table 4.10 Heterogeneity test for research question one

Q-value	df (Q)	p-value	I-squared	Tau Squared	Standard Error	Variance	Tau
1428	217	<.001	84.80	0.35	0.05	0.00	0.59

The second question is "What is the variance of the true effects?". Since it is not possible to measure true effect sizes, we estimate true effect variance by observed effect size variance. Therefore, it was estimated by T^2 which is the between-study variance. If we had an infinitely large sample of studies, then observed variance and true variance become the same (T^2=1). In this meta-analysis, the T^2=0.35 means that 35% of variance stems from a variation of the true effects (between-study variance). Tao refers to the estimation of the actual standard deviation, which is 0.59.

Finally, I^2 value provides to investigate the question that, "What proportion of the observed variance reflects real differences in effect size?" In this study, 84.8% of observed variance reflects the true variance. This result implies that there is pretty much heterogeneity in findings, and the possible source of heterogeneity should be investigated through further analyses. In the scope of this study, this heterogeneity will be investigated through moderator analyses.

4.2.2 Overall Effectiveness of Cognitive Conflict

What is the overall effectiveness of cognitive conflict on science achievement compared to traditional teaching methods?

4.2.2.1 Unit of Analysis

We accepted each primary study as a unit of analysis. Every primary study has only one mean effect size value for calculations. We also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 150 primary studies. The literature also proposes that the random-effects model is more reasonable for analyzing moderator variables for educational study contexts. That is why the random-effects model was used to analyze the heterogeneity.

4.2.2.2 Publication Bias

In Table 4.11, it is stated that articles have a lower effect size value than doctoral dissertations, master thesis, and conference papers. We do not reject the hypothesis that there is no difference between articles and other types. This is an unexpected result of literature and publication bias. This result implies a decreasing possibility for publication bias. More specifically, doctoral dissertations have the largest effect size value. This is very similar to the result of the first question.

Table 4.11 The number of studies and effect sizes in different publication types and corresponding point estimates for research question two.

Variables/ Subgroups	k	%	g	95% CI		Heterogeneity			
				Lower Limit	Upper Limit	p	T^2	I^2	R^2
Publication Type						0.08	0.34	84.57	0.02
Journal Article	91	65	1.03	0.93	1.18				
Doctoral Dissertation	33	20	1.39	1.15	1.63				
Master Thesis	20	12	0.99	0.72	1.26				
Proceeding	6	3	0.95	0.54	1.65				
Overall	150	100	1.09	0.99	1.19				

k:Number of studies; %: Percent; *g:* Hedges' g; CI:Confidence interval

The forest plot (Appendix E) illustrates this meta-analysis's distribution and precision. Other visual and statistical methods to investigate publication bias are funnel plot, trim and fill method, Rosenthal and Orwin's FSN regression, and rank tests. In Figure 4.5, the funnel plot is constructed for the random-effects model to observe the distribution of primary studies and the scope. For this study, the shape of the inverse funnel represents an asymmetric distribution of studies for the second research question. This visual tool implies missing studies at the left part of the funnel. This idea can be more evidence-based by the trim and fill method. This method provides an adjusted overall mean effect by estimating potentially missing studies (Table 4.12).

Table 4.12 Adjusted mean values for random-random and fixed-random models

Result	Model	k	Studies Trimmed	Direction for mean	g	Lower Limit	Upper Limit
Observed	Random- Random	150	-	Right	1.10	1.00	1.21
Adjusted	Random- Random	163	13	Right	1.21	1.09	1.33
Observed	Random- Random	150	-	Left	1.10	1.00	1.21
Adjusted	Random- Random	150	-	Left	1.10	1.00	1.21
Observed	Fixed-Random	150	-	Right	1.10	1.00	1.21
Adjusted	Fixed-Random	150	-	Right	1.10	1.00	1.21
Observed	Fixed-Random	150	-	Left	1.10	1.00	1.21
Adjusted	Fixed-Random	150	51	Left	0.69	0.58	0.81

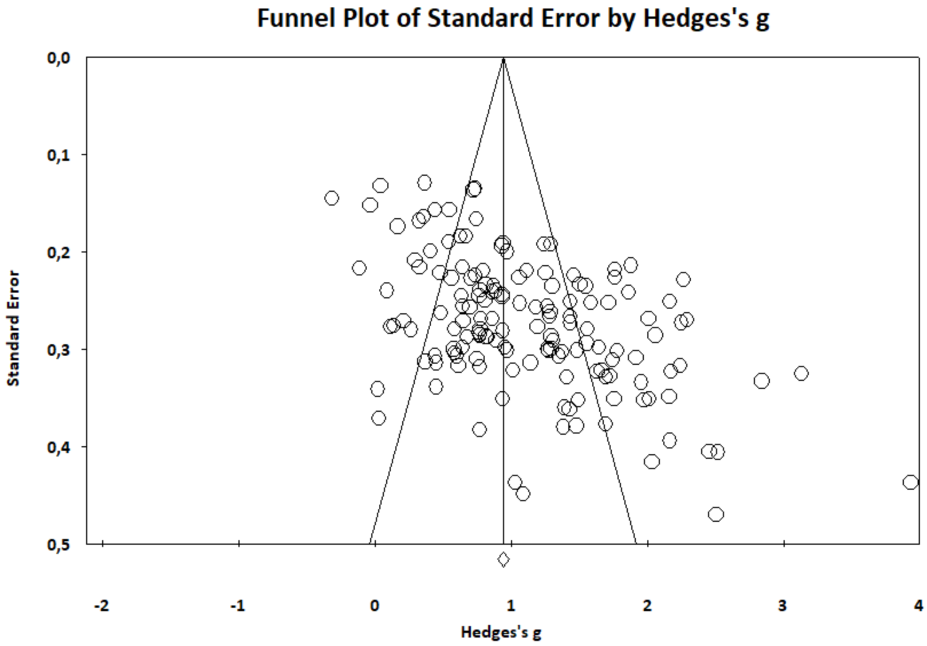


Figure 4.5 Funnel plot of the studies included in the sample of the second research question based on the random-effects model

For the random-random model (Figure 4.6), 13 missing studies were estimated at the right side of the mean to adjust the overall effect. The adjusted mean increase was from 1.103 to 1.210 (Table 4.12). This model provides a symmetric distribution of all studies around the mean by adding mirror values. It yields more conservative results for fixed-fixed models but yields less conservative for fixed-random models. Therefore, the literature proposes to report the fixed-random model additionally.

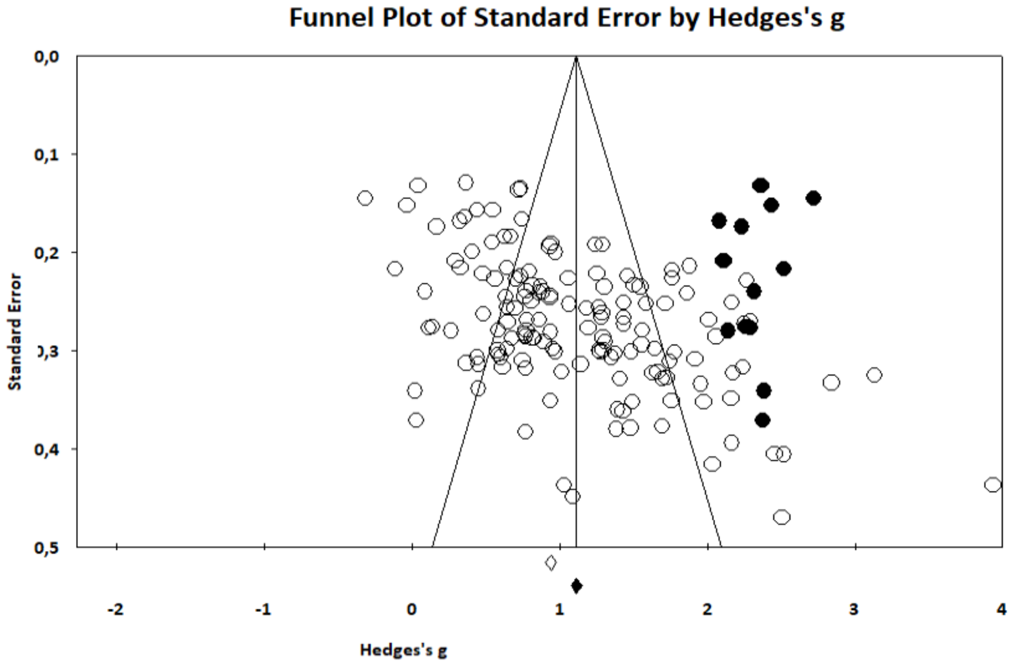


Figure 4.6 Funnel plot of the studies included in the sample of the second research question based on the random effects model

During the fixed-random model, researcher should create a mirror image about the center of the funnel plot for the fixed-effect method. But, meta-analysis should be done by the random-effects method to estimate the adjusted overall mean effect. For the scope of this study, 51 missing studies at the left of the mean were estimated. This is a very serious number for this sample. As a limitation of this model, Vevea et al. (2019) stated that the fixed-random model gives over-conservative results in some situations. For example, if the studies with small effect sizes get together SD of the mean decrease. This leads to an increase in the number of extreme studies that depend on the SD of the sample. Therefore, the missing study number increases dramatically. The sample for the cognitive conflict strategies also reflects this sensitivity, as shown in Figure 4.7. Therefore, the fixed-random model yields over-conservative results.

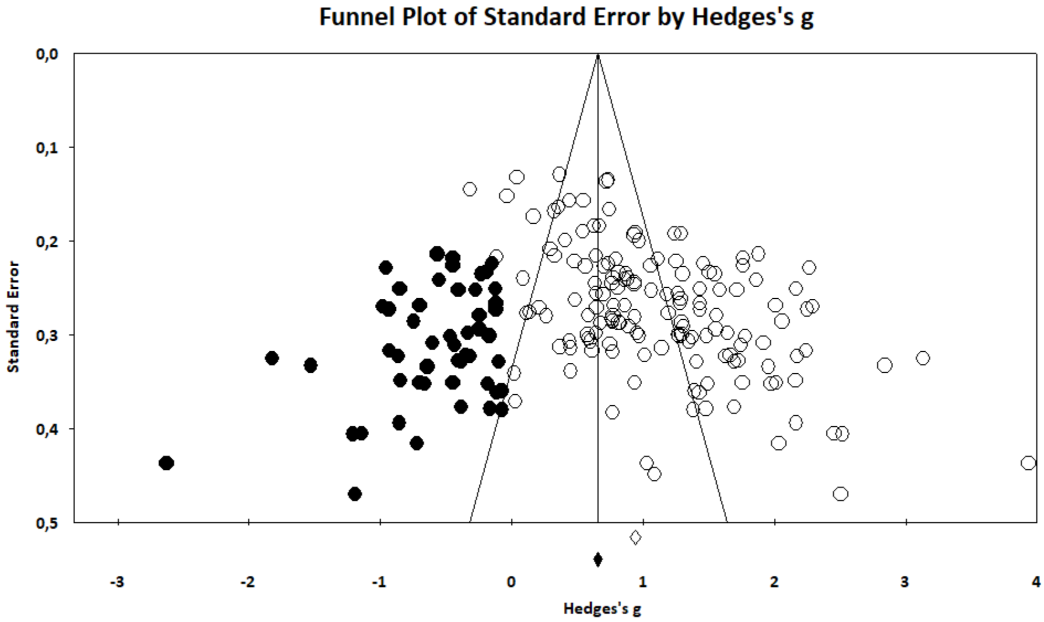


Figure 4.7 Funnel plot of the studies included in the sample of the second research question based on fixed effects models

RoBMA is a very comprehensive analysis of publication bias. This analysis compared and averaged 36 models, including PET, selection models, and tail and cut-off points. Therefore, this analysis informs researchers most reliably and comprehensively. This analysis takes about 45 minutes, which is comprehensive and heavy content for computers. This quantitative correction analysis informs that the adjusted mean value should be 0.94, which is smaller than the unadjusted value (1.10) but still a large effect size (Table 4.13)

Table 4.13 Robust BMA model-averaged estimates

			95% CI	
	Mean	Median	Lower	Upper
Effect size (δ)	0.94	0.97	0.57	1.11
Heterogeneity (τ)	0.54	0.53	0.45	0.69

The above analysis informs that the unadjusted overall effect value is 1.10 and the adjusted value is 0.94, which informs around 16% of the change in the overall mean. Therefore, we may infer that there is a publication bias effect in a small effect direction, but this correction does not change the overall mean dramatically. When the funnel plot is examined in detail, it is seen that the asymmetry is generally because studies with larger sample sizes have lower effect size values in the analysis. In this sense, we need to check other sensitivity methods. According to Rosenthal's Fail-Safe N method, there should be 5728 more studies with a mean of 0.000 to bring a p-value less than 0.05, which is a trivial result. It can be said that there is no publication bias in this meta-analysis concerning Rosenthal's FSN. The cut-off number of Mullen et al. (2001) is 745 for this meta-analysis result from the (5N+10) formula. The below table (Table 4.14) is constructed with respect to a 0.05 alpha level with a two-tail.

Table 4.14 Rosenthal's FSN for all studies

Z-value for observed studies	49.55
p-value for observed studies	0.00
Alpha	0.05
Tails	2.00
Z for alpha	1.96
Number of observed studies	150
Fail-safe N	5728

Another method for publication bias is Orwin's FSN which is based on practical significance for effect size values. According to Orwin's Fail-Safe N method, there should be 1262 more studies with a mean 0.000 to bring Hedges' g value less than 0.10, which is a trivial result (Table 4.15). This result implies that this meta-analysis can neglect publication bias regarding Orwin's FSN.

Table 4.15 Orwin’s FSN for all studies

Hedges’g in observed studies	0.94
Criterion for trivial Hedges’ g	0.1
Mean Hedges’ g in missing studies	0.00
Number of observed studies	150
Fail-safe N	1262

Although there is some asymmetric funnel plot for publication bias, we can conclude that publication bias has not so effective on treatment effect for the sample in meta-analysis with respect to FSN methods and visual analysis for funnel plot.

Another method is Egger’s Regression Test, one of the most sensitive methods for publication bias based on the symmetry of the funnel plot. This test implies an asymmetric plot for this meta-analysis, which implies a publication bias. We reject the null hypothesis that “there is no funnel plot asymmetry (β_0 = 0)”. In this sense, we may state that there is a moderate publication bias for the scope of Egger's regression analysis (Table 4.16). Therefore, there should be a more comprehensive evaluation to disclose the existence of publication bias.

Table 4.16 Egger’s regression test results for all studies

Intercept	5.28
Standard error	0.66
95% lower limit (2-tailed)	3.98
95% lower limit (2-tailed)	6.60
t value	8.00
Df	148
p-value (2-tailed)	<.001

4.2.2.3 Overall Mean Effect Size and Corresponding Statistical Test

Null Hypothesis: Ho: δ1 = δ2

The mean of all observed effect sizes for studies investigating the effect of Cognitive Conflict Methods on science achievement is equal to the mean effect sizes obtained from studies using traditional methods.

The random-effects model was used in this meta-analysis to observe the effect of CCS compared to traditional teaching methods. The overall effect size is 1.103 with a confidence interval 95% of 1.00 and 1.21, which is large effect size. The null hypothesis related to research question one was rejected at the alpha level of .05 (p<.001), indicating that the mean of all true effect sizes for cognitive conflict significantly differs from traditional teaching methods (Table 4.17).

Table 4.17 Overall effect size details for research question two.

Model	*k*	*g*	*SE*	Var.	95% CI		Statistical value	
					Lower Limit	Upper Limit	*z*	*p*
Random	150	1.10	0.05	0.03	1.00	1.21	20.48	<.001

k:Number of studies; *g:* Hedges' g; SE: Standard error; Var: Variance; CI:Confidence interval

4.2.2.4 Power Analysis

For this meta-analysis; the variance for the effect size (Hedges *'g*) is 0.03 based on the random-effects model, so the parameter λ is:

$$\lambda = \frac{\delta}{\sqrt{V}} = \frac{1{,}103}{\sqrt{0{,}03}} = 6.368$$

Then, power is calculated with an alpha level of 0.05 as:

$$\text{Power} = 1 - \Phi\,(c_\alpha - \lambda) + \Phi\,(-c_\alpha - \lambda) = 1$$

In Excel,

Power = 1-NORMSDIST(1.96- 6.368)+ NORMSDIST (-1.96 - 6.368) =1.000

where c is the critical value of Z associated with a significance level (thus, for α=0.05, c_α= 1.96) and Φ is the normal distribution function which returns a one-tailed probability from the standardized normal distribution. Due to a large number of sample sizes, the power of the test is very high with respect to single studies. The power of 1.000 implies that the Type II error probability is almost zero. It means " fail to detect a real treatment effect."

$$\text{B} = 1\text{- Power} = 0.$$

In other words, the power to detect an effect size of 1.103 is almost 1.

4.2.2.5 Heterogeneity Analysis

The heterogeneity in this chapter is the heterogeneity between true effect sizes rather than observed effect sizes. Since observed variation includes both true variation and random error, it is not practical to measure heterogeneity between observed effect sizes. There are some statistical procedures to answer the question about heterogeneity as;

Is there evidence of heterogeneity in true effect sizes?
What is the variance of the true effects?
What are the substantive implications of this heterogeneity?
What proportion of the observed heterogeneity is real?

The above questions were investigated under the statistical values sequentially as the *Q* statistic (a measure of weighted squared deviations), the between-studies variance (T^2), and the ratio of true heterogeneity to total observed variation (I^2). The first statistical method is *Q* statistics, based on the chi-squared distribution. We investigate the question, "Is there evidence of heterogeneity in true effect sizes?". The hypothesis is that " All studies share a common effect size " is tested so it is reported a p-value for any observed value of *Q* for degrees of freedom k–1. Shortly, it is set alpha at 0.05, with a p-value less than alpha, indicating to reject of the null hypothesis (Table 4.18, $p<.05$). It was concluded that the studies do not share a common effect size. There is significant heterogeneity between true effect size values.

Table 4.18 Heterogeneity test for research question two

Q-value	df (Q)	*P*	I^2	T^2	*SE*	Variance	*T*
1017	149	<.001	85.35	0.36	0.06	<.01	0.60

The second question is, "What is the variance of the true effects? " Since it is not possible to measure true effect sizes, we estimate true effect variance by observed effect size variance. Therefore, it was estimated by T^2 which is the between study variance. if we had an infinitely large sample of studies than observed variance and true variance become the same (T^2=1). In this meta-analysis the T^2=0.36 means that about 36% of variance stems from a variation of the true effect (between-study variance). Tao refers to the estimation of the actual standard deviation which is 0.56.

Finally, I^2 value provides to investigate the question, “What proportion of the observed variance reflects real differences in effect size?”. In this study, 85.35% of observed variance reflects the true variance. This result implies that there is pretty much heterogeneity in findings and the possible source of heterogeneity should be investigated through further analyses. In the scope of this study, this heterogeneity will be investigated through moderator analyses.

4.2.3 Overall Effectiveness of Cognitive Bridging

What is the overall effectiveness of cognitive bridging on science achievement compared to traditional teaching methods?

4.2.3.1 Unit of Analysis

We accepted each primary study as a unit of analysis. Every primary study has only one mean effect size value for calculations. We also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. Thirty primary studies were included to examine the fifth research question. The literature also proposes that the random-effects model is more reasonable for analyzing moderator variables for educational study contexts. That is why the random-effects model was used to analyze the heterogeneity.

4.2.3.2 Publication Bias

In Table 4.19, it is stated that articles have lower effect size value than doctoral dissertations, master thesis, and conference papers. We do not reject the hypothesis that there is no mean difference between articles and other types in the heterogeneity analyses. This is the unexpected result with respect to literature and publication bias. This result implies a decreasing possibility for publication bias. More specifically, doctoral dissertations have the largest effect size value. This is very similar to the result of the first question. On the other hand, the number of students in three subgroups is smaller than the five value. Therefore, heterogeneity analysis cannot provide reliable evidence due to an insufficient number of samples for each group. We should check the other methods.

Table 4.19 The number of studies, effect sizes in different publication types, and corresponding point estimates for research question three.

Variable	k	%	g	95% CI		Heterogeneity			R^2
				Lower Limit	Upper Limit	p	T^2	I^2	
Publication Type						0.43	0.31	81.9	<.01
Journal Article	22	73	1.13	0.86	1.40				
Doctoral Dissertation	2	7	1.36	0.09	2.63				
Master Thesis	4	13	0.64	0.36	0.92				
Proceeding	2	7	0.93	0.53	1.33				
Overall	30	100	0.91	0.73	1.08				

k:Number of studies; %: Percent; g: Hedges' g; CI:Confidence interval

We can also observe the effect size distribution concerning the precision, which is inversely proportional to standard error by forest plots. Appendix F illustrates the forest plots for studies in this meta-analysis for the third research question.

Other visual and statistical methods are provided to investigate publication bias, such as funnel plot, trim and fill method, Rosenthal and Orwin's FSN methods, regression and rank tests. In Figure 4.8, the funnel plot is constructed for the random-effects model to observe the distribution of primary studies and the scope. For this study, the shape of the inverse funnel represents an asymmetric distribution of studies for the second research question. This visual tool implies that there is an increasing heterogeneity in the below part of the funnel. But it is not fit for inverse funnel shape, so there are missing studies. It can be more evidence-based by the trim and fill method. This method adjusts the overall mean effect by estimating potentially missing studies.

Table 4.20 Adjusted mean values for random-random and fixed-random models

	Model	k	Studies Trimmed	Direction for mean	Hedges' g	Lower Limit	Upper Limit
Observed	Random- Random	30	-	Right	1.063	0.843	1.283
Adjusted	Random- Random	32	2	Right	1.122	0.900	1.344
Observed	Random- Random	30	-	Left	1.063	0.843	1.283
Adjusted	Random- Random	30	-	Left	1.063	0.843	1.283
Observed	Fixed-Random	30	-	Right	1.063	0.843	1.283
Adjusted	Fixed-Random	30	-	Right	1.063	0.843	1.283
Observed	Fixed-Random	30	-	Left	1.063	0.843	1.283
Adjusted	Fixed-Random	30	-	Left	1.063	0.843	1.283

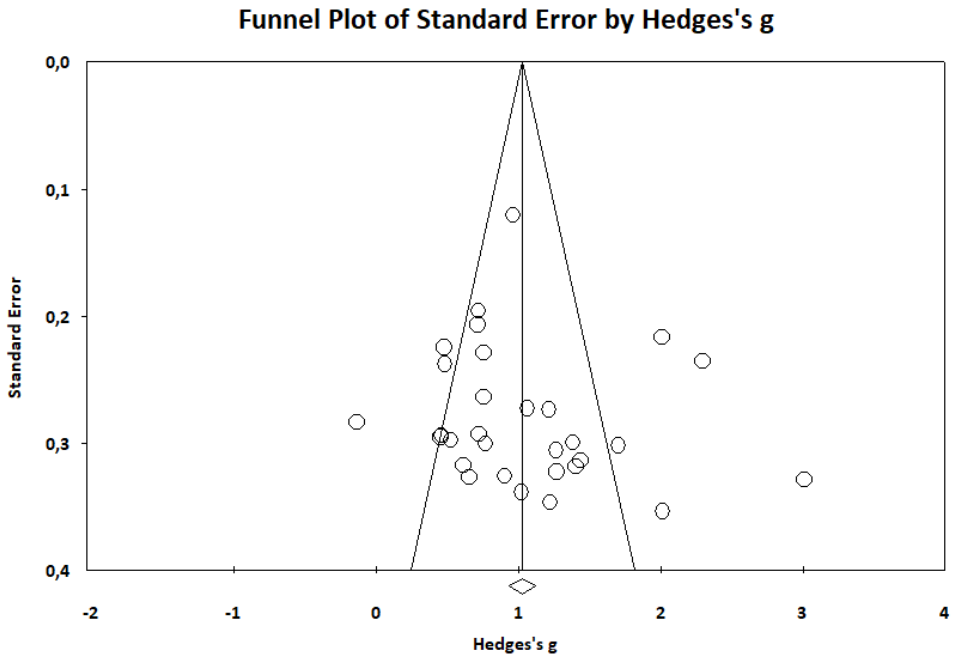


Figure 4.8 Funnel plot of the studies included in the sample of third research question based on random-effects model

For the random-random model (Figure 4.9), two missing studies were estimated at the right side of the mean to adjust the overall effect. The adjusted mean increased from 1.063 to 1.122 (Table 4.20). This model tries to provide symmetric distribution of all studies around the mean by adding mirror values. It yields a more conservative result for the fixed-fixed model but yields less conservative for the fixed-random model. Therefore, the literature proposes to report the fixed-random model additionally.

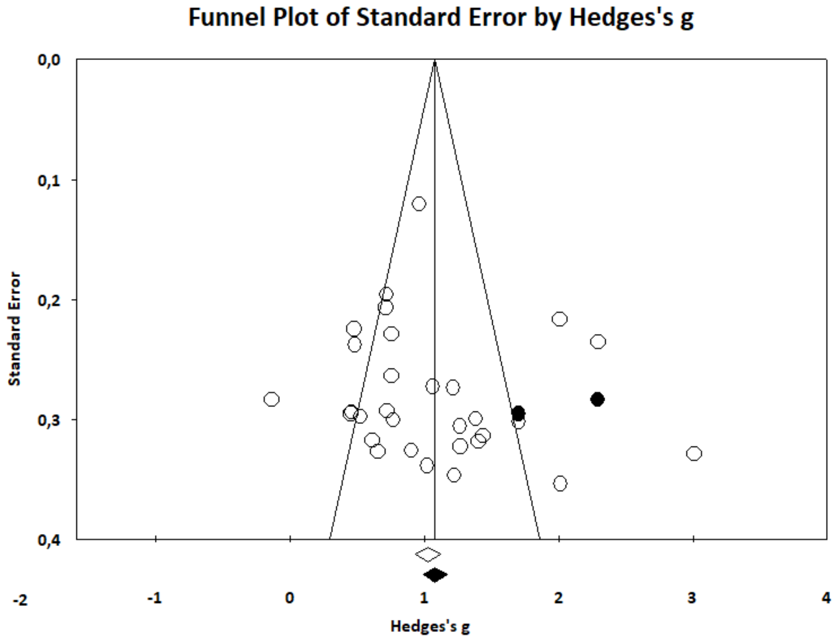


Figure 4.9 Funnel plot for the third research question based on random-random model

During the fixed-random model, the researcher should create a mirror image of the center of the funnel plot concerning the fixed-effect method. But, meta-analysis should be done by the random-effects method to estimate the adjusted overall mean effect. For the scope of this study, no missing studies at the left or right of the mean were estimated (Figure 4.10). This result implies that there can be three additional missing studies at the right of the mean for the random-random model.

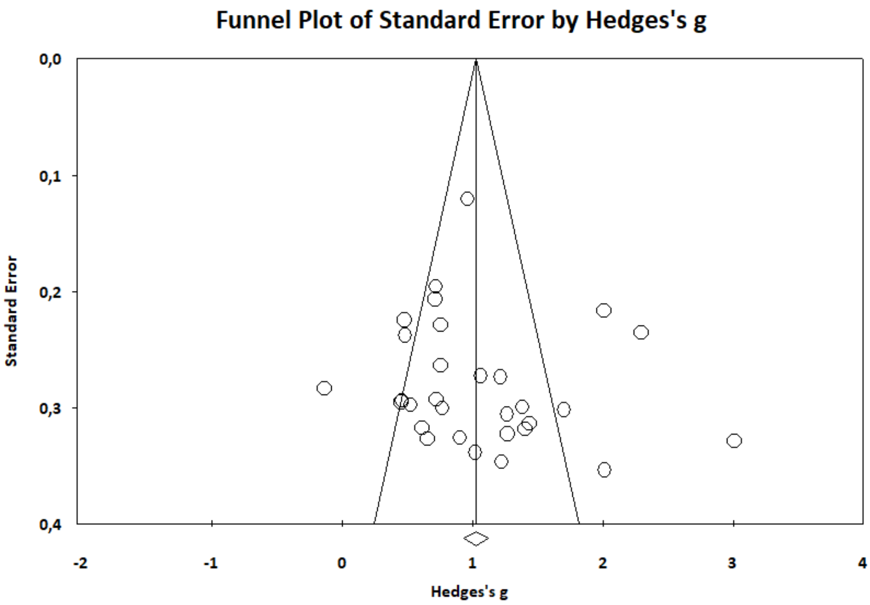


Figure 4.10 Funnel plot of the studies included in the sample of the third research question based on the fixed-random model.

RoBMA is a comprehensive analysis of publication bias problems. This analysis compared and averaged 36 different models including PET, selection models, different tail and cut-off points. Therefore, this analysis informs researchers most reliably and comprehensively. This analysis takes about 45 minutes, which is comprehensive and heavy content for computers. This quantitative correction analysis informs that the adjusted mean value should be 0.87 which is smaller than the unadjusted value (1.06) but still a large effect size (Table 4.21).

Table 4.21 Robust BMA model averaged estimates

			95% CI	
	Mean	Median	Lower	Upper
Effect size (g)	0.87	0.93	0.00	1.17
Heterogeneity	0.53	0.50	0.35	0.92

The above analysis informs that the unadjusted overall effect value is 1.06 and the adjusted value is 0.87 which informs around 13% of change in the overall mean.

Therefore, we may infer that there is a publication bias effect in a small effect direction but this correction does not change the overall mean dramatically. When the funnel plot is examined in detail, it is seen that the asymmetry is generally since studies with larger sample sizes have lower effect size values in the analysis. In this sense, we need to check other methods. According to Rosenthal's Fail-Safe N method, there should be 3570 more studies with a mean 0.000 to bring p-value less than 0.05 which is a trivial result (Table 4.22). It can be said that the publication bias effect in this meta-analysis can be neglected for Rosenthal's FSN. The cut-off number of Mullen et al. (2001) is 145 for this meta-analysis result from (5N+10) formula.

Table 4.22 Rosenthal's FSN for all studies included in meta-analysis

Z-value for observed studies	21.46
p-value for observed studies	0.000
Alpha	0.05
Tails	2.00
Z for alpha	1.96
Number of observed studies	30
Fail-safe N	3570

Another method for publication bias is Orwin's FSN which is based on practical significance for effect size values. This method provides a more conservative result for publication bias. According to Orwin's Fail-Safe N method, there should be 280 more studies with a mean 0.000 to bring Hedges' g value to less than 0.100 which is a trivial result (Table 4.23). It can be said that the publication bias effect in this meta-analysis can be neglected with respect to Orwin's FSN.

Table 4.23 Orwin's FSN for all studies included in meta-analysis

Hedges'g in observed studies	1.03
Criterian for trivial Hedges' g	0.1
Mean Hedges' g in missing studies	0.00
Number of observed studies	30
Fail-safe N	280

Another method is Egger's Regression Test which is the most conservative method for publication bias which is based on the symmetry for funnel plot. This test also supported the idea that the funnel plot for this meta-analysis is symmetric which implies that there is publication bias (Table 4.24). The hypothesis is that " We do not reject the null hypothesis that " *there is no funnel plot asymmetry* ($\beta_0 = 0$)".

Table 4.24 Egger's regression test results for all studies included in the meta-analysis**.**

Intercept	1.10
Standard error	1.59
95% lower limit (2-tailed)	-2.17
95% lower limit (2-tailed)	437
t value	0.69
df	28
p value (2-tailed)	0.496

The result of the four methods inform that there is no significant publication bias affecting on treatment effect for this meat-analysis study. Although there are some visual evidence for publication bias, we can conclude that there is no significant publication bias for the sample in meta-analysis with respect to statistical and visual methods.

4.2.3.3 Overall Mean Effect Size and Corresponding Statistical Test

Null Hypothesis: Ho: $\delta 1 = \delta 2$

The mean of all observed effect sizes for studies investigating the effect of cognitive bridging on science achievement is equal to the mean effect sizes which are obtained from studies using traditional methods.

This meta-analysis used a random-effects model to observe the effect of cognitive bridging compared to traditional teaching methods. The overall effect size is 1.06 with a confidence interval of 95% of 0.84 and 1.28 which is a large effect size (Table 4.25). The null hypothesis related to research question one is rejected at the alpha level of .05 (p<.001), indicating that the mean of all true effect sizes for Cognitive Bridging Methods is significantly different from traditional teaching methods

Table 4.25 Overall effect size details and corresponding statistical test for research question three.

Model	*k*	*g*	SE	Variance	95% CI		Statistical test	
					Lower Limit	Upper Limit	*z*	*p*
Random	30	1.06	0.11	0.01	0.84	1.28	9.48	<.001

k:Number of studies; *g:* Hedges' g; SE: Standard error; CI:Confidence interval

4.2.3.4 Power Analysis

One of the advantages of meta-analysis studies is yielding high power due to large number of the sample size used in primary studies. This power is calculated by collecting each study power. Therefore, it is expected to yield very high power for meta-analysis studies. For this meta-analysis; the variance for the effect size (Hedges'*g)* of 1.063 is 0.015 based on random-effects model, so the parameter λ is:

$$\lambda = \frac{\delta}{\sqrt{V}} = \frac{1.063}{\sqrt{0.015}} = 8.679$$

where δ is the true effect size, and V_δ its variance.

Then, power is calculated with an alpha level of 0.05 as:

$$\text{Power} = 1 - \Phi\,(c_\alpha - \lambda) + \Phi\,(-c_\alpha - \lambda)$$

In Excel,

Power = 1-NORMSDIST(1.96- 8.679)+ NORMSDIST (-1.96 - 8.679) =0.98

where c is the critical value of Z associated with a significance level (thus, for α=0.05, c_α= 1.96) and Φ is the normal distribution function which returns a one-tailed probability from the standardized normal distribution. Due to a large sample size, the power of the test is very high with respect to single studies. The power of

1.000 implies that the Type II error probability is almost zero. It means " fail to detect a real treatment effect"

$$B = 1\text{- Power} = 0.01.$$

In other words, the power to detect an effect size of 1.063 is 98 %.

4.2.3.5 Heterogeneity Analysis

The heterogeneity in this chapter is the heterogeneity between true effect sizes rather than observed effect sizes. Since observed variation includes both true variation and random error, it is not practical to measure heterogeneity between observed effect sizes. There are some statistical procedures to answer the question about heterogeneity as;

Is there evidence of heterogeneity in true effect sizes?
What is the variance of the true effects?
What are the substantive implications of this heterogeneity?
What proportion of the observed heterogeneity is real?

The above questions were investigated under the statistical values sequentially as the Q statistic (a measure of weighted squared deviations), the between-studies variance (T^2) and the ratio of true heterogeneity to total observed variation (I^2). The first statistical method is Q statistics which is based on the chi-squared distribution. We investigate the question, "Is there evidence of heterogeneity in true effect sizes?". The hypothesis is that " All studies share a common effect size " is tested so it is reported a p-value for any observed value of Q for degrees of freedom k–1. Shortly, it is set alpha at .05, with a p-value less than alpha indicate to reject the null hypothesis (Table 4.26, $p<.05$). It was concluded that the studies do not share a common effect size. There is significant heterogeneity between true effect size values.

Table 4.26 Heterogeneity test for research question three

Heterogeneity							
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
156	28	<.001	81.39	0.30	0.11	0.01	0.46

The second question is **"**What is the variance of the true effects? ". Since it is not possible to measure true effect sizes, we estimate true effect variance by observed effect size variance. Therefore, it was estimated by T^2 which is the between-study variance. if we had an infinitely large sample of studies than observed varience and true variance become the same (T^2=1). In this meta-analysis, the T^2=0.298 means that 29.8% of variance stems from variation of the true effect (between-study variance). Tao refers to the estimation of actual standard deviation which is 0.46.

Finally, I^2 value provides to investigate the question that "What proportion of the observed variance reflects real differences in effect size?". In this study 81.39% of observed variance reflects the true variance. This result implies that there is pretty much heterogeneity for findings and the possible source of heterogeneity should be investigated through further analyses.In the scope of this study, this heterogeneity will be investigated through moderator analyses.

4.2.4 Overall Effectiveness of Ontological Category Shift

What is the effectiveness of the ontological category shift on science achievement compared to traditional teaching methods?

4.2.3.1 Unit of Analysis

We accepted each primary study as a unit of analysis. Every primary study has only one mean effect size value for calculations. We also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. Nine primary studies were included in the analyses. The literature proposes that the random-effects model is more reasonable for analyzing moderator variables for educational study contexts. That is why the random-effects model was used to analyze the heterogeneity.

4.2.4.2 Publication Bias

In Table 4.27, it is stated that articles have a higher effect size value than doctoral dissertations but lower than a master thesis. From the heterogeneity analysis, we do not reject the hypothesis that no significant difference exists between published and unpublished study effect size values. This is an unexpected result of literature and publication bias. This result implies the decreasing possibility of publication bias. More specifically, master thesis has the largest effect size value. This is similar to the result of the first, second and third research questions. On the other hand, the number of students in two subgroups is smaller than the five value. Therefore, heterogeneity analysis cannot provide reliable evidence due to an insufficient number of samples for each group. We should check the other methods.

Table 4.27 The number of studies, effect sizes in different publication types, and corresponding point estimates for research question four.

Variables//Subgroup	*k*	%	*g*	95% CI Lower Limit	95% CI Upper Limit	Heterogeneity *p*	T^2	I^2	R^2
Publication Type						0.25	0.22	72.6	<.01
Journal Article	7	78	0.92	0.48	1.35				
Doctoral Dissertation	1	11	0.04	-1.10	1.15				
Master Thesis	1	11	1.15	0.46	1.83				
Overall	9	100	0.88	0.50	1.26				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval

We can also observe the effect size distribution concerning the precision, which is inversely proportional to standard error by forest plots. Appendix G illustrates the forest plots for studies in this meta-analysis for the fourth research question. We can observe the effect size distribution regarding precision, which is inversely proportional to standard error by forest plots. Since the number of studies is few, it is unreliable to evaluate the relation between precision and effect size values by forest plot examination. Therefore, we need to control other methods.

Other visual and statistical methods provide to investigate publication bias, such as funnel plot, trim and fill method, Rosenthal and Orwin's FSN regression, and rank tests. In Figure 4.11, the funnel plot is constructed with respect to the random-effects model to observe the distribution of primary studies and the scope. For this study, the shape of the inverse funnel represents an asymmetric distribution of studies for the second research question. This visual tool implies an increasing heterogeneity in the below part of the funnel. But it is not fit for inverse funnel shape, which means that there are missing studies. It can be more evidence-based by the trim and fill method. This method provides to adjust the overall mean effect by estimating potentially missing studies.

Table 4.28 Adjusted mean values for trim and fill method for random-random and fixed-random models

	Model	*k*	Studies	Direction	Hedges' g	Lower	Upper
Observed	Random- Random	9	-	Right	0.88	0.50	1.13
Adjusted	Random- Random	9	-	Right	0.88	0.50	1.13
Observed	Random- Random	9	-	Left	0.88	0.50	1.13
Adjusted	Random- Random	9	-	Left	0.88	0.50	1.13
Observed	Fixed-Random	9	-	Right	0.88	0.50	1.13
Adjusted	Fixed-Random	9	-	Right	0.88	0.50	1.13
Observed	Fixed-Random	12	3	Left	0.58	0.20	0.96
Adjusted	Fixed-Random	9	-	Left	0.88	0.50	1.13

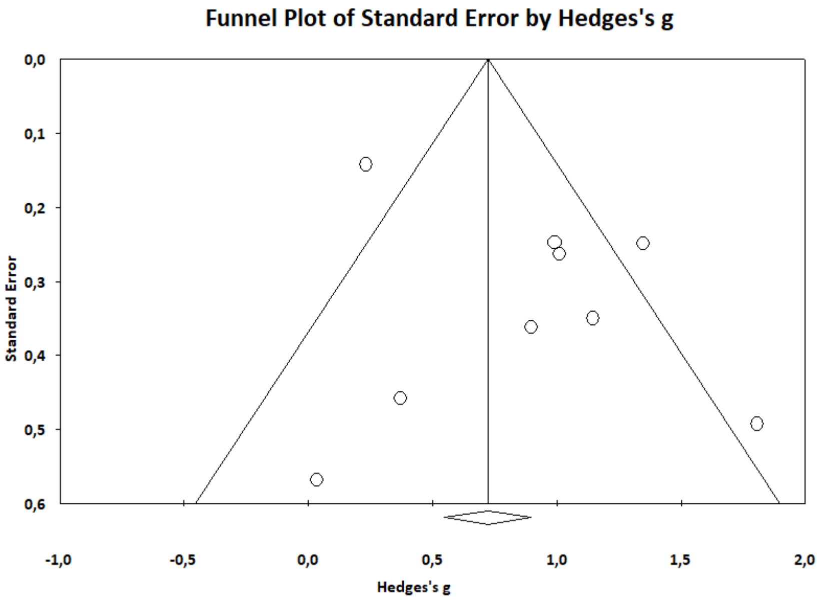


Figure 4.11 Funnel plot of the studies included in the sample of the fourth research

For the random-random model (Figure 4.12), no missing studies were estimated at the right or left side of the mean to adjust the overall effect. The adjusted mean is 0.88 (Table 4.28). This model tries to provide symmetric distribution of all studies around the mean by adding mirror values. It yields a more conservative result for the fixed-fixed model but less conservative for the fixed-random model. Therefore, the literature proposes to report the fixed-random model additionally.

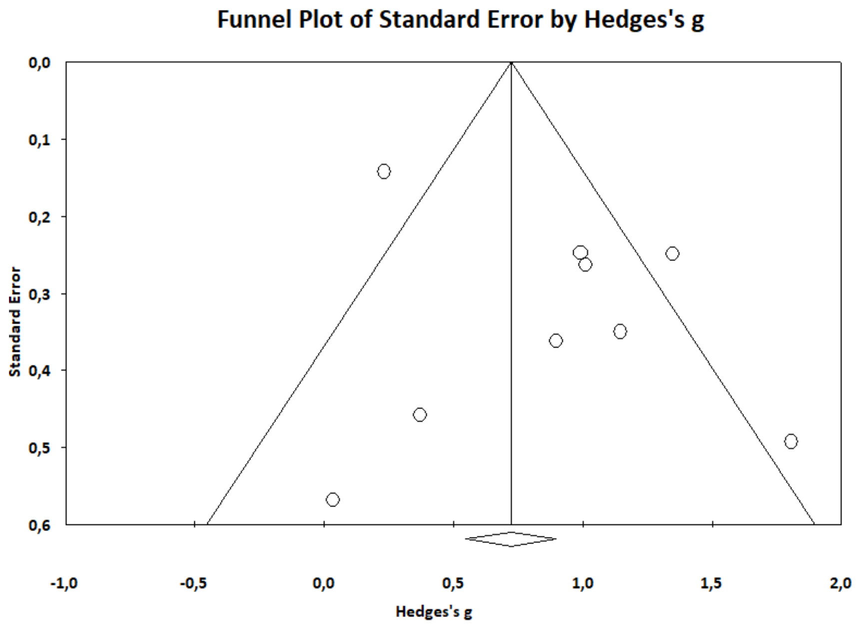


Figure 4.12 Funnel plot of the studies included in the sample of the fourth research question based on the random-random model

During the fixed-random model, researcher should create a mirror image of the center of the funnel plot for the fixed-effect method. But, meta-analysis should be done by the random-effects method to estimate the adjusted overall mean effect. For the scope of this study, three missing studies at the left of the mean were estimated (Figure 4.13). This result implies that there can be three additional missing studies at the left of the mean for the fixed-random model. The adjusted mean decreased from 0.88 to 0.50 (Table 4.28).

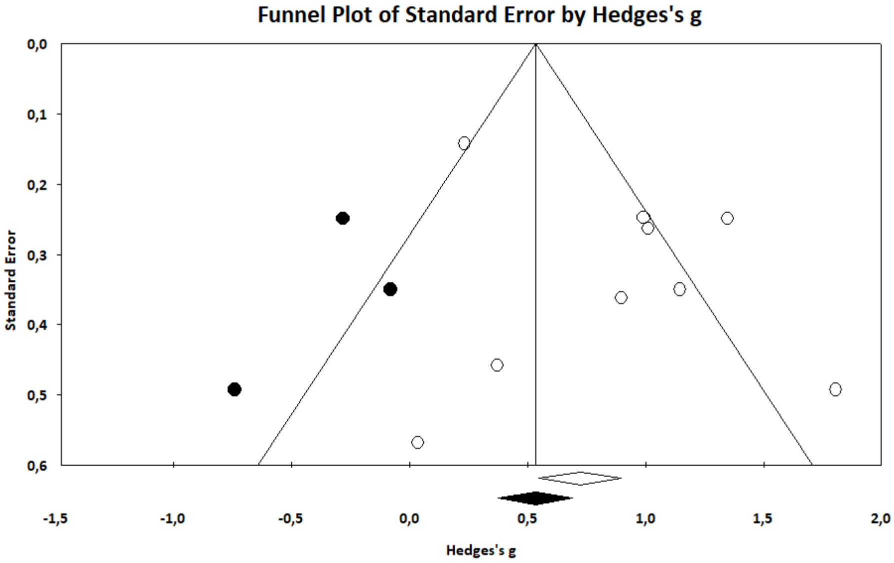


Figure 4.13 Funnel plot of the studies included in the sample of the fourth research question based on the fixed-random model

Finally, RoBMA is a very critical comprehensive analysis of the publication bias problem. This analysis compared and averaged 36 different models, including PET, selection models, and different tail and cut-off points. Therefore, this analysis informs researchers most reliable and comprehensive way. This analysis takes about 10 minutes, which is comprehensive and heavy content for computers. This quantitative correction analysis informs that the adjusted mean value should be 0.69, which is smaller than the unadjusted value (0.88) but still a medium effect size (Table 4.29).

Table 4.29 Robust BMA model-averaged estimates

			95% CI	
	Mean	Median	Lower	Upper
Effect size (δ)	0.69	0.75	0.00	1.12
Heterogeneity (τ)	0.38	0.34	0.08	0.89

When the funnel plot is examined in detail, it is seen that the asymmetry is generally since studies with larger sample sizes have lower effect size values in the analysis. In this sense, we need to check other methods. According to Rosenthal's Fail-Safe N method, there should be 103 more studies with a mean 0.000 to bring a p-value less than 0.05, which is a trivial result (Table 4.30). It can be said that there is no publication bias in this meta-analysis with respect to Rosenthal's FSN. The cut-off number of Mullen et al. (2001) is 55 for this meta-analysis result from (5N+10) formula.

Table 4.30 Rosenthal's FSN method

Z-value for observed studies	8.41
p-value for observed studies	0.00
Alpha	0.05
Tails	2.00
Z for alpha	1.96
Number of observed studies	9
Fail-safe N	157

Another method for publication bias is Orwin's FSN which is based on practical significance for effect size values. This method provides more conservative results for publication bias. According to Orwin's Fail-Safe N method, there should be 56 more studies with a mean of 0.000 to bring Hedges' *g* value less than 0.100 which is a trivial result (Table 4.31). It can be said that there is no publication bias in this meta-analysis with respect to Orwin's FSN since it is higher than the cut-off value (45).

Table 4.31 Orwin's FSN for all studies included in the meta-analysis

Hedges'g in observed studies	0.88
Criteria for trivial Hedges' g	0.1
Mean Hedges' g in missing studies	0.00
Number of observed studies	9
Fail-safe N	56

Another method is Egger's Regression Test which is the most conservative method for publication bias and is based on the symmetry of the funnel plot. This test implies a symmetric plot for this meta-analysis, which implies no publication bias (Table

4.32). The hypothesis is that we do not reject the null hypothesis that "*there is no funnel plot asymmetry (β_0 = 0)*".

Table 4.32 Egger's regression test results for all studies

Intercept	2.12
Standard error	1.45
95% lower limit (2-tailed)	-1.32
95% lower limit (2-tailed)	5.55
t value	1.46
df	7
p-value (2-tailed)	0.19

The four methods show no significant publication bias affecting the treatment effect for this meta-analysis study. Although there is some visual evidence for publication bias, we can conclude that there is no significant publication bias for the sample in meta-analysis with respect to statistical and visual methods.

4.2.4.3 Overall Mean Effect Size and Corresponding Statistical Test

Null Hypothesis: Ho: $\delta 1 = \delta 2$

The mean of all observed effect sizes for studies investigating the effect of Ontological Category Shift Methods on science achievement is equal to the mean effect sizes which are obtained from studies using traditional methods. This meta-analysis uses a random-effects model to observe the effect of CBM compared to traditional teaching methods. The overall effect size is 1.06 with a confidence interval 95% of 0.50 and 1.26 which is large effect size. The null hypothesis related to research question one is rejected at the alpha level of 0.05 (Table 4.33, $p<.001$), indicating that the mean of all true effect sizes for ontological category shift is significantly different from traditional teaching methods.

Table 4.33 Overall effect size details and corresponding statistical test for research question three.

Model	k	g	SE	Variance	95% CI Lower Limit	95% CI Upper Limit	Statistical test z	Statistical test p
Random	9	0.88	0.19	0.04	0.50	1.26	4.55	<.001

k: Number of studies; g: Hedges' g; SE: Standard error; CI: Confidence interval

4.2.4.4 Power Analysis

One of the advantages of meta-analysis studies is yielding high power due to large sample size used in primary studies. This power is calculated by collecting each study power. Therefore, it is expected to yield very high power for meta-analysis studies. For this meta-analysis; the variance for the effect size (Hedges'g) of 1.144 is 0.002 based on the random-effects model, so the parameter λ is:

$$\lambda = \frac{\delta}{\sqrt{V}} = \frac{0.877}{\sqrt{0.037}} = 4.559$$

where δ is the true effect size, and V_δ its variance.

Then, power is calculated with an alpha level of 0.05 as:

$$\text{Power} = 1 - \Phi\,(c_\alpha - \lambda) + \Phi\,(-c_\alpha - \lambda)$$

In Excel,

Power = 1-NORMSDIST(1.96- 4.56)+ NORMSDIST (-1.96 - 4.56) =0.98

where c is the critical value of Z associated with a significance level (thus, for α=0.05, c_α= 1.96) and Φ is the normal distribution function which returns a one-tailed probability from the standardized normal distribution. Due to large sample size, the power of the test is very high for single studies. The power of 1.000 implies that the Type II error probability is almost zero. It means " fail to detect a real treatment effect."

$$\text{B} = 1\text{- Power} = 0.12$$

In other words, the power to detect an effect size of 0.877 is 0.98.

4.2.4.5 Heterogeneity Analysis

The heterogeneity in this chapter is the heterogeneity between true effect sizes rather than observed effect sizes. Since observed variation includes both true variation and random error, measuring heterogeneity between observed effect sizes is not practical. There are some statistical procedures to answer the question about heterogeneity as;

Is there evidence of heterogeneity in true effect sizes?
What is the variance of the true effects?
What are the substantive implications of this heterogeneity?
What proportion of the observed heterogeneity is real?

The above questions were investigated under the statistical values sequentially as the Q statistic (a measure of weighted squared deviations), the between-studies variance (T^2), and the ratio of true heterogeneity to total observed variation (I^2). The first statistical method is Q statistics, based on the chi-squared distribution. We investigate the question, "Is there evidence of heterogeneity in true effect sizes?". The hypothesis is that " All studies share a common effect size " is tested so it is reported a p-value for any observed value of Q for degrees of freedom k – 1. Shortly, it is set alpha at 0.05, with a p-value less than alpha, indicating to reject of the null hypothesis (Table 4.34, $p<.05$). It was concluded that the studies do not share a common effect size. There is significant heterogeneity between true effect size values.

Table 4.34 Heterogeneity test for research question three

Heterogeneity							
Q-value	df (Q)	p	I^2	T^2	T	Variance	SE
29	8	<.001	72.62	0.22	0.47	0.03	0.17

The second question is, "What is the variance of the true effects? ". Since it is not possible to measure true effect sizes, we estimate true effect variance by observed effect size variance. Therefore, it was estimated by T^2 which is the between-study variance. if we had an infinitely large sample of studies, then observed variance and true variance would become the same (T^2=1). In this meta-analysis, the T^2=0.218 means that 21.8% of variance stems from a variation of the true effect (between-study variance). Tao refers to the estimation of the actual standard deviation which is 0.467. Finally, I^2 value provides to investigate the question, “What proportion of the observed variance reflects real differences in effect size?”. In this study 72.6% of observed variance reflects the true variance. This result implies that there is pretty much heterogeneity in findings and the possible source of heterogeneity should be investigated through further analyses. In the scope of this study, this heterogeneity will be investigated through moderator analyses.

4.3 Moderator Analyses

In this part, we examined 23 moderators with simple meta-regression processes. The relationship between each explanatory moderator and treatment effect was analyzed individually. Additionally, the combined impact of whole moderators was analyzed simultaneously to yield a simultaneous model that best explains the heterogeneity in effect size distribution.

4.3.1 Analyses for Types of CCS

What is the role of conceptual change strategies (cognitive bridging, cognitive conflict, or ontological category change) on the effectiveness of CCS on science achievement?

4.3.1.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 189 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.1.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:
Means of all true effect sizes of cognitive conflict, cognitive bridging, and ontological category shift strategies are equal to each other.

Table 4.35 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. Descriptively, the mean effect sizes for conflict and bridging methods are higher than ontology-based methods. But conflict and bridging methods are not different from each other. As a result, the null hypothesis was not rejected and indicated that the mean effect sizes for conceptual change strategies are not significantly different at 0.05 (p> .05).

Table 4.35 The results of heterogeneity analysis within subgroups for type of CCS.

Variable	*k*	%	*g*	SE	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	R^2	I^2
Type of CCS							0.62	0.35	0.00	84.53
Cognitive Bridging	30	12	1.06	0.11	0.84	1.28				
Cognitive Conflict	150	71	1.10	0.05	1.00	1.21				
Ontological Change Shift	9	4	0.88	0.19	0.50	1.26				
Unspecified**	29	13	1.21	0.13	0.95	1.47				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 84.53 which illustrates that 84.5% of total variance results from study variance and 15.5% from sampling error. This result also implies the existence of other moderator variables that have an impact on treatment effect.

4.3.1.3 Explained Variance by Individual Model

The individual model enables to test of the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.36, we tested the significance of the individual model. We have one covariate as CCSs. The analysis showed that Q=0.95 with df=2 and p=.622, so it implies that the predictive model probably does not explain the variance in mean effect size. Therefore, it can be said that CCSs have no impact on the treatment effect if we do not control the other confounders. This individual model does not explain any variance.

4.3.1.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. We tested whether the common effect size for all studies is the same or not. For this model, we rejected the null hypothesis that all CCSs do not have the same common effect size (Table 4.36, $p<.05$). Therefore, confounding variables should affect the common effect size except for CCSs. In this sense, it is needed to look simultaneous model including more confounders that explain the total variance in true effect sizes more.

Table 4.36 Meta-regression results on the effect of different types of conceptual change strategies

Set	Covariate	β	SE	95% Low Limit	95% Up. Limit	z	p	Heterogeneity		
								Q	df	p
Type of CCS	Intercept	1.07	0.12	0.83	1.30	8.94	<.001	0.95	2	0.622
	Conflict	0.04	0.13	-0.22	0.29	0.29	0.77			
	Ontology	-0.19	0.26	-0.69	0.32	-0.72	0.47			
Model								0.95	2	0.622
Residual								1202	186	<.001
Total								1203	188	<.001

β:Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.2 Analyses for Material

What is the role of material (computer-based, hands-on, and text-based) on the effectiveness of CCS on science achievement?

4.3.2.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.2.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of computer-based, hands-on, and text materials are equal.

Table 4.37 expresses the results of heterogeneity analyses for heterogeneity between each subgroup. There is significant heterogeneity within subgroups. As a descriptive evaluation, the mean effect sizes for hands-on materials are higher than the other materials, and the computer-based materials have the least mean effect size value. As a result, the null hypothesis was rejected, indicating that the mean effect sizes for materials are significantly different at 0.05 (p> .05).

Table 4.37 Heterogeneity analysis within subgroups for material type.

Variable	*k*	%	*g*	SE	95% CI		Heterogeneity			R^2
					Low. Limit	Up. Limit	*p*	T^2	I^2	
Material							0.05	0.34	84.50	0.02
Computer	32	16	0.87	0.11	0.70	1.12				
Hands-on	69	32	1.23	0.08	1.07	1.39				
Text-based	117	52	1.09	0.06	0.97	1.21				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval

The overall I-squared is 84.5, which illustrates that 84.5% of total variance results from study variance and 15.5% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.2.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.38, we can test the significance of the individual model. We have one covariate as material. The analysis shows that Q=6.16 with df=2 and p=.046, implying that the predictive model probably explains variance in mean effect size. Therefore, it can be said that material has an impact on treatment effect if we do not control the other confounders. This individual model only explains the total variance by 1.8%.

4.3.2.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. We tested whether the common effect size for all studies is the same or not. For this model, we rejected the null hypothesis that all materials do not have the same common effect size (Table 4.38, $p<.05$). Therefore, there should be confounding variables that affect the common effect size except the type of material. In this sense, it is needed to look simultaneous model including more confounders that explain the total variance in true effect sizes more.

Table 4.38 Simple meta-regression analysis within subgroups for the type of material.

						Heterogeneity		
Set	Covariate	β	SE	z	p	Q	df	p
	Intercept	0.88	0.12	7.60	<.001			
Medium	Hands-on	0.34	0.14	2.46	0.01	6.16	2	0.046
	Text	0.21	0.13	1.57	0.12			
Model						6.16	2	0.046
Residual						1387	215	<.001
Total						1428	217	<.001

β:Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.3 Analyses for Publication Type

What is the role of publication type (Article, Doctoral dissertation, Master thesis, Conference paper) on the effectiveness of CCS on science achievement?

4.3.3.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies to examine the fifth research question. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.3.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of computer-based, hands-on, and text materials are equal to each other.

Table 4.39 expresses the results of simple meta-regression analyses for heterogeneity between subgroups. There is no significant heterogeneity within subgroups. More specifically, the mean effect sizes for doctoral dissertations are higher than the other types. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for publication types are not significantly different at 0.05 (p> .05).

Table 4.39 Heterogeneity analysis within subgroups for publication type

Variable	*k*	%	*g*	SE	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Publication Type							0.09	0.34	84.57	0.02
Journal Article	133	62	1.06	0.06	0.95	1.17				
Doctoral Dissertation	40	19	1.35	0.12	1.12	1.57				
Master Thesis	35	14	1.01	0.10	0.82	1.19				
Conference Proceeding	10	5	0.98	0.16	0.68	1.29				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval

The overall I-squared is 84.6, which illustrates that 84.6% of total variance results from study variance and 15.4% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.3.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.40, we can test the significance of the individual model. We have one covariate as publication type. The analysis shows that Q=6.73 with df=3 and p=.081, so it implies that the predictive model probably does not explain the variance in mean effect size. Therefore, it can be said that publication type has no impact on treatment effect if we do not control the other confounders. This individual model only explains the total variance by 0.16%.

4.3.3.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested the common effect size for all studies is the same or not. For this model, we rejected the null hypothesis that all CCSs do not have the same common effect size (Table 4.40, p<.05). Therefore, confounding variables should affect the common effect size except for publication type. In this sense, it is needed to look simultaneous model including more confounders that explain the total variance in true effect sizes more.

Table 4.40 Simple meta-regression analysis for the type of mater

Set	Covariate	*β*	SE	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Publication Type	Intercept	1.06	1.17	18.88	<.001			
	Dissertation	0.28	0.51	2.39	0.02	6.73	3	0.081
	Master	-0.04	0.20	-0.35	0.73			
	Proceeding	-0.06	0.35	-0.30	0.76			
Model						6.73	3	0.081
Residual						1386	214	<.001
Total						1428	217	<.001

β:Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.4 Analyses for Region

What is the role of region (Africa, America, Asia, Europe, Turkey) on the effectiveness of CCS on science achievement?

4.3.4.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.4.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of Africa, America, Asia, Europe, and Turkey are equal to each other.

Table 4.41 expresses the results of heterogeneity analyses between each subgroup. There is significant heterogeneity within subgroups. As a descriptive evaluation, the mean effect sizes for studies done in Turkey are higher than in the other regions. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for different regions are significantly different at 0.05 (p< .05).

Table 4.41 Heterogeneity analysis within subgroups for the type of region

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Region							<.001	0.27	81.12	0.22
Africa	8	3	0.88	0.20	0.48	1.28				
America	28	13	0.68	0.09	0.49	0.86				
Asia	23	11	0.79	0.12	0.56	1.02				
Europe	12	6	0.66	0.18	0.31	1.01				
Turkey	147	67	1.28	0.05	1.18	1.38				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval

The overall I-squared is 81.1, which illustrates that 81.1% of total variance results from study variance and 18.9% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.4.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.42, we can test the significance of the individual model. According to Table 4.42, we have one covariate as country/region. The analysis shows that Q=40.0 with df=4 and p<.001, implying that the predictive model explains the variance in mean effect size. Therefore, it can be said that the region impacts the treatment effect if we do not control the other confounders. This individual model explains the total variance by 21.8 %.

4.3.4.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested the common effect size for all studies is the same or not. For this model, we rejected the null hypothesis that all regions do not have the same common effect size (Table 4.42, p<.05). Therefore, there should be confounding variables that affect the common effect size except for the region. In this sense, it is needed to look simultaneous model including more confounders that explain the total variance in true effect sizes more.

Table 4.42 Simple regression analysis within subgroups for the region

Set	Covariate	β	SE	z	p	Heterogeneity		
						Q	df	p
Region	Intercept	0.88	0.21	4.30	<.001	40.0	4	<.001
	America	-0.19	0.23	-0.81	0.42			
	Asia	-0.09	0.24	-0.37	0.71			
	Europe	-0.23	0.26	-0.86	0.39			
	Turkey	0.39	0.21	1.87	0.06			
Model						40.0	4	<.001
Residual						1128	213	<.001
Total						1418	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.5 Analyses for Subject Domain

What is the role of the subject domain (Biology, Chemistry, Physics) on the effectiveness of CCS on science achievement?

4.3.5.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, We tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.5.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of Biology, Chemistry, and Physics are equal to each other.

Table 4.43 expresses the results of heterogeneity analyses between each subgroup. There is significant heterogeneity within subgroups. More specifically, the mean effect sizes for chemistry are higher than the other subjects. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for different subject domains are significantly different at 0.05 ($p < .05$).

Table 4.43 Heterogeneity analysis within subgroups for the subject domain

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			
					Low Limit	Up. Limit	*p*	T^2	I^2	R^2
Subject Domain							<.001	0.30	82.73	0.14
Biology	42	19	0.82	0.09	0.64	0.99				
Chemistry	85	39	1.37	0.07	1.23,	1.51				
Physics	91	42	0.98	0.07	0.86	1.10				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; *g:* Hedges' g; SE: Standard error; CI: Confidence interval

The overall I-squared is 82.7, which illustrates that 82.7% of total variance results from study variance and 17.3% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.5.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.44, we can test the significance of the individual model. We have one covariate as the subject domain. The analysis shows that Q=28.27 with df=3 and p<.001, implying that the predictive model explains the variance in mean effect size. Therefore, it can be said that subject domain has an impact on the treatment effect if we do not control the other confounders. This individual model explains the total variance by 13.5%.

4.3.5.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested the common effect size for all studies is the same or not. For this model, we rejected the null hypothesis that all subject domains do not have the same common effect size (Table 4.44, $p<.05$). Therefore, confounding variables should affect the common effect size except for the subject domain. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.44 Simple regression analysis within subgroups for the subject domain

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Subject Domain	Intercept	0.82	0.09	8.73	<.001	28.27	3	<.001
	Chemistry	0.54	0.12	4.69	<.001			
	Physics	0.19	0.12	1.64	0.10			
Model						28.27	3	<.001
Residual						1239	214	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.6 Analyses for Question Type

What is the role of question type (mix, objective, and open-ended) on the effectiveness of CCS on science achievement?

4.3.6.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 145 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.6.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of mix, objective and open-ended question types are equal to each other.

Table 4.45 expresses the results of heterogeneity analyses between each subgroup. There is significant heterogeneity within subgroups. More specifically, the mean effect sizes for the objective type tests are higher than the other types. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for different question types are significantly different at 0.05 (p< .05).

Table 4.45 Heterogeneity analysis within subgroups for question type.

Variable	*k*	%	*g*	SE	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Question type							<.001	0.32	82.24	0.14
Mix**	73	35	1.09	0.07	0.94	1.23				
Objective	113	51	1.21	0.06	1.09	1.33				
Open-ended	32	14	0.75	0.10	0.55	0.94				
Overall	218	100	1.10	0.04	1.01	1.19				

** This subgroup has not been included in heterogeneity analysis.
k: Number of studies; %: percent; g: Hedges' g ; SE: Standard error

The mixed type does not include regression analysis. The overall I-squared is 82.2, which illustrates that 82.2% of total variance results from study variance and 17.8% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.6.3 Explained Variance by Individual Model

The individual model enables to test of the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.46, we can test the significance of the individual model. We have one covariate as the question type. The analysis shows that Q=12.85 with df=1 and p<.001, implying that the predictive model probably explains the variance in mean effect size. Therefore, it can be said that question type has an impact on the treatment effect if we do not control the other confounders. This individual model explains the total variance by 13.5%.

4.3.6.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all question types do not have the same common effect size (Table 4.46, $p<.05$). Therefore, there should be confounding variables that affect the common effect size except for the question type. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.46 Simple regression analysis within subgroups for question types.

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Question type	Intercept	1.21	0.06	20.25	<.001	12.85	1	<.001
	Open-ended	-0.45	0.13	-3.58	<.001			
Model						14.54	1	<.001
Residual						805	143	<.001
Total						919	144	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.7 Analyses for Educational Levels

What is the role of educational levels (elementary, middle school, high school, undergraduate) on the effectiveness of CCS on science achievement?

4.3.7.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.7.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:
Means of all true effect sizes of elementary, middle, high school, and undergraduate are equal to each other.

Table 4.47 expresses the results of heterogeneity analyses between each subgroup. There is significant heterogeneity within subgroups. As a descriptive evaluation, the mean effect sizes for high school students are higher than the other educational levels. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for different educational levels are significantly different at 0.05 (p<.05).

Table 4.47 Heterogeneity analysis within subgroups for educational levels

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Educational Levels							0.03	0.34	84.43	0.02
Elementary	13	5	0.96	0.18	0.64	1.29				
Middle	50	49	1.03	0.07	0.90	1.16				
High school	101	23	1.24	0.07	1.10	1.39				
Undergraduate	54	24	0.92	0.09	0.76	1.12				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g ; SE: Standard error

The overall I-squared is 84.4 which illustrates that 84.4% of total variance results from study variance and 15.6% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.7.3 Explained Variance by Individual Model

The individual model enables to test the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.48, we can test the significance of the individual model. We have one covariate as educational levels. The analysis shows that Q=9.27 with df=3 and p=.026, implying that the predictive model probably explains the variance in mean effect size. Therefore, it can be said that educational levels have an impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 2.2%.

4.3.7.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all education levels do not have the same common effect size (Table 4.48, p<.05). Therefore, confounding variables should affect the common effect size except for educational levels. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.48 Simple regression analysis within subgroups for educational levels

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Education Level	Intercept	0.97	0.18	5.40	<.001	9.27	3	0.026
	High school	0.27	0.19	1.42	0.16			
	Middle	0.08	0.20	0.40	0.69			
	Undergrad.	-0.04	0.20	-0.21	0.84			
Model						9.27	3	0.026
Residual						1374	214	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.8 Analyses for Instrument Type

What is the role of instrument type (adapted test, preexisting test, researcher developed test) on the effectiveness of CCS on science achievement?

4.3.8.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.8.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of elementary, middle, high school and undergraduate are equal to each other.

Table 4.49 expresses the results of heterogeneity analyses for heterogeneity between each subgroup. There is significant heterogeneity within subgroups. More specifically, the mean effect sizes for researcher-developed tests are higher than the other types. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for different instrument types are significantly different at 0.05 (p< .05).

Table 4.49 Heterogeneity analysis within subgroups for instrument type.

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			
					Low Limit	Up. Limit	*p*	T^2	I^2	R^2
Instrument type							0.04	0.35	84.42	0.02
Adapted test	23	10	1.01	0.13	0.76	1.26				
Preexisting test	40	19	0.88	0.10	0.68	1.09				
Researcher developed test	155	71	1.17	0.05	1.07	1.28				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of study; %: percent; g: Hedges' g value; SE: Standard error

The overall I-squared is 84.4 which illustrates that 84.4% of total variance results from study variance and 15.6% from sampling error. This also implies the existence of other moderator variables that impact treatment effect.

4.3.8.3 Explanined Varience by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.50, we can test the significance of the individual model. We have one covariate as instrument type. The analysis shows that Q =6.73 with df=2 and p=.035, implying that the predictive model probably explains the variance in mean effect size. Therefore, it can be said that instrument type have an impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 2.3%.

4.3.8.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all instrument types do not have the same common effect size (Table 4.50, p<.05). Therefore, confounding variables should affect the common effect size except for the instrument type. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.50 Simple regression analysis within subgroups for instrument type

Set	Covariate	β	SE	z	p	Heterogeneity		
						Q	df	p
Instruent type	Intercept	1.04	0.14	7.25	<.001	6.73	2	0.035
	Preexisting	-0.16	0.18	-0.87	0.38			
	Researcher developed	0.14	0.15	0.93	0.35			
Model						6.73	2	0.035
Residual						1379	215	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.9 Analyses for Experimental Design

What is the role of experimental design (Poor experimental, Quasi experimental, and True experimental) on the effectiveness of CCS on science achievement?

4.3.9.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.9.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of elementary, middle, high school and undergraduate are equal to each other.

Table 4.51 expresses the results of heterogeneity analyses between each subgroup. There is significant heterogeneity within subgroups. More specifically, the mean effect sizes for quasi-experimental designs are higher than the other designs. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for experimental designs are significantly different at 0.05 ($p < .05$).

Table 4.51 Heterogeneity analysis within subgroups for experimental designs

Variable	*k*	%	*g*	SE	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Experimental Design							<.001	0.33	83.77	0.08
Poor	23	10	0.88	0.10	0.68	1.09				
Quasi	162	75	1.22	0.05	1.12	1.33				
True	33	15	0.64	0.08	0.49	0.79				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of study; %: percent; g: Hedges' g value; SE: Standard error

The overall I^2 is 83.8, which illustrates that 83.8% of total variance results from between-study variance and 16.2% results from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.9.3 Explained Variance by Individual Model

The individual model enables to test of the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.52, we can test the significance of the individual model. We have one covariate as an experimental design. The analysis shows that Q=23.7 with df=2 and p<.001, so it implies that the predictive model probably explains the variance in mean effect size. Therefore, it can be said that experiment design has an impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 7.2% .

4.3.9.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all experiment designs do not have the same common effect size (Table 4.52, p<.05). Therefore, there should be confounding variables that affect the common effect size except for the experiment design. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.52 Simple regression analysis within subgroups for experiment design

Set	Covariate	β	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Experiment Design	Intercept	0,90	0.14	6,57	<.001	23.72	2	<.001
	Quasi	0,32	0.15	2,16	0.03			
	True	-0,25	0.18	-1,41	0.16			
Model						23.72	2	<.001
Residual						1304	212	<.001
Total						1373	214	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.10 Analyses for Teacher Training

What is the role of teacher training (Unstated or stated) on the effectiveness of CCS on science achievement?

4.3.10.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.10.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of stated and unstated are equal to each other.

Table 4.53 expresses the results of heterogeneity analyses for heterogeneity between each subgroup. There is significant heterogeneity within subgroups. More specifically, the mean effect sizes for the studies that conduct the teacher training process are higher than the unstated group. As a result, the null hypothesis is rejected. It indicates that the mean effect sizes for studies that state conducting teacher training are significantly different from the unstated condition at the level of 0.05 (p< .05).

Table 4.53 Heterogeneity analysis within subgroups for teacher training

Variable	*k*	%	*g*	*SE*	95% CI Low Limit	95% CI Up. Limit	Heterogeneity *p*	Heterogeneity T^2	Heterogeneity I^2	R^2
Teacher Training							<.001	0.32	83.52	0.09
Unstaed	107	49	1.08	0.06	0.96	1.21				
Stated	111	51	1.12	0.06	1.00	1.25				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error

The overall I-squared is 83.5, which illustrates that 83.5% of total variance results from study variance and 16.5% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.10.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.54, we can test the significance of the individual model. According to Table 4.54, we have one covariate as teacher training. The analysis shows that Q=16.68 with df=2 and p<.001, implying that the predictive model probably explains the variance in mean effect size.

Therefore, it can be said that teacher training has an impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 9 %.

4.3.10.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that trained or untrained teachers do not have the same common effect size (Table 4.54, $p<.05$). Therefore, there should be confounding variables that affect the common effect size except teacher training. In this sense, it is necessary to look at simultaneous models, including more confounders that explain the total variance in true effect sizes.

Table 4.54 Simple regression analysis within subgroups for teacher training

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Teacher Training	Intercept	0.93	0.06	15.49	<.001	16.68	1	<.001
	yes	0.35	0.09	4.08	<.001			
Model						16.68	1	<.001
Residual						1311	216	<.001
Total						1373	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.11 Analyses for School Type

What is the role of school type (private and public) on the effectiveness of CCS on science achievement?

4.3.11.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 176 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.11.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of private and public are equal to each other.

Table 4.55 expresses the results of simple meta-regression analyses for heterogeneity between subgroups. There is no significant heterogeneity within subgroups. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for school types are not significantly different at 0.05 (p> .05).

Table 4.55 Heterogeneity analysis within subgroups for school type

Variable	*k*	*g*	*SE*	95% CI		Heterogeneity			R^2
				Low Limit	Up Limit	*p*	T^2	I^2	
School Type						0.79	0.35	85.22	<.01
Private	11	1.03	0.15	0.73	1.33				
Public	165	1.10	0.05	1.01	1.2				
Unspecified**	42	1.10	0.11	0.89	1.31				
Overall	218	1.10	0.04	1.01	1.19				

** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 85.2, which illustrates that 85.2% of total variance results from study variance and 14.8% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.11.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.56, we can test the significance of the individual model. We have one covariate as school type. The analysis shows that Q=0.07 with df=1 and p=.790, implying that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that school type has no impact on treatment effect if we do not control the other confounders. This individual model does not explain any variance.

4.3.11.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all school types do not have the same common effect size (Table 4.56, $p<.05$). Therefore, confounding variables should affect the common effect size except for school type. In this sense, it is necessary to look at simultaneous models, including more confounders that explain the total variance in true effect sizes.

Table 4.56 Simple meta-regression analysis within subgroups for school type

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
School type	Intercept	1.05	0.20	5.24	<.001	0.07	1	0.79
	Public	0.06	0.21	0.27	0.79			
Model						0.07	1	0.79
Residual						1176	174	<.001
Total						1176	165	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.12 Analyses for School Location

What is the role of school location (rural and urban) on the effectiveness of CCS on science achievement?

4.3.12.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 174 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.12.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes of rural and urban are equal to each other.

Table 4.57 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for school locations are not significantly different at 0.05 (p> .05).

Table 4.57 Heterogeneity analysis within subgroups for school location

Variable	*k*	*g*	*SE*	95% CI		Heterogeneity			
				Low Limit	Up. Limit	*p*	T^2	I^2	R^2
School location						0.30	0.36	85.34	<.01
Rural	22	0.96	0.12	0.72	1.20				
Urban	157	1.13	0.05	1.03	1.24				
Unspecified **	39	1.08	0.12	0.89	1.36				
Overall	218	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' *g* value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 85.3, which illustrates that 85.3% of total variance results from study variance and 14.7% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.12.3 Explained Variance by Individual Model

The individual model enables to test of the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.58, we can test the significance of the individual model. According to Table 4.58, we have one covariate as school location**.** The analysis shows that Q=1.09 with df=1 and p=.296, so it implies that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that school location has no impact on treatment effect if we do not control the other confounders. This individual model does not explain any variance.

4.3.12.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all school locations do not have the same common effect size (Table 4.58, p<.05). Therefore, confounding variables should affect the common effect size except for school location**.** In this sense, it is necessary to look at simultaneous models, including more confounders that explain the total variance in true effect sizes.

Table 4.58 Simple meta-regression within subgroups for school location

Set	Cov.	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
School location	0.97	0.14	0.69	<.01	0.97	1.09	1	0.30
	0.16	0.15	-0.14	0.30	0.16			
Model						1.09	1	0.30
Residual						1207	177	<.001
Total						1208	178	<.001

Cov: Covariate; β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.13 Analyses for Sampling Method

What is the role of the sampling method (Nonrandom sampling and random sampling) on the effectiveness of CCS on science achievement?

4.3.13.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.13.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of sampling method are equal to each other.

Table 4.59 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for the sampling method are not significantly different at 0.05 (p> .05).

Table 4.59 Heterogeneity analysis within subgroups for sampling method

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			R^2
					Low. Limit	Up. Limit	*p*	T^2	I^2	
Sampling method							0.32	0.35	84.71	0.01
Nonrandom	196	88	1.09	0.05	1.00	1.18				
Random	22	10	1.24	0.15	0.95	1.52				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 84.7, which illustrates that 84.7% of total variance results from study variance and 15.3% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.13.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.60, we can test the significance of the individual model. We have one covariate as the sampling method. The analysis shows that Q=0.99 with df=1 and p=.319, so it implies that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that the sampling method has no impact on the treatment effect if we do not control the other confounders. This individual model explains the total variance by 0.1 %, and this moderator does not explain 99.9% of the total variance.

4.3.13.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all sampling method types do not have the same common effect size (Table 4.60, p<.05). Therefore, there should be confounding variables that affect the common effect size except the sampling method. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.60 Simple meta-regression analysis within subgroups for sampling method

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Sampling Method	Intercept	1.09	0.05	23.28	<.001	0.99	1	0.319
	Random	0.15	0.15	1.00	0.32			
Model						0.99	1	0.319
Residual						1412	216	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.14 Analyses for Researcher Effect

What is the role of researcher effect (Not teacher, One of the teachers, Only teacher) on the effectiveness of CCS on science achievement?

4.3.14.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 168 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.14.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of researcher effect are equal to each other.

Table 4.61 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for the researcher's effect are not significantly different at 0.05 (p> .05).

Table 4.61 Heterogeneity analysis within subgroups for researcher effect

Variable	*k*	*g*	*SE*	95% CI		Heterogeneity			R^2
				Low. Limit	Up. Limit	*p*	T^2	I^2	
Researcher Effect						0.57	0.36	84.64	<.01
Not teacher	108	1.20	0.07	1.07	1.33				
One of the teachers	12	0.99	0.17	0.65	1.32				
Only teacher	48	1.14	0.09	0.97	1.30				
Unspecified**	50	0.88	0.09	0.72	1.05				
Overall	218	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' *g* value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 84.6, which illustrates that 84.6% of total variance results from between-study variance and 15.4% results from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.14.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.62, we can test the significance of the individual model. We have one covariate as the researcher effect. The analysis shows that Q=1.11 with df=2 and p=.573, implying that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that researcher effect has no impact on the treatment effect if we do not control the other confounders. This individual model does not explain any variance.

4.3.14.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all groups do not have the same common effect size (Table 4.62, $p<.05$). Therefore, there should be confounding variables that affect the common effect size except for the researcher effect. In this sense, it is needed to look at simultaneous models including more confounders that explain the total variance in true effect sizes more comprehensively.

Table 4.62 Simple meta-regression analysis within subgroups for researcher effect

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Researcher Effect	Intercept	1.20	0.06	18.83	<.001	1.11	2	0.573
	One of the teachers	-0.20	0.20	-1.00	0.32			
	Only teacher	-0.06	0.12	-0.51	0.61			
Model						1.11	2	0.573
Residual						1074	165	<.001
Total						1096	167	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.15 Analyses for Teacher Effect

What is the role of teacher effect (Different teacher or same teacher) on the effectiveness of CCS on science achievement?

4.3.15.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 178 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.15.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of teacher effect are equal to each other.

Table 4.63 expresses the results of heterogeneity analyses for each subgroup.
There is no significant heterogeneity within subgroups. As a result, the null hypothesis is not rejected. It indicates that the mean effect sizes for teacher effect are not significantly different at the level of 0.05 (p> .05).

Table 4.63 Heterogeneity analysis within subgroups for teacher effect

Variable	*k*	*g*	*SE*	95% CI		Heterogeneity			
				Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Teacher Effect						0.60	0.37	85.12	<.01
Different teachers	44	1.12	0.09	0.94	1.30				
Same teacher	134	1.83	0.06	1.07	1.30				
Unspecified**	40	0.81	0.08	0.64	0.97				
Overall	218	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 85.1, which illustrates that 85.1% of total variance results from study variance and 14.9% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.15.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.64, we can test the significance of the individual model. We have one covariate as the teacher effect. The analysis shows that Q=0.27 with df=1 and p=.601, implying that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that teacher effect has no impact on the treatment effect if we do not control the other confounders. This individual model explains the total variance by 0%, and this moderator does not explain 100 % of the total variance.

4.3.15.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we reject the null hypothesis that all groups do not have the same common effect size (Table 4.64, p<.05). Therefore, there should be confounding variables that affect the common effect size except for the teacher effect. In this sense, it is needed to look at simultaneous models including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.64 Simple meta-regression analysis within subgroups for teacher effect

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Teacher Effect	Intercept	1.13	0.10	11.21	<.001	0.27	1	0.60
	Same teacher	0.08	0.12	0.69	0.49			
Model						0.27	1	0.60
Residual						1182	176	<.001
Total						1182	177	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.16 Analyses for Number of Tiers

What is the role of the number of tiers (one-tier, two-tier, or more) on the effectiveness of CCS on science achievement?

4.3.16.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 204 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.16.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of the number of tiers are equal to each other.

Table 4.65 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis was not rejected and indicated that the mean effect sizes for the number of tiers are not significantly different at 0.05 (p>0.05).

Table 4.65 Heterogeneity analysis within subgroups for number of tiers

Variable	*k*	%	*g*	*SE*	95% CI		Heterogeneity			R^2
					Low. Limit	Up. Limit	*p*	T^2	I^2	
Number of Tiers							0.22	0.34	84.50	<.01
1	164	75	1.13	0.05	1.03	1.23				
2	31	15	0.91	0.11	0.69	1.12				
3	15	6	1.13	0.15	0.84	1.42				
Mix	8	4	1.18	0.31	0.58	1.78				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 84.5, which illustrates that 84.5% of total variance results from study variance and 15.5% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.16.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.66, we can test the significance of the individual model. We have one covariate as the number of tiers**.** The analysis shows that Q=3.08 with df =2 and p=.215, so it implies that the predictive model probably does not explain any variance in mean effect size.

Therefore, it can be said that number of tiers has no impact on the treatment effect if we do not control the other confounders. This individual model does not explain any variance.

4.3.16.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test the common effect size for all studies is the same or not. For this model, we reject the null hypothesis that all groups do not have the same common effect size (Table 4.66, p<.05). Therefore, confounding variables should affect the common effect size except for the number of tiers. In this sense, it is needed to look at simultaneous models including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.66 Simple meta-regression analysis within subgroups for number of tiers

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Number of Tiers	Intercept	1.13	0.05	22.28	<.001	3.08	2	0.215
	2	-0.22	0.13	-1.73	0.08			
	3	0.01	0.18	0.07	0.95			
Model						3.08	2	0.215
Residual						1325	201	<.001
Total						1361	203	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.17 Analyses for Treatment Verification

What is the role of treatment verification (stated or unstated) on the effectiveness of CCS on science achievement?

4.3.17.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.17.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of treatment verification are equal to each other.

Table 4.67 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis was not rejected and indicated that the mean effect sizes for treatment verification are not significantly different at the level of 0.05 (p>.05).

Table 4.67 Heterogeneity analysis within subgroups for treatment verification

Variable	*k*	*%*	*g*	*SE*	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	*p*	T^2	I^2	R^2
Treatment Verification							0.57	0.35	84.8	<.01
Unstated	107	49	1.08	0.06	0.96	1.21				
Stated	111	51	1.13	0.06	1.00	1.25				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error

The overall I-squared is 84.8, which illustrates that 84.8% of total variance results from study variance and 15.2% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.17.3 Explained Variance by Individual Model

The individual model enables to test predictive model that explains any of the variances in effect size for a single moderator. In Table 4.68, we can test the significance of the individual model. We have one covariate as treatment verification**.** The analysis shows that Q=0.32 with df=1 and p=.570 so it implies that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that treatment verification has no impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 0%, and 100% of the total variance is not explained by this moderator.

4.3.17.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all groups do not have the same common effect size (Table 4.68, $p<.05$). Therefore, there should be confounding variables that affect the common effect size except for verification. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.68 Simple meta-regression analysis within subgroups for treatment verification

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Verification	Intercept	1.08	0.06	16.88	<.001	0.32	1	0.57
	yes	0.05	0.09	0.57	0.57			
Model						0.32	1	0.57
Residual						1424	216	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.18 Analyses for Measuring Outcome

What is the role of outcome measure type (conceptual change or general achievement) on the effectiveness of CCS on science achievement?

4.3.18.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we also calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 208 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.18.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effect sizes within subgroups of outcome measure are equal to each other. Table 4.69 expresses the results of heterogeneity analyses between each subgroup. There is no significant heterogeneity within subgroups. As a result, the null hypothesis was not rejected and indicated that the mean effect sizes for outcome measures are not significantly different at 0.05 (p> .05).

Table 4.69 Heterogeneity analysis within subgroups for outcome measure type

Variable	*k*	%	*g*	*SE*	95% CI Low. Limit	95% CI Up. Limit	Heterogeneity *p*	Heterogeneity T^2	Heterogeneity I^2	R^2
Measuring outcome							0.17	0.36	85.1	<.01
Conceptual Change	192	90	1.12	0.05	1.03	1.22				
General Achievement	16	6	0.87	0.14	0.60	1.14				
Mix**	10	4	1.04	0.17	0.70	1.37				
Overall	218	100	1.10	0.04	1.01	1.19				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error
** This subgroup has not been included in heterogeneity analysis.

The overall I-squared is 85.1, which illustrates that 85.1% of total variance results from between-study variance and 14.9% result from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.18.3 Explained Variance by Individual Model

The individual model enables to test of the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.70, we can test the significance of the individual model. We have one covariate as outcome measure type**.** The analysis shows that Q=1.93 with df=1 and p=.165, so it implies that the predictive model probably does not explain any variance in mean effect size. Therefore, it can be said that outcome measure has no impact on treatment effect if we do not control the other confounders. This individual model explains the total variance by 0%, and 100% of the total variance is not explained by this moderator.

4.3.18.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. We rejected the null hypothesis that all groups do not have the same common effect size (Table 4.70, $p<.05$). Therefore, confounding variables should affect the common effect size except for measuring outcome. In this sense, it is needed to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.70 Simple meta-regression analysis within subgroups for outcome measuring type

Set	Cov.	β	SE	z	p	Heterogeneity		
						Q	df	p
Measuring Outcome	Intercept	1.12	0.05	23.53	<.001	1.93	1	0.165
	General Achievement	-0.24	0.17	-1.39	0.17			
Model						1.93	1	0.165
Residual						1379	206	<.001
Total						1389	207	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.19 Analyses for Sample Size

What is the role of sample size (24-396) on the effectiveness of CCS on science achievement?

4.3.19.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We include 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.19.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effects for each sample with different sample sizes are equal to each other.

Table 4.71 expresses the results of heterogeneity analyses for sample size. There is significant heterogeneity for samples with different sample sizes. As a result, the null hypothesis was not rejected and indicates that the mean effect sizes for sample size are not significantly different at 0.05 (p> .05).

Table 4.71 Heterogeneity analysis for sample size

Variable	k	g	SE	95% CI		Heterogeneity			R^2
				Lower Limit	Upper Limit	p	T^2	I^2	
Sample Size						0.20	0.33	84.07	0.05
16-46	45	1.09	0.09	0.92	1.26				
47-56	45	1.07	0.09	0.90	1.24				
57-72	45	1.23	0.10	1.04	1.42				
73-100	41	1.18	0.10	0.97	1.383				
102-396	42	0.93	0.10	0.73	1.12				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error

The overall I-squared is 84.1, which illustrates that 84.1% of total variance results from study variance and 15.9% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.19.3 Explained Variance by Individual Model

The individual model enables testing the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.72, we can test the significance of the individual model. We have one covariate as sample size. The analysis shows that Q =5.98 with df =4 and p<.001, so it implies that the predictive model probably explains significant variance in mean effect size. Therefore, it can be said that sample size has an impact on treatment effect if we don't control the other confounders. This individual model explains the total variance by 5%.

4.3.19.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all groups do not have the same common effect size (Table 4.72 $p<.05$). Therefore, confounding variables should affect the common effect size except for sample size. In this sense, it is necessary to look at simultaneous models including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.72 Simple meta-regression analysis within subgroups for sample size

Set	Cov.	β	SE	z	p	Heterogeneity		
						Q	df	p
Sample Size	1.10	0.10	0.91	0.00	1.10	5.98		<.001
	-0.03	0.14	-0.30	0.84	-0.03			
	0.13	0.14	-0.15	0.36	0.13			
	0.07	0.14	-0.20	0.61	0.07			
	-0.18	0.14	-0.45	0.20	-0.18			
Model						5.98	4	0.20
Residual						1298	213	<.001
Total						1428	217	<.001

Cov: Covariate; β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.20 Analyses for Intervention Length

What is the role of intervention length on the effectiveness of CCS on science achievement?

4.3.20.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 144 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.20.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effects for each sample with different intervention lengths are equal to each other.

Table 4.73 expresses the results of heterogeneity analysis for heterogeneity for intervention length. There is significant heterogeneity for samples with different intervention lengths. As a result, the null hypothesis was rejected and indicated that the mean effect sizes for studies with varying intervention lengths are significantly different at 0.05 (p> .05).

Table 4.73 Heterogeneity analysis for intervention length

Variables		k	β	*95% CI*		p	T^2	I^2	R^2
				Low. Limit	Up. Limit				
Intervention Length	0-48 course hour	144	0.02	0.01	0.03	<.001	0.31	83.46	0.11

The overall I-squared is 83.5, which illustrates that 83.5% of total variance results from study variance and 16.5% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.20.3 Explained Variance by Individual Model

The individual model enables testing the predictive model that explains any of the variances in effect size for a single moderator. In Table 4.74, we can test the significance of the individual model. We have one covariate as intervention duration. The analysis shows that Q=15.81 with df=1 and p<.001, so it implies that the predictive model probably explains significant variance in mean effect size. Therefore, it can be said that intervention length has an impact on treatment effect if we don't control the other confounders. This individual model explains the total variance by 10 %.

4.4.20.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that instructions using different intervention lengths do not have the same common effect size (Table 4.74, p<.05). Therefore, confounding variables should affect the common effect size except for intervention length. In this sense, it is necessary to look at simultaneous models including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.74 Simple meta-regression analysis within subgroups for intervention length

Set	Covariate	β	SE	z	p	Heterogeneity		
						Q	df	p
Intervention Length	Intercept	0.84	0.08	10.12	<.001	15.81	1	<.001
	Intervention Length	0.02	0.01	3.98	<.001			
Model						15.81	1	<.001
Residual						858	142	<.001
Total						958	143	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.21 Analyses for Publication Year

What is the role of publication year (1989-2000, 2001-2005, 2006-2010, 2011-2015, 2016-2020) on the effectiveness of CCS on science achievement?

4.3.21.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.21.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effects for each sample with different publication years are equal to each other.

Table 4.75 expresses the results of heterogeneity analysis for different years intervals. There is a significant heterogeneity for effect size means for studies with different year intervals. As a result, the null hypothesis is rejected and indicates that the mean effect sizes for the year intervals are significantly different at 0.05 (p<.05).

The overall I-squared is 82.9 which illustrates that 82.9 % of total variance results from study variance and 17.1% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

Table 4.75 Heterogeneity analysis for the publication year

Variable	k	%	g	SE	95% CI		Heterogeneity			
					Low. Limit	Up. Limit	p	T^2	I^2	R^2
Publication Year							<.001	0.31	82.94	0.12
1989-2000	14	6	0.95	0.16	0.35	0.78				
2001-2005	48	22	1.05	0.19	0.88	1.28				
2006-2010	73	34	1.27	0.18	1.15	1.45				
2011-2015	69	32	1.14	0.18	0.95	1.24				
2016-2020	14	6	1.13	0.23	0.49	1.00				

k: Number of studies; %: percent; g: Hedges' g value; SE: Standard error

4.3.21.3 Explained Variance by Individual Model

The individual model enables testing the predictive model that explains any variances in effect size for a single moderator. In Table 4.76, we can test the significance of the individual model. According to Table 4.76, we have one covariate as a year. The analysis shows that Q=0.53 with df=1 and p =0.467, implying that the predictive model probably does not explain significant variance in mean effect size. Therefore, it can be said that year has no impact on treatment effect if we don't control the other confounders. This individual model explains the total variance by 12%.

4.3.21.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all groups do not have the same common effect size (Table 4.76, p<.05). Therefore, confounding variables should affect the common effect size except for publication year. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.76 Simple meta-regression analysis for publication year

Set	Subgroups	β	SE	z	p	Heterogeneity		
						Q	df	p
Publication Year	Intercept	0.58	0.16	3.61	<.001	0.53	1	0.47
	2001-2005	0.49	0.19	2.65	0.01			
	2006-2010	0.72	0.18	4.05	<.001			
	2011-2015	0.51	0.18	2.86	<.001			
	2016-2020	0.17	0.23	0.72	0.47			
Model						22.59	4	0.02
Residual						1298	213	<.001
Total						1428	217	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.22 Analyses for Intervention Intensity

What is the role of intervention intensity on the effectiveness of CCS on science achievement?

4.3.22.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 142 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.22.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:

Means of all true effects for each sample with different intervention intensities are equal to each other.

Table 4.77 expresses the results of heterogeneity analyses for intervention intensity. There is no significant heterogeneity between study sample means with different intervention intensities. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes for treatment intensity are not significantly different at 0.05 (p> .05).

Table 4.77 Heterogeneity analysis for intervention intensity

Variables		k	β	95% CI		p	T^2	I^2	R^2
				Low Limit	Up. Limit				
Intervention Intensity	1-8 course hours per week	142	0.02	-0.06	0.107	0.56	0.37	85.58	<.01

The overall I-squared is 85.6, which illustrates that 85.6% of total variance results from study variance and 14.4% from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.22.3 Explained Variance by Individual Model

The individual model enables testing the predictive model that explains any variances in effect size for a single moderator. In Table 4.78, we can test the significance of the individual model. We have one covariate as treatment intensity. The analysis shows that Q=0.33 with df=1 and p=.665, implying that the predictive model probably does not explain significant variance in mean effect size. Therefore, it can be said that treatment intensity has no impact on treatment effect if we don't control the other confounders. This individual model does not explain any variance.

4.3.22.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we test whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that instructions using different intervention intensities do not have the same common effect size (Table 4.78, p<.05). Therefore, confounding variables should affect the common effect size except for intervention intensity. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.78 Simple meta-regression analysis within subgroups for intervention intensity

Set	Covariate	*β*	*SE*	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Treatment Intensity	Intercept	1.09	0.13	8.15	<.001	0.33	1	0.57
	Treatment Intensity	0.02	0.04	0.43	0.66			
Model						0.33	1	0.57
Residual						962	136	<.001
Total						963	137	<.001

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

4.3.23 Analyses for Class Size

What is the role of class size interval (8-22, 23-26, 27-30, 31-38, 39-87) on the effectiveness of CCS on science achievement?

4.3.23.1 Unit of Analysis and Model

We accepted each primary study as unit of analysis so that we used only one mean effect size value for calculations for a single study. Additionally, we calculated a single weighted effect size value for studies with more than one effect size, which is inversely proportional to the standard error of effect size values. We included 218 primary studies for analyses. The population that each primary study was obtained is unique in social sciences. That is why the random-effects model was used to analyze the heterogeneity. Therefore, it is more meaningful to analyze moderator variables in this model. Additionally, more than five effect sizes for each subgroup are needed to provide valid calculations for the overall mean effect size (Borenstein et al., 2009). For the scope of this study, we analyzed the correlations between moderators and treatment effects through simple and simultaneous meta-regression analyses. In this way, we tried to observe the effect of different moderator variables that are likely to impact the treatment effect. Thus, how much influence each moderator variable has on the treatment effect can be observed.

4.3.23.2 Simple Meta-regression Analyses

In primary studies, there may be lots of factors that have an impact on mean effect size simultaneously. Multiple regression analysis provides to investigate the effect of potential confounders by isolating other variables. The exact process is still valid for meta-regression analysis. Therefore, it is critical to investigate different confounders by including enough studies in meta-analysis to isolate the unique effect of each confounding variable. We also examined 23 confounding variables on the overall

treatment effect. In this part, we discussed the moderators individually. In the second part, we discussed the moderators simultaneously.

Null Hypothesis:
Means of all true effects for each sample with different class size intervals are equal to each other.

Table 4.79 expresses the results of a simple meta-regression analysis for heterogeneity for class sizes. There is no significant heterogeneity for sample means with different class sizes. As a result, the null hypothesis is not rejected and indicates that the mean effect sizes with varying class sizes are not significantly different at the level of 0.05 ($p > .05$).

Table 4.79 Heterogeneity analysis for class size

Variables		*k*	*β*	95% CI		*p*	T^2	I^2	R^2
				Low Limit	Up. Limit				
Class Size	16-87	218	-0.00	-0.01	1.69	0.69	0.35	84.73	<.01

β: Coefficient; SE: Standard error; Q: Total heterogeneity; df: Degree of freedom

The overall I-squared is 84.7, which illustrates that 84.7 % of total variance results from study variance and 15.3 % from sampling error. This result also implies the existence of other moderator variables that have an impact on the treatment effect.

4.3.23.3 Explained Variance by Individual Model

The individual model enables testing the predictive model that explains any variances in effect size for a single moderator. In Table 4.80, we can test the significance of the individual model. We have one covariate as class size. The analysis shows that Q=5.74 with df=4 and p=.219, implying that the predictive model probably does not explain significant variance in mean effect size. Therefore, it can be said that class size has no impact on the treatment effect if we don't control the other confounders. This individual model explains the total variance by 3%, and this moderator does not explain 97% of the total variance.

4.3.23.4 Residual

The residual is a test for consistency with the assumptions of the fixed-effect model. That is, we tested whether the common effect size for all studies is the same. For this model, we rejected the null hypothesis that all groups do not have the same common effect size (Table 4.80, p<.05). Therefore, there should be confounding variables that affect the common effect size except for class size. In this sense, it is necessary to look at simultaneous models, including more confounders that comprehensively explain the total variance in true effect sizes.

Table 4.80 Simple meta-regression analysis within subgroups for class size

Set	*k*	*β*	SE	*z*	*p*	Heterogeneity		
						Q	*df*	*p*
Class Size	8-22	0.95	0.09	10.16	<.001	5.74	4	0.22
	23-26	0.10	0.14	0.75	0.45			
	27-30	0.32	0.14	2.29	0.02			
	31-38	0.20	0.14	1.44	0.15			
	39-87	0.17	0.14	1.25	0.21			
Model						5.74	4	0.22
Residual						1357	213	<.001
Total						1428	217	<.001

k: Number of study; %: percent; g: Hedges' g value; SE: Standard error

4.4 Multiple Meta-Regression Analyses

In many ways, the reasons for dispersion on effect sizes are masked by unconsidered moderators that are effective on true variance in mean effects. The potential pitfall of simple meta-regression is that acknowledging unisolated moderators causes an overly simplistic assessment of the effect value. Multiple meta-regression analyses tried to determine the possible confounders to estimate the effect of the independent variables on dependent variables. Researchers should include each possible explanatory moderator in the regression model to explain the observed heterogeneity between effect size values by increasing the explained variance into the total variance (R^2) ratio. On the other hand, increasing the number of moderators causes a decrease in sample size for each subgroup. Therefore, it is critical to provide the optimum number of moderators that yield the optimum R^2 value. At the same time, the significance value within subgroups of each moderator in the general model should be taken into consideration.

Three different multiple regression methods yield the best general model for a defined moderator set: simultaneous, stepwise, and sequential. Each method has different strengths and weaknesses during heterogeneity analysis. Therefore, for the scope of this study, the combination of these three methods is used to define the best general model for the heterogeneity analysis process. We have 23 explanatory moderators that are yielded from primary studies. But, it is not efficient and appropriate to include all moderators at the same time in the general model. Some of the moderators can be coded for less number of studies. On the other hand, the multiple regression process compares the common study findings. This situation causes us to lose data for the general model. That is why it is reasonable to use moderators with effect size values that are coded for 218 primary studies for the scope of this meta-analysis. In this sense, the number of moderators for the final moderator set is limited in the general model. In the first stage, a simultaneous model is intended to use to observe the combined effect. The CMA version 3.0 program allows at most 11 moderators to analyze simultaneously. We have 23 moderators that can be included in analyses. Therefore, We tried to determine practically and theoretically important moderators that explain the total variance, which was the general model. This process needs to search previous literature on the effectiveness of intervention methods from a meta-analytic perspective. Additionally, the simple meta-regression analysis results should also be consistent with previous findings. These moderators are region, subject domain, education level, instrument type, teacher training, sample size, question type, experiment design, publication type, material type, intervention length, and publication year. Nevertheless, the combined effect of these moderators does not give precise evidence about the best general model and the individual effect of moderators in the model.

It should also be investigated on the question, "which moderators are really effective for treatment in simultaneous models?". "Is there any significant correlation between moderators?" These analyses provide more comprehensive knowledge about each moderator in the model. There are indexes that inform us about the correlation matrix for moderators, like tolerance and VIF indexes. The VIF (variance inflation factor) index is the degree of multicollinearity in the moderator set during the multiple regression analyses. This value should not exceed 10 to inform uncorrelated independent subgroups (Cohen et al., 2003). On the other hand, the VIF or tolerance indexes do not give evidence about the moderator correlations but rather inform us about the correlations between subgroups."The VIF index is the $1/(1-R_i^2)$, where R_i^2 is the squared multiple correlation coefficient between column i and the remaining columns of the design matrix. VIF is not very informative as some variables are algebraically connected to each other" (Harrel, 2015, p.79). Therefore, it may not be observed multicollinearity between moderators with VIF index may not be observed in the distribution of effect sizes, p-values, and R^2 values. The tolerance is just an inverse proportion of VIF, which is 1/VIF. In this sense, investigating moderators by practical applications can give more reliable evidence about the degree of correlations for moderators in simultaneous analyses.

Simultaneous Model

No isolated treatment effect yields a precise result since CCS also interacts with other explanatory variables simultaneously. In this sense, it is essential to set a model that includes explanatory variables to explain the true impact of treatment intervention. This final state of treatment effect will be adjusted by controlling potentially important factors. Therefore, describing the potentially important moderators improves our understanding of the true treatment effect. For the scope of this study, the theoretical background, simple and multiple meta-regression analyses enable to derive a simultaneous model that best explains the variation in effect sizes.

The inclusion and exclusion of moderators in the simultaneous model are critical in that theoretical background, and statistical evidence should be balanced to set the most reasonable scheme. For the scope of this study, firstly, moderators related to the randomization process and internal validity threats are prominent to test the effect of study quality on achievement. In any area of education literature, the quality of the intervention process may interact with the findings (Glass & Smith, 1978). Previous meta-analyses also infer that study quality interacts with effect value, namely, poor study quality considerably inflates the effect size (Chadwick, 1997; Cuijpers et al., 2009). In this sense, design characteristics related to the type of experiment design (poor, quasi, true experimental) are an effective moderator in literature.

Additionally, White (1988) and Johnson & Johnson (2000) proposed that teacher training impacted the efficiency of any instructional method. The efficacy of a method is strictly related to the implementer's proficiency. In this sense, better teacher training results in a higher effect size value. It is important that the implementer should be adequately trained on the implementation process to become evident of the effect of the method. The simple and multiple meta-regression analyses also support that experimental design and teacher training are significant independent moderators on achievement. Therefore, including these two moderators in the general model is substantial. The Table 4.81 demonstrates the impact of the combined effect of two moderators on achievement by reporting notable R^2 values.

Table 4.81 The multiple meta-regression model for moderators related to design characteristics

No	Variables	Study Characteristics	Simple R^2	*P*	*k*	Multiple R sq.
1	Experiment design	Design	0.08	<.001	218	0.16
	Teacher training	Design	0.10	<.001	218	

Secondly, the publication characteristics' effect is frequently tested in meta-analytic studies. The previous meta-analytic CCS studies (Armağan, 2011; Gelen, 2015) propose that doctoral dissertations have a higher mean value. Dissertations are expected to give more prominence to controlling confounding variables across other types. Namely, doctoral dissertations provide better control for internal validity threads. That is why the overall effect value is possibly influenced by publication type. The multiple meta-regression analyses also support the theoretical background, so dissertations have a higher effect value than other types (Table 4.82). Therefore, we included in the general model to moderate the effect of CCS on student achievement.

Table 4.82 The general model that includes moderators related to design and publication characteristics

No	Variables	Study Characteristics	Simple R^2	*P*	*k*	Multiple R sq.
2	Experimental design	Design	0.08	.005	218	0.27
	Teacher training	Design	0.10	.012	218	
	Publication type	Publication	0.02	.026	218	
	Publication year	Publication	0.12	.010	218	

The impact of *the* publication year should be considered in that year is a significant publication character in meta-analytic studies on education (Bayraktar, 2000; Chadwick, 1997). For the scope of this study, we coded publication year as the year in which the treatment was applied. As theoretically, there are two bases for this impact. Firstly, new studies may take into consideration the limitations of early studies during application so that new studies provide a more planned process. This may inflate the effect value for new studies (Armağan,2011). Secondly, the popularity of methods gives direction to researchers. In some year intervals, CCS had investigated more intensely to disclose the true impact of the method on the learning process. This trend can be observed by the number of studies throughout the years (Table 4.2). This situation might cause an increase in the representativeness of the sample for the population and provide more precise findings in those years. Therefore year may become an effective moderator on achievement scores. The statistical evidence and early meta-analysis on CCS also support this hypothesis strictly (Armağan, 2011; Gelen, 2015). A distinctive pattern for those year intervals was observed related to the popularity of the method and overall effect value (Table 4.2). Thus, year is a prominent independent moderator on student achievement for CCS and should be tested in the general model. There are strong indications that study intervention characteristics are effective moderators in literature (subject domain, intervention length, and material type). In particular, different subject domains have different impacts on treatment interventions for CCS (Armağan, 2011; Mufit et al., 2020). The statistical findings support the idea that different domains may create different treatment impacts on achievement even the other moderators controlled (Table 4.83). On the other hand, multiple regression analyses do not support the efficacy of material type. Additionally, the missing data on intervention length conceal the role of this moderator.

Table 4.83 The general model that includes moderators related to design, publication, and intervention characteristics

No	Variables	Study Characteristics	Simple R^2	*p*	*k*	Multiple R sq.
3	Experimental design	Design	0.08	<.001	218	0.32
	Teacher training	Design	0.09	.003	218	
	Publication type	Publication	0.02	.047	218	
	Publication Year	Publication	0.12	.002	218	
	Subject Domain	Intervention	0.14	.001	218	

Finally, it can be prioritized that the possible effect of sample characteristics like region (Africa, Asia, America, Europe, and Turkey) is a critical issue. The simple and multiple meta-regression analyses also support that region is a significant independent moderator on achievement even after controlling publication characteristics, experimental design types, teacher training process, and subject domains (Table 4.84). Therefore it is substantial to include region moderator in the general model.

Table 4.84 The general model that includes moderators related to design, publication, intervention, and sample characteristics

Model	Variables	Study Characteristics	Simple R^2	*p*	*k*	Multiple R^2
4*	Experimental design	Design	0.08	.010	218	0.35
	Teacher training	Design	0.09	.008	218	
	Publication type	Publication	0.02	.025	218	
	Publication Year	Publication	0.12	.046	218	
	Subject Domain	Intervention	0.14	.005	218	
	Region	Sample	0.24	.029	218	

Additionally, the treatment process can be accepted as an effective moderator of achievement. One important hypothesis is that the short duration of the intervention causes a lack of deeper comprehension of knowledge for students. That is why increasing intervention length enables most likely to boost effect value (White, 1988). In this sense, simple meta-regression analyses inform that the intervention length is an effective moderator. But, the coded sample in this moderator is smaller than other moderators. Therefore, it is reasonable to test this moderator finally. On the other hand, multiple meta-regression imply that intervention length also interacts with other moderators (Table 4.85). Therefore, it is not an individual moderator rather, it reflects the effect of studying other moderators. In this sense, it is not reasonable to include it in the general model.

Table 4.85 The general model that includes moderators related to design, publication, intervention, and sample characteristics

No	Variables	Study Characteristics	Simple R^2	*P*	*k*	Multiple R sq.
6	Experimental design	Design	0.08	0.02	144	0.35
	Teacher training	Design	0.09	0.02	144	
	Publication type	Publication	0.02	0.29	144	
	Year	Publication	0.12	0.03	144	
	Domain	Intervention	0.14	0.07	144	
	Region	Sample	0.24	0.50	144	
	Intervention length	Intervention	0.11	0.11	144	

Figure 4.14 demonstrates the simultaneous model yielded by 23 explanatory moderators and 218 primary studies. According to the general model, region, domain, experimental design, teacher training, year, and publication type impact the effectiveness of CCS on student achievement. The multiple meta-regression results imply that this model can explain about 35% heterogeneity.

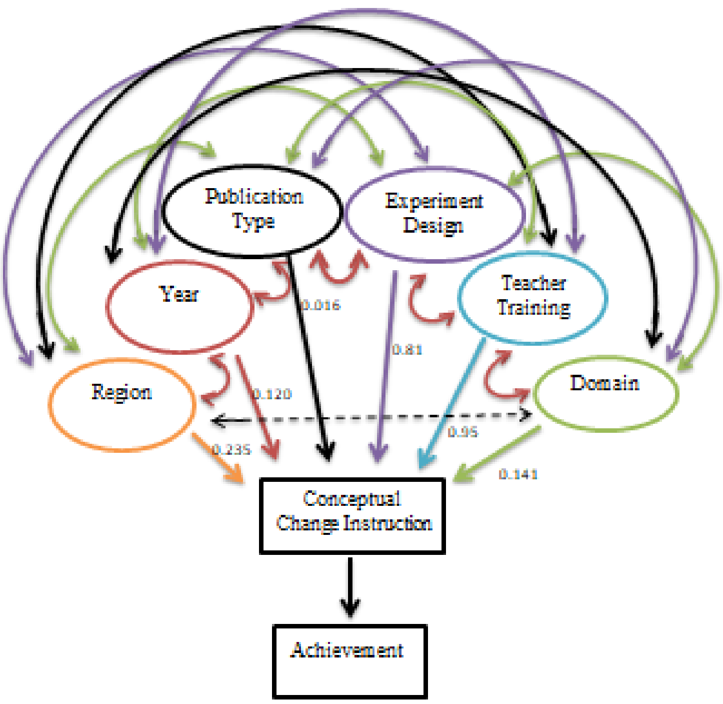


Figure 4.14 Path representation of simultaneous model on treatment effect

4.5 Publication Bias

As we explained before, we ran several analyses to examine the degree to which publication bias was a threat to the validity of this meta-analysis. First, Orwin's fail-safe N was 1853, much larger than the total number of studies (218) included in this meta-analysis. Therefore, the results of the meta-analysis seem robust to publication bias. Then, we examined the funnel plot for small study bias. Figure 4.2 shows the funnel plot, which appears asymmetrical: as the standard error increases, the effect sizes tend to have larger effect sizes. Egger's regression test also points out an asymmetry ($t(216)= 8.21$, two-tailed $p<.001$). The Trim and Fill method imputes 70 studies to the left of the mean to eliminate this asymmetry, resulting in an adjusted effect size of 0.71 when using a fixed-fixed model. On the other hand, the random-random model yields an adjusted effect size of 1.23, imputing 24 studies to the right of the mean. The asymmetry in the funnel plot and the trim and fill method indicate a small study bias, which may or may not result from a publication bias. Therefore, we also performed selection models to calculate an adjusted effect size. The selection model gives us an adjusted g = 0.93, 95% CI [0.75, 1.11] for random effects estimate. Finally, we conducted a Robust Bayesian Meta-Analysis using 36 models to calculate a robust mean effect size. This analysis gives us a robust mean effect size of g = 0.93. The adjusted effect sizes estimated by the selection method and Robust Bayesian Meta-Analysis are close to each other and are still considered large effects. Accordingly, we conclude that publication bias does not change our interpretation of the mean effect size. Table 4.86 presents the results of our analyses to examine publication bias.

Table 4.86 Summary of the adjustment analyses for publication bias

Strategy	k	Unadjusted Overall Effect Size [95% CI]	Adjusted Overall Effect Size: RoBMA	Selection model (Random model) [95% CI]	Trim and Fill (Fixed-random) [95% CI]	Trim and Fill (Random-random) [95% CI]
Overall Conceptual Change	218	1.10, [1.01,1.19]	0.93	0.92, [0.75,1.11]	0.71 [0.75, 1.11]	1.23 [1.13, 1.33]
Cognitive Conflict	150	1.10, [1.00,1.21]	0.94	0.93, [0.71,1.15]	0.69 [0.58, 0.81]	1.21 [1.09, 1.33]
Cognitive Bridging	30	1.06, [0.84,1.28]	0.87	0.91, [0.48,1.35]	1.06 [0.90, 1.34]	1.12 [0.99, 1.44]
Ontological Category Shift	9	0.88, [0.50,1.26]	0.69	0.85, [0.75,1.05]	0.58 [0.20, 0.96]	0.88 [0.50, 1.25]

k: Number of studies; *CI:* Confidence interval.

4.6 Heterogeneity Analysis

The heterogeneity analyses yielded statistically significant results for not only the effect size distribution of all 218 studies ($Q(217)=1428.10, p<.001$) but also that of cognitive conflict ($Q(149)=1017.18, p<.001$), cognitive bridging ($Q(29)=155.83$, p<.001), and ontological category shift ($Q(8)=29.21, p<.001$). As shown in Table 4.87, all I^2 values are above 80% except for ontological category shift (73%), which means that most of the total variance corresponds to the between-study variance for all distributions. In addition, large tau squared values provide evidence of high heterogeneity within the distributions, resulting in a wide prediction interval for each distribution. In other words, each strategy has a significant and large amount of true heterogeneity.

Table 4.87 The heterogeneity analyses result for CCSs.

					Heterogeneity			
Strategy	*k*	*g*	*p*	95% CI	T^2	I^2	Q_T	95% PI
Overall Conceptual Change	218	1.10	<.001	[1.01, 1.19]	.35	84.80	1428.10	[0.19,2.38]
Cognitive Conflict	150	1.10	<.001	[1.00, 1.21]	.36	85.35	1017.18	[0.18,2.03]
Cognitive Bridging	30	1.06	<.001	[0.84, 1.28]	.30	81.39	155.83	[0.20,1.92]
Ontological Category Shift	9	0.88	<.001	[0.50, 1.26]	.22	72.62	29.21	[0.07,1.69]

k: Number of studies; g: Hedges' g; Q_T: Total heterogeneity; CI =Confidence interval; PI= Prediction interval; T^2:Between study variance; I^2: The ratio of true variance to the total variance.

CHAPTER 5

DISCUSSION AND CONCLUSION

In this meta-analysis, we aimed to synthesize the studies investigating the effect of CCS on science achievement in more than 30 years. By doing so, we sought not only to estimate the degree to which overall CCS affects science achievement but also to examine if there is any difference in the effectiveness of different types of CCS. In this sense, we compared three main types of CCS extensively studied in the literature that derived from three knowledge perspectives: cognitive conflict, cognitive bridging, and ontological category shift. Furthermore, we conducted moderator analyses to explain the heterogeneity in the distribution of true effect sizes by conducting simple and multiple meta-regression analyses. We performed a RoBMA adjustment analysis for publication bias sensitivity.

We tried to standardize the effect values by including primary studies that used control group design as traditional instruction. So we aimed to yield more consistent findings by controlling the effect of traditional instruction across the studies.

In this context, we combined the results of 218 studies involving 18,051 students from four continents. This broad sample of studies consists of published and unpublished studies conducted in a wide range of time-interval from 1989 to 2020. We also investigated publication, design, sample, intervention, and measurement characteristics. The studies were collected through a systematic and comprehensive literature search that 147 studies yielded from Turkey, 28 from America, 23 from Europe, 12 from Asia, and 8 from Africa. We believe that the sample of this meta-analysis is highly representative of CCS studies in science education literature.

5.1 The Effect of CCS on Science Achievement

Our findings show that CCS has a large overall effect on science achievement. This result is consistent across all CCS types. Although the mean effect size for studies using ontological category shift is slightly smaller than others, the difference is neither statistically significant nor explains a considerable heterogeneity of true effect sizes. In other words, each type of CCS consistently has a large effect on science achievement. However, the distributions of effect sizes are highly heterogeneous, resulting in a wide prediction interval for the mean effect size for each type of CCS. Therefore, we conducted moderator analyses using some variables related to study characteristics.

The overall effect size indicating the effectiveness of CCS is consistent with previous meta-analyses in the literature (Armağan, 2011; Gelen, 2015; Guzzetti et al., 1993; Schroeder & Kucera, 2021). In other words, our findings, similar to previous meta-analyses, confirm that CCS significantly affects science achievement regardless of publication, sample, design, intervention, or measurement characteristics. Although the scope and sample of the prior meta-analyses are not the same as ours, the consistent findings revealed in all these meta-analyses provide strong evidence for the effectiveness of conceptual change strategies.

As we underlined before, what is unique about this meta-analysis is that we compared the relative effectiveness of each conceptual change strategy. Thus, one of the key findings in this meta-analysis is that each conceptual change strategy consistently has a large effect on science achievement. The meta-analyses of Armağan (2011) and Gelen (2015), focusing mainly on cognitive conflict, estimate the overall effect size to be 1.18 and 1.13, respectively. These values are very close to the one estimated for the cognitive conflict in this meta-analysis (g=1.10). However, in a recent meta-analysis, Schroeder and Kucera (2021) synthesized the studies on the refutational text and gave us an overall effect size of 0.45 for science, which is smaller than the one in this meta-analysis (g=1.09 for text-based CCS).

There could be several reasons for this discrepancy, one of which is the sample characteristics of the meta-analyses. Schroeder and Kucera reviewed 33 studies published in North America, whereas we synthesized 218 studies, 147 of which come from Turkey and 28 from North America. The overall mean value for studies from North America is very consistent with Schroeder and Kucera's findings. In this sense, findings are consistent when we consider the regional impact.

We also performed several analyses to examine the effect of the publication bias on our results. The high values that yielded from Rosenthal and Orwin's Fail-safe N show that the results were robust to publication bias. Furthermore, both selection models and Robust Bayesian Meta-Analysis estimate large adjusted effect sizes close to the unadjusted one. Thus, we conclude that publication bias does not considerably affect our interpretation of the results in this meta-analysis.

5.2 The Role of Study Characteristics on the Effectiveness of CCS

Main effect analyses are extremely important to draw a comprehensive conclusion about the effectiveness of conceptual change strategy. On the other hand, there is a significant distribution among effect values. The amount of this variation is measured through heterogeneity analyses, which inform us that there is significant variation between study variances. In this sense, there are prominent moderators that provide evidence on reasons for heterogeneity in effect size values. This practical knowledge improves understanding of the possible roles of moderators for the effectiveness of CCS in student achievement. We examined several moderator variables to explain the heterogeneity in true effect sizes. We categorized these variables into five broad characteristics: publication, sample, design, intervention, and measurement (Table 5.1).

In terms of the role of publication characteristics, publication type is one of the most criticized moderators in literature (Borenstein et al., 2009). As a theoretical perspective, published and unpublished studies yield different effect values due to the publication bias. This criticism implies that researchers have less tendency to publish studies with nonsignificant results. So, articles have a higher effect size value than unpublished studies like dissertations, master theses, or conference papers. On the other hand, the coding process of publication bias is not straightforward. Most of the conference papers, master theses, and doctoral dissertations could also be published. In this sense, it is not easy to code these sources as unpublished. Therefore, testing the theoretical argument on publication bias is a controversial issue. Nevertheless, the previous meta-analyses support this issue by reporting very divergent results for publication type moderator (Kaçar, 2021; Üstün, 2012). In the scope of this study, simple meta-regression analyses disclosed that publication type is not a significant moderator when included in the analysis alone. However, it turns out to be significant when it was analyzed with other variables using multiple meta-regressions (Table 4.84). Doctoral dissertations have a larger mean effect size than other publication types. We need to note that if a dissertation or thesis was also published in a journal, we included the dissertation or thesis since it generally provided us with much more detailed information about the study than journal articles. That is to say, the results revealed from some of these dissertations were also published in journals, but we still labeled them as dissertations. Thus, this classification does not directly represent comparing published and unpublished studies.

Meta-analytic studies have also criticized the publication year impact for publication characteristics (Armağan, 2011; Bayraktar, 2000). Initially, we have to say that the publication year expressed here is the year in which the treatment was applied. The mean effect size of the studies conducted between 2006 and 2010 is larger than the others. This result is also consistent when we consider simultaneously the theoretically important moderators like design and sample characteristics. Previous meta-analyses appear to show a similar pattern. For example, Guzzetti (1993), as one of the earliest meta-analyses on conceptual change, covers the studies between 1981 and 1991, but most of them were conducted in the second half of this time interval.

The findings of this meta-analysis indicate an overall effect size of 0.85. In addition, Armağan (2011) classified the primary studies into three groups separated by five years. The mean effect sizes are 0.98, 1.08, and 1.25 for the studies conducted between 1995-1999, 2000-2005, and 2006-2010, respectively. There could be several reasons for this trend, such as the developments in the conceptual change theory or improved experimental conditions in more recent studies. Yet, we do not have any empirical evidence to support these explanations in this meta-analysis.

Study quality is another important issue in meta-analysis literature. In any area of education literature, the quality of the intervention process may interact with the findings (Glass & Smith, 1977). Previous studies also infer that the better study designs reveal the true effect value more clearly (Chadwick, 1997). We aimed to investigate the effect of study quality by including moderators related to design characteristics like the design of the experiment and the teacher training. As a result of simple meta-regression, increasing study quality causes to a decrease in mean effect value. For instance, the quasi-experimental studies have a much larger mean effect size than the true-experimental studies. Unlike quasi-experimental design, true-experimental design requires random assignment of the subjects to treatment groups, which is an influential way to control threats to internal validity (Fraenkel et al., 2012, p. 266). Therefore, any difference between the mean effect sizes of true and quasi-experimental studies may result from the degree to which they control the threats to internal validity. And if this is so, we might be overestimating the overall effect size because of the weakness of the poor and quasi-experimental designs.

Furthermore, the studies reporting the teacher training reveal a larger mean effect size than those that did not. Teacher training is essential to boost implementation fidelity. Carroll et al. (2007) claim that the variations in the degree to which implementation fidelity is achieved in primary studies might explain the heterogeneity in their results. Additionally, the efficacy of a method is strictly related to the implementer's proficiency. In this sense, better teacher training result in a higher effect size value. It is vital that the implementer should be adequately trained in the implementation process to become evident of the effect of the method. From this perspective, our findings show that the conceptual change strategies work more effectively when the implementation fidelity is elevated through teacher training. Multiple meta-regression with different explanatory moderators also supports this evidence. That is, if we control other explanatory moderators with the training process one by one simultaneously, the training process is still influential on student science achievement.

Investigating the intervention characteristics is one of the main themes of this study. Subject domain is one of the most important moderating variables regarding the intervention characteristics. The studies in Chemistry have a considerably larger mean effect size than those in Physics and Biology. In contrast, the studies in Biology have the smallest mean effect size among these three subject domains. Multiple meta-regression with explanatory characteristics also supports this evidence. If we control other moderators with subject domain simultaneously, it is still effective on student achievement. So that the subject domain can be a pivotal independent moderator on student achievement for CCS. This result is also compatible with previous meta-analyses that used domain as a moderator in conceptual change strategy and problem-based methods (Arık & Yılmaz, 2020; Armağan, 2011; Gelen, 2015; Üstün, 2012). But there is no clear theoretical argument about why CCS works more effectively in literature in a specific subject domain. Thus, further qualitative and quantitative analyses could enable to disclosure of this issue more comprehensively.

Intervention length also moderates the effect of CCS on science achievement. One important hypothesis is that the short duration of the intervention causes a lack of deeper comprehension of knowledge for students (White, 1988). The effect size increases as the intervention length increases, indicating that it takes time to observe the actual impact of conceptual change strategies. Previous meta-analytic results on the conceptual change also support this finding (Chadwick, 1997; Gelen, 2015). On the other hand, controlling the other moderators with multiple meta-regression analyses is critical. Simultaneous analyses also show that length seems effective and has a significant and individual impact on treatment. This finding is reasonable because students and teachers may require time to get used to a new method, which is considerably different from what they usually experience in their classes.

Intervention intensity can be another moderator that makes sense to readers as theoretically significant in that more treatment dosage in a short period may yield higher student achievement. On the other hand, simple and multiple meta-regression analyses imply the ineffectiveness of treatment intensity. Intended results are broadly reasonable to reviews of recent studies attaining no consistent association between treatment intensity and student achievement (Kim et al., 2021; Kraft et al., 2018; Lynch et al., 2019). Therefore, it can be concluded that treatment intensity has no significant impact on student achievement.

The region is the only variable that explains a significant amount of heterogeneity in terms of sample characteristics. The studies conducted in Turkey have a significantly larger mean effect size than those in other studies. Previous meta-analyses on the conceptual change also support this finding. For example, most of the primary studies synthesized in Armağan (2011) and Gelen (2015) were conducted in Turkey. The overall effect sizes estimated in these meta-analyses are large and close to the one in this meta-analysis. Schroeder and Kucera (2021), on the other hand, cover the studies conducted in North America and estimate an overall effect size of 0.45 for science. As noted previously, this could be one of the reasons for the discrepancy between their findings and ours. This result is parallel to previous meta-analyses also (Arık & Yılmaz, 2020; Gözüyeşil & Dikici, 2014; Üstün, 2012). This moderator implies that conceptual change strategies work best in Turkey and worst in Europe and America. But, the critical question is why the region is an effective moderator. Simple meta-regression analyses only give rough evidence about the possible explanatory moderators. Analyses imply that the region moderator can also interact with other moderators. In this sense, multiple meta-regression could provide more comprehensive evidence of the region's effectiveness. For example, the studies in Turkey are mainly designed as quasi-experimental (88%), use objective-type questions for assessment (92%), and most teachers were trained by researchers (56%). On the other hand, in Europe and America, researchers have used less frequently quasi-experimental designs (17%, 39 %) and objective-type questions for assessment (28 %, 43%), and very few teachers were trained (27%, 44%). In this sense, region can be an over-effective moderator due to the interaction with these moderators, even if region seems most effective in a simple regression model. The effect of region can also be related to the intervention, design, and sample charactersitics. On the other hand, even though we control design and publication characteristics with region moderator, the region is still an effective moderator on student achievement. It is important to note that this variable remains significant in the final model created using multiple meta-regression analyses. In other words, it is a substantial moderator above and beyond the other variables in the model, which includes some other essential moderators. It would seem reasonable that this finding arises from another moderator we could not include in this meta-analysis.

One of the important intervention characteristics concerning simple meta-regression analysis is the material (computer, hands-on, or text-based). Material is a particular moderator for the scope of this study. Students can use different materials during intervention process like hands-on materials (laboratory activities, real tools, or chemicals), computer-based materials (simulation, demo, or video), or text-based materials (texts, maps, or drawings). Multiple intelligence theory proposes to include different types of materials to appeal to various attributes of students during educational interventions. So the properties of instructional tools may interact with student achievement. Some students may prefer computers, but some prefer real experimental environments. That is why the materials that students interact during the intervention can make a difference on student achievement. Simple meta-regression analyses support this theoretical argument by showing significant heterogeneity between different material types. This result implies that hands-on materials work better for student achievement than computer-based and text-based materials. Nevertheless, the multiple meta-regression analyses give more extensive evidence on this issue. This moderator is also affected by other moderators and has no significant and individual impact on the treatment effect. This final state of knowledge reveals that material does not significantly change CCS's effectiveness on student achievement.

Outcome measure type (misconception test or general achievement test) is also a special moderator for this study as measurement characteristics. Some primary studies use general achievement tests to measure the effect of CCS. On the other hand, conceptual change strategy aims to eliminate misconceptions rather than to grasp new knowledge. In this sense, it is expected that the studies that used a misconception test for assessment should provide a higher effect value than studies that used general achievement tests. As a result of simple meta-regression analyses, results derived from misconception tests had trivial differences from the achievement test scores. On the other hand, due to the moderator interactions, this hypothesis merits further scrutiny by carefully coding this issue to yield more valid analyses. For further investigation, we controlled explanatory moderators in multiple meta-regression. Consequently, there is no significant difference in student science achievement scores for different outcome measure types.

Education level (elementary, middle, high, university) should also be investigated for the effectiveness of different methods. In this sense, it was included in both simple and multiple meta-regression analyses to provide a more comprehensive understanding of this moderator. Some meta-analytic studies indicated that educational interventions work better in lower-level grades like preschool and elementary levels (Arık & Yılmaz, 2020; Kim et al., 2021). On the other hand, there are also inconclusive meta-analytic findings (Armağan, 2011; Deniz, 2019). In this study, the simple meta-regression analyses imply that CCS works better at the high school level. Nevertheless, the multiple regression analyses give more comprehensive evidence on this issue since explanatory moderators are taken into consideration simultaneously. This moderator is also affected by other moderators and has no significant and individual impact on the treatment effect. This final state of knowledge revealed that education levels do not significantly change student achievement.

Instrument type (adapted test, pre-existing test or researcher-developed test) is one of the effective moderators in literature. The validated standardized tests are more robust against random errors and researcher bias (Bayraktar, 2000). Therefore, it is expected to have a more negligible role in the treatment effect for systematic errors. Additionally, researcher-developed instruments are more purposeful and related to learning outcomes. In this sense, it is reasonable to expect differences between standardized and researcher-developed tests. According to simple meta-regression analyses, the CCS has the highest overall mean value if researchers use a researcher-developed test for assessment rather than other test types. There is significant heterogeneity between moderators. Nevertheless, multiple meta-regression with explanatory moderators did not support this evidence. Instrument type was not an effective moderator if we controlled study quality or publication characteristics. As a result, instrument types do not yield a significant impact on student achievement for CCS.

One of the explanatory moderators in sample characteristics is the sample size. As a theoretical argument, studies with a small sample size enable researchers to control moderators better and may increase treatment effect (Kulik et al., 1985). It is so that better controlling processes yield higher treatment scores (Chadwick, 1997). The simple meta-regression analyses informed that sample size did not impact treatment efficiency. The existing meta-analyses yielded controversial findings on this issue (Armağan, 2011; Bayraktar, 2000; Gelen, 2015). For the scope of this study, we controlled the effective characteristics (design and publication characteristics) simultaneously to observe the true effect of the sample size. The final state of knowledge revealed that sample size is not an effective moderator for CCS.

One of the effective moderators for simple meta-regression analysis is the question type for the assessment process (open-ended, objective type, or mix) representing measurement characteristics. The conceptual change strategy is affected by question type significantly. The mixed question type (including both open-ended and objective types) is also included in the analyses. The results imply that using objective-type questions during the assessment process enables to yield of higher effect value than open-ended and mix-type questions. The mix-type is also significantly higher than the open-ended question type. On the other hand, these results are not compatible with previous meta-analyses (Arık & Yılmaz, 2020; Üstün, 2012). In this sense, multiple meta-regression is crucial to yield more valid and comprehensive knowledge. Multiple meta-regression with explanatory moderators supports previous studies and implies that assessment with objective-type questions might overestimate the effectiveness of conceptual change strategies. The type of assessment instrument also appears to be significant, but it did not explain a considerable variance. That is, if we control design characteristics with question type moderator, it is not effective on student achievement. The final stage of investigation reveals that different question types in the assessment process do not cause any significant heterogeneity in student achievement for CCS.

Finally, the type of conceptual change strategy (cognitive conflict, cognitive bridging, or ontological category shift) is also critical and a special moderator for the scope of this study. Different conceptual change perspectives do not entirely reject each other but lead to a dispersion of the possible roles of prior knowledge through the conceptual change process. Different instructional implications of CCS yield very divergent effectiveness on achievement (Brown, 1995; Slotta & Chi, 2006; Smith et al., 1993; Tsai, 2003; Zohar & Kravetsky, 2005). Therefore, different instructional strategies of conceptual change perspective may have different profiles on student achievement and cause significant variation in effect size values. We derived three common types of strategy from the characteristics of 218 primary studies and theoretical knowledge introduced by Posner et al., 1982; Smith et al., 1993 and Chi et al. 1992; 1993; 1994; 2002; 2008. At the same time, we also validated this grouping using primary study authors' feedback, which included studies in this meta-analysis. The list of authors that provide feedback on this issue is also attached to the appendix. We also investigated the effect of strategies individually in the main effect analyses.Additionally, their impact on the heterogeneity in total variance was analyzed by using simple and multiple meta-regression analyses. These analyses reveal that there is no significant heterogeneity among these groups. This moderator is also affected by other moderators and has no significant and individual impact on the treatment effect. As a result, there is no significant difference between these strategies on science achievement.

The hierarchical multiple meta-regression analysis yielded a parsimonious model including six moderators, explaining 35.1% of the heterogeneity. This model allows us to estimate the effect of a particular moderator variable above and beyond the other variables in the model. In this model, quasi-experimental studies have a larger mean effect size than other studies even when we control for the type and year of the publication, the subject domain, the region where the studies were conducted, and whether the teachers were trained before the intervention.

To sum up, we identified some moderator variables affecting the effectiveness of conceptual change strategies on science achievement in terms of publication, sample, design, intervention, and measurement characteristics. Furthermore, we created a parsimonious model using multiple meta-regression to examine the unique moderating effect of some important variables. On the other hand, based on our meta-analysis, the conceptual change strategies significantly improve science achievement regardless of any study characteristics.

Table 5.1 Summary of the results yielded by simple meta-regression analyses

Study Characteristics	Variables	Heterogeneity				
		Q	p	T^2	I^2	R^2
Publication Characteristics	Publication Type	6.73	.089	.34	84.57	.02
	Publication Year*	22.59	.001	.32	83.30	.12
Sample Characteristics	Region*	48.33	<.001	.27	81.12	.24
	Sample Size	5.98	.200	.33	84.80	.05
	Class Size	5.74	.219	.34	84.31	.03
	Education Level*	9.27	.026	.34	84.43	.02
	School Location	1.27	.296	.36	85.34	<.01
	School Type	0.07	.790	.35	85.22	<.01
Design Characteristics	Experimental Design*	23.72	<.001	.33	83.77	.08
	Sampling Method	0.99	.319	.35	84.71	.01
	Researcher Effect	8.39	.573	.36	84.64	<.01
	Teacher Effect	9.68	.601	.37	85.12	<.01
	Treatment Verification	0.32	.570	.35	84.84	<.01
	Teacher Training*	16.68	.000	.32	83.52	.10
Intervention Characteristics	Type of CCS	0.95	.622	.35	84.53	<.01
	Material*	6.16	.046	.34	84.50	.02
	Subject Domain*	28.27	<.001	.30	82.73	.14
	Intervention length*	15.81	<.001	.31	83.46	.11
	Intervention Intensity	0.33	.564	.37	85.58	<.01
Measurement Characteristics	Instrument Type*	3.76	.035	.35	84.42	.02
	Question Type*	2.95	<.001	.32	82.24	.14
	Number of Tiers	3.12	.215	.34	84.50	<.01
	Type of Outcome Measure	2.04	.165	.36	85.07	<.01

Q: Total heterogeneity; T^2: Between study variance; I^2: Ratio of the true variance to the total variance; R^2: Ratio of the explained variance to the total variance.
*The moderators explain a significant amount of heterogeneity.

5.3 Implications for Theory and Practice

In earlier sections, we limited ourselves to discussions of the results of quantitative analyses. In this section, we will question our findings from the theoretical perspective and discuss some issues from a practical point of view. One of our frustrations in this meta-analysis is that we used only science achievement as an outcome variable because most studies focused on this variable. However, the theoretical framework of each strategy gives different messages about how to learn science. Cognitive bridging, for example, appreciates what is already known, while cognitive conflict does the opposite. This may cause a difference in students' epistemologies about how to learn science. Similarly, several motivational factors, especially self-efficacy beliefs, can be affected by consistent falsification of what students already know during implementing cognitive conflict strategy (Smith et al., 1993). However, few studies focus on motivational or epistemological factors as outcome variables. Although the three conceptual change strategies do not significantly differ in the degree to which they affect students' science achievement, they may still diverge considering their effects on other outcome variables.

Treatment fidelity emerges as another issue for the available primary studies in the literature on conceptual change, as argued by researchers in other domains (Moncher & Prinz, 1991; Swanson et al., 2013). Unfortunately, the primary studies hardly provide evidence about the fidelity of the treatments. All of the conceptual strategies require carefully designed procedures for implementation. In the case of cognitive conflict strategy, anomalous data or discrepant events have been used to trigger cognitive conflict. Nevertheless, several researchers reported that the intended conflict might not be achieved (Chan et al., 1997; Chinn & Brewer, 1993). According to these researchers, students' responses to anomalous data can range from unadopted to comply with anomalous data. Even though carefully selected anomalous data cannot guarantee a meaningful conflict, it is the best tool to initiate a conflict. Some researchers have included data about how students' responses to anomalous data contribute to their gains from the cognitive conflict strategy (Lee & Byun, 2012). We need more of this kind of study to perform analyses examining the heterogeneity of the effects of cognitive conflict strategies.

Similarly, the cognitive bridging strategy relies on analogies, especially anchoring/bridging analogies. Anchoring analogies have some unique qualities. Furthermore, calling them analogies can be misleading because they are not analogies but the concept itself. For example, the analogy of "a hand pressing on a spring" is used to eliminate students' misconception that "inanimate objects do not exert force." This analogy was used in the preliminary study by Clement et al. (1989) to overcome students' misconception about the reaction force acting on a book resting on a table. Without going into details, both analog (the force acting on a hand pushing the spring) and target concepts (the force acting on a book resting on the table) are related to the same concept, the reaction force. The difference is that the force on the hand exerted by the spring is more noticeable for students compared to the force on the book by the table. However, in traditional analogies, the analog and the target concepts are different from each other. When using these analogies, it is critical to discuss the dissimilarities and the similarities between the analog and the target concepts. It is always possible for students to pick up new misconceptions out of the traditional analogies if the necessary precautions are not taken (Spiro et al., 1988). However, this is not the case for anchoring "analogies" because they are not analogies. In primary studies, we encounter traditional analogies as well as anchoring ones. More detailed discussions about the similarities and differences between the analog and the target concepts are needed.

On the other hand, ontological category shift seems to be the most novel approach among the conceptual change strategies. This is probably why the number of studies testing ontological category shifts' effectiveness is far smaller than the ones for the other strategies. Furthermore, most of the studies on ontological categories focus on the concepts already categorized by Chi and her colleagues (1992; 1993; 2002).

Identifying the ontological categories of different concepts seems to be one of the challenging tasks of this strategy. Several researchers engaging in such a task, especially for physics concepts, have been arguing about the difficulties of locating them into distinct categories (Chiu et al., 2005; Çoruhlu & Çepni, 2015; Yang et al., 2012). All these discussions are fruitful in creating awareness about the ontological nature of concepts and providing feedback for the ontological category theory.

In terms of instructional practices, it is evident that students have misconceptions that are hard to eliminate. In this sense, the most efficient way of overcoming this barrier is using the CCS instead of traditional strategies regardless of subject, material, and conceptual change strategy. The effectiveness of the CCS is not related to the intensity of the process. It is more important how many course hours are applied. For this reason, it is important to conduct more instruction in optimum time intervals rather than in short-term, intensive applications. Furthermore, the effectiveness of instruction does not change in large or small classes, contrary to common belief, but the effect of other variables such as teacher's qualification and sample characteristics cannot be ignored. Thus, teachers should not avoid practicing CCS in large classes if there is no classroom management or insufficient material problem.

5.4 Limitations and Future Directions

The first limitation of this meta-analysis is the language of the primary studies we included. We can generalize the results of this meta-analysis only to the studies published in either English or Turkish. In addition, any meta-analysis is limited by the scope of the primary studies synthesized in that meta-analysis. We could only examine the effect of conceptual change strategies on science achievement since most primary studies focus on the cognitive dimension rather than the effective one. However, conceptual change strategies may affect the motivational constructs as well. For example, cognitive conflict may affect the students' self-efficacy due to the dissatisfaction it aims to create. Furthermore, it may also affect students' epistemology since it focuses on unproductive preconceptions rather than productive ones. Yet, the outcomes of the conceptual change strategies in the affective domain are often neglected in the literature. Another limitation of any meta-analysis is the potential effect of publication bias on the findings. However, several analyses show that our results seem robust to the publication bias. Therefore, it would not change our interpretation of the overall effect size in this meta-analysis. Finally, the findings revealed by the moderator analyses should be interpreted cautiously. Because of the associational nature of these analyses, we cannot claim a cause-effect relationship directly.

Each type of conceptual change strategy appears to be effective when it is used in an appropriate context. Yet, there is a need for studies comparing the effectiveness of these strategies on the same misconception. Since each conceptual change strategy is based on a different approach, its effectiveness might change according to the nature of the misconception. In addition, the conceptual change strategies may have different effectiveness measured by the retention test. Further studies can be conducted to examine the delayed effects of these strategies.

5.5 Conclusions

In the scope of this comprehensive meta-analytic study, the provided knowledge derived from simple and multiple meta-regression analyses infer significant conclusions for researchers and policymakers. Even though the intended moderator significantly impacts simple meta-regression, the final state of knowledge has reached by using both multiple meta-regression results and theoretical knowledge. We briefly tabulated our findings in Table 5.2. Totally 218 primary studies (18,051 individual samples), including articles, dissertations, thesis, and conference papers, were utilized for the following conclusions;

- CCS has a large effect on science achievement compared to traditional teaching methods.
- Cognitive conflict strategy has a large effect on science achievement when compared to traditional teaching methods.
- Cognitive bridging strategy has a large effect on science achievement when compared to traditional teaching methods.
- Ontological category shift strategy has a large effect on science achievement when compared to traditional teaching methods.
- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by different regions such as Africa, America, Asia, Europe, and Turkey.
- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by subject domain: chemistry reveals a larger effect size than physics and biology.

- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by different publication types: doctoral dissertations reveal a larger effect size than articles, master thesis, and conference papers.

- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by different publication year intervals: studies done between 2010-2014 yield higher effect values than other year intervals.

- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by intervention length: the longer the intervention length provides the higher the effect size values.

- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by teacher training: when implementers are trained, it reveals a larger effect size compared to unstated training process conditions.

- The estimated effect size values for the effectiveness of CCS on student achievement are moderated by experimental design: quasi-experimental designs provide larger effect sizes than poor experimental designs.

- There is no statistical evidence for the effectiveness of the type of CCS, sample size, class size, school type, school location, question types, instrument types, educational levels, materials, sampling method, number of tiers, type of outcome measure, researcher effect, teacher effect, treatment verification, or treatment intensity on student achievement.

Table 5.2 The results of simultaneous analyses and conclusions

Code Character	Code Name	Regression Results	Simple R^2	Conclusions
Publication Characteristics	Publication Type	Ineffective in Simple MR, effective in MR	0.016	Studies done as doctoral dissertations provide higher effect value than other types (article, master thesis, proceeding)
	Publication year	Effective in simple and multiple MR	0.120	Studies done at different year intervals (1989-2000, 2001-2005, 2006-2010, 2011-2015, 2016-2020) have an impact on the effectiveness of CCS
Sample Characteristics	Region	Effective in simple and multiple MR	0.235	Studies in Turkey yield higher effectiveness than in other regions (Africa, America, Asia, Europe).
	Sample Size	Effective in simple MR and ineffective in multiple MR	0.025	Sample size interval (16-46, 47-56, 57-72, 73-100, 102-396) has no impact on the effectiveness of CCS
	Class Size	Ineffective for both Simple and Multiple MR	0.030	Class size interval (8-22, 23-26, 27-30, 30-38, 38-87) has no impact on the effectiveness of CCS
	Educational level	Effective in simple MR and ineffective in multiple MR	0.022	Educational level (elementary, middle, high school, university) has no impact on the effectiveness of CCS
	School Location	Ineffective for both Simple and Multiple MR	0.000	School location (rural, urban) has no impact on the effectiveness of CCS
	School Type	Ineffective for both Simple and Multiple MR	0.000	School type (public, private) has no impact on the effectiveness of CCS
Design Characteristics	Experiment design	Effective in simple and multiple MR	0.081	CCS is more effective in quasi-experimental designs (poor, quasi, true)
	Sampling Method	Ineffective for both Simple and Multiple MR	0.008	The sampling method (random, non-random) has no impact on the effectiveness of CCS
	Researcher Effect	Ineffective for both Simple and Multiple MR	0.000	The researcher (one of the teachers, only teacher, not the teacher) has no impact on the effectiveness of CCS

Table 5.2 (Continued)

	Teacher Effect	Ineffective for both Simple and Multiple MR	0.000	Teacher (same teacher, different teacher) has no impact on the effectiveness of CCS
	Treatment Verification	Ineffective for both Simple and Multiple MR	0.000	Verification (stated, unstated) has no impact on the effectiveness of CCS
	Teacher Training *	Effective in simple and multiple MR	0.095	CCS is more effective when teachers were trained.
Intervention Characteristics	Type of CCS	Ineffective for both Simple and Multiple MR	0.000	Type of CCS (cognitive conflict, cognitive bridging, ontological category shift) has no impact on the effectiveness of CCS
	Material	Effective in simple MR and ineffective in multiple MR	0.018	Material(text-based, hands-on, computer-based) has no impact on the effectiveness of CCS
	Subject Domain*	Effective in simple and multiple MR	0.140	Chemistry reveals a larger effect size than physics, and physics has larger values than biology.
	Intervention length*	Effective in simple MR and multiple MR	0.112	CCS is more effective when intervention length increases (1- 48 course hours)
	Intervention Intensity	Ineffective for both Simple and Multiple MR	0.000	Intervention intensity (1-8 course hours per week) has no impact on the effectiveness of CCS
Measurement Characteristics	Outcome Measure Type	Ineffective for both Simple and Multiple MR	0.002	Measuring outcome (general achievement test, misconception test) has no impact on the effectiveness of CCS
	Assessment Instrument Type	Effective in simple MR and ineffective in multiple MR	0.023	Instrument type (researcher-developed, adapted, pre-existed) has no impact on the effectiveness of CCS
	Question type	Effective in simple MR and ineffective in multiple MR	0.080	Question type (open-ended, objective, mix) has no impact on the effectiveness of CCS
	Number of tiers	Ineffective for both Simple and Multiple MR	0.003	The number of tiers (one, two, three) has no impact on the effectiveness of CCS

REFERENCES

Abdullah, M.F., Ibrahim, M., & Zulkifli, H. (2017). Resolving the misconceptions on big data analytics implementation through the government research institute in Malaysia. *IoTBDS*, 261-266. https://doi.org/10.5220/0006293902610266

Alvi, M. (2016). A manual for selecting sampling techniques in research. https://mpra.ub.uni-muenchen.de/60138/

Appelbaum, M., Cooper, H., Kline, R.B., Mayo-Wilson, E., Nezu, A.M., & Rao, S.M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *The American Psychologist, 73*(1), 3-25.

Arık, S., & Yılmaz, M. (2020). The effect of constructivist learning approach and active learning on environmental education: A meta-analysis study. *International Electronic Journal of Environmental Education, 10*(2), 44-84.

Armağan, F. Ö. (2011). *Kavramsal değişim metinlerinin etkililiği: Meta-analiz çalışması* (Publication No. 279751) [Doctoral dissertation, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

Arthur, W., Bennett, W., & Huffcutt, A. I. (2001). Conducting meta-analysis using SAS. Mahwah, NJ: Erlbaum.

Ausubel, D.P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

Bahar, M. (2003). Misconceptions in biology education and conceptual change strategies. *Kuram ve Uygulamada Eğitim Bilimleri / Educational Sciences: Theory & Practice, 3*(1),55-64.

Bai, R., Lin, L., Boland, M.R., & Chen, Y. (2021). A robust bayesian copas selection model for quantifying and correcting publication bias. *arXiv: Methodology*. https://arxiv.org/abs/2005.02930

Bailar, J.C. (1995). The practice of meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 149– 157.

Baillargeon, R. (2004). Infants' reasoning about hidden objects: evidence for event-general and event-specific expectations. *Developmental Science, 7*(4), 391-414

Balcı, A. (2012). Türkiye'de çocuk edebiyatı üzerine hazırlanan lisansüstü tezler hakkında bir meta-analiz çalışması. *Mustafa Kemal University Journal of Social Sciences Institute, 9*(17), 195-206.

Barke, H. D., Hazari, A., & Yitbarek, A. (2009). *Misconceptions in chemistry: addressing perceptions in chemistry*. Springer.

Bartoš, F. Maier, M , Quintana, D. S., & Wagenmakers, E. J. (2020). PEESE, and robust bayesian meta-analysis. Adjusting - selection models, PET- for publication bias in JASP & R Project. 1-31. https://doi.org/10.31234/osf.io/75bqn

Bayraktar, Ş. (2000). *A meta-analysis study on the effectıveness of computer-assisted ınstructıon in science education* (Publication No. 398527) [Doctoral dissertation, Ohio State University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

Beerenwinkel, A., Parchmann, I., & Gräsel, C. (2011). Conceptual change texts in chemistry teaching: A study on the particle model of matter. *International Journal of Science and Mathematics Education, 9*(5), 1235–1259.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias for meta-analysis: prevention, assessment and adjustments*. John Wiley & Sons Ltd.

Borenstein, M. (2009). Effect size for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), T*he handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Borenstein, M., & Hedges, L. V. (2019). The effect size for meta-analysis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.). *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 207-245). Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). Comprehensive Meta-analysis, Version 3.0. [Computer software]. Biostat. https://www.meta-analysis.com

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9,* 333–342.

Brown, D. E. (1995, April). *Theories in pieces? The nature of students' conceptions and current issues in science education* [paper presentation], National Association for Research in Science Teaching, San Francisco, CA.

Burbules, N. C., & Linn, M. C. (1988). Response to contradiction: Scientific reasoning during adolescence. *Journal of Educational Psychology, 80,* 67-75.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition and Instruction, 9,* 117-123.

Carey, S. (2008). The *origin of concepts. The process of conceptual change.* MIT Press.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science: IS, 2,* 40 - 40.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144.

Chadwick, D. K. H. (1997). *Computer-assisted instruction in secondary mathematics classrooms: a meta-analysis* (Publication no. 9809425) [Doctoral dissertation, Drake University]. ProQuest Dissertation. https://www.proquest.com

Chamber, S. K., & Andree, T. (1995). Are conceptual change approaches to learning science effective for everyone? Gender, prior subject, matter interest and learning about electricity. *Contemporary education psycholog*. *20*(4), 377-391. https://doi.org/10.1006/ ceps.1995.1025

Chan, C., K., K. (1993). Effects of conflict and knowledge-building approach on conceptual change [Unpublished doctoral dissertation]. The University of Toronto.

Chan, C., Burtis J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15*(1), 1-40.

Chen, C., & Yang, Y. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review, 26,* 71-81. https://doi.org/10.1016/j.edurev.2018.11.001

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Eds.), *Cognitive models of science: Minnesota studies in the philosophy of science*, (pp.129-186). University of Minnesota Press.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In Vosniadou, S. (2008). *International handbook of research on conceptual change.* (1st ed., pp. 61-83). Taylor & Francis.

Chi, M. T. H., & Roscoe, R. D. (2002).The process and challenges of conceptual change. In M. Limón & L. Mason (Eds.). *Reconsidering conceptual change: Issues in theory and practice,* (pp. 3-27). Kluwer Academic Publishers.

Chi, M. T. H., & Slotta, J. D., (1993). The ontological coherence of intuitive physics. *Cognition and Instruction, 10*(2-3), 249-260.

Chi, M. T.H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction,* 4(1), 27-43. https://doi.org/10.1016/0959-4752(94)90017-5

Chinn, C.A., & Brewer, W.F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1 - 49. https://doi.org/ 10.2307/1170558

Choi, S. W., & Lam, D.M.H. (2015). Statistically speaking: Funnels for publication bias – have we lost the plot? *Anaesthesia, 71*(3), 338-341

Christmann, E.P., Badgett, J.L., & Lucking, R.A. (1997). Microcomputer-based computer-assisted ınstruction within different subject areas: A statistical deduction. *Journal of Educational Computing Research, 16*, 281 - 296.

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research*. *53*(4). 445-459. https://doi.org/10.3102/00346543053004445

Clement, J., Brown, D. E., & Zietsman, A. (1989). *Not all preconceptions are misconceptıons: Finding "anchoring conceptions for grounding instruction on students' intuitions* [paper presentation], American Educational Research Association, SanFrancisco, CA, 27-31. https://doi.org/10.1080/0950069890110507

Cleophas, T.J., & Zwinderman, A.H. (2017). *Modern meta-Analysis: Review and update of methodologies*. Springer International Publishing

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates Publishers.

Cooper, H. (2017). *Research synthesis and meta- Analysis: A step-by-step approach, (*5th ed.*).* Sage Publications.

Cooper, H., & Hedges, L.V (2019). Formulating a problem. In Cooper, H., Hedges, L.V., & Valentine, J.C. *The handbook of research synthesis and meta-analysis*, (3th ed. pp.17-48). Russell Sage Foundatıon.

Cooper, H., Hedges, L. V., & Valentine, J. C., (2019).*The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

Cumming, G. (2012). *Understanding the new statistics*: *Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Çil, E., & Çepni, S. (2012). The effectiveness of the conceptual change approach, explicit reflective approach, and course book by the ministry of education on the views of the nature of science and conceptual change in light unit. *Kuram ve Uygulamada Egitim Bilimleri, 12*, 1107-1113.

Dechartres, A.,, Ludovic, T., TimorF., & Philippe R. (2016). Empirical evaluation of which trial characteristics are associated with treatment effect estimates. *Journal of Clinical Epidemiology, 77*(1), 24–37.

diSessa, A., A. (1993). Responses. *Cognition and Instruction, 10* (2 &3), 261-280.

diSessa, A. A. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10). 1155-1191. https://doi.org/10.1080/ 0950069980201002

diSessa, A. A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice*, 29-60. Dortrecht: Kluwer.

diSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy. In Vosniadou, S. (Eds.). *International handbook of research on conceptual change*. (pp. 35-60). Taylor & Francis.

Driscoll, M. P. (2000). *Psychology of learning for instruction* (2nd ed.). Allyn & Bacon.

Draper, D., Light, R.J., & Pillemer, D.B. (1987). Summing up: The science of reviewing research. *Journal of the American Statistical Association, 82*, 349.

Dreyfus, A., Jungwirth, E., & Eliovitch, R. (1990). Applying the cognitive conflict strategy for conceptual change - some implications, difficulties, and problems. *Science Education, 74*, 555-569.

Driver, R. (1992). Pupils alternative frameworks in science. *European Journal of Science Education, 18*, 365-392.

Driver, R., & Easley, J. (1978). Pupils and paradigm. A review of literature related to the concept development in adolescent science student. *Studies of Science Education, 5*, 61-84.

Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change. In S. Vosniadou (Eds.), *International handbook of research on conceptual change* (pp. 629-647) . Routledge.

Durlak, J. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917-928.

Dusek, J. B. (1971). Experimenter bias in performance in children at a simple motor task. *Developmental Psychology, 4*(1p1), 55.

Duval, S., & Tweedie, R. (2000a). A nonparametric" trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89-98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Duval, S. (2005). The Trim and Fill Method. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. (pp. 127-143). John Wiley & Sons.

Eggen, P., & Kauchak, D. (2004). *Educational psychology: Windows, classrooms.* Pearson Prentice Hall.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge University Press.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice,40* (5), 532–538.

Fisher, K. M. (1985). A misconception in biology: Amino acids and translation. *Journal of Research in Science Teaching, 22* (1), 562.

Fletcher-Flinn, C.M., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*, 219 - 241.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Freedman, M. P., (1997). Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching, 34*, 343- 357.

Geban, Ö., & Bayır, G. (2000). Effect of conceptual change approach on students understanding of chemical change and conservation of matter. *Hacettepe University Journal of Education, 19*, 79-84.

Gelen, B (2015). *Fen eğitiminde kavramsal değişim yaklaşımına dayalı öğretim tekniklerinin öğrencilerin kavramsal anlamaları üzerine etkisi: meta analiz çalışması* (Publication No. 616353) [Master thesis, Istanbul University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi].

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Glass, G. V. (2006). Meta-analysis: The quantitative synthesis of research findings. Green, J.L., Camilli, G., Elmore, P.B., Skukauskaitė, A., & Grace, E. (Eds.). *Handbook of complementary methods in education research.* (1st ed.), 427-439. Lawrence Erlbaum Assocıates

Glass, G. & Smith, M., (1978). Meta-analysis of research on the relatıonshıp of class size and achıevement. The class size and instructıon project. *AERA, 1*(1), 2-19. https://doi.org/10.2307/1164099

Gleser, L.J., & Olkin, I. (1996). Models for estimating the number of unpublished studies. *Statistics in medicine, 15 23*, 2493-507.

Gravetter, F. and Wallnau, L. (2014) *Essentials of statistics for the behavioral* sciences (8th ed.). Cengage Learning.

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ: British Medical Journal, 331*, 1064 - 1065.

Griffiths, A. K., & Preston, K. R. (1992). Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. *Journal of Research in Science Teaching, 29*(6), 611-628.

Gronau, Q.F., Heck, D.W., Berkhout, S.W., Haaf, J.M., & Wagenmakers, E. (2021). A primer on bayesian model-averaged meta-analysis. *Advances in Methods and Practices in Psychological Science, 4*.

Gupta, A., Hammer, D., & Redish, E. F. (2010). The case for dynamic models of learners' ontologies in physics. *Journal of the Learning Sciences, 19*, 285–321. https://doi.org/10.1080/10508406.2010.491751

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*(2), 116-159. https://doi.org/10.2307/747886

Haidich, A. (2010). Meta-analysis in medical research. *Hippokratia. 14* (1), 29-37.

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perceptions of cognitive structure influence instructional perceptions and intentions. *The Journal of Learning Science, 5*(2), 97-127. https://doi.org/10.1207/s15327809jls0502_1

Harrel, F. E. (2015). Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis (2nd ed.). Springer International Publishing.

Hawkins, D. (1985). The nature of the problem. In M. Apelman (Ed.), *Critical barrier phenomenon in elementary science*. University of North Dakota, Center for Teaching and Learning.

Hedges, L.V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107 - 128.

Hedges, L. V., & Olkin, I. (1985).*Statistical methods for meta-analysis*. Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504.

Hedges, L. V., & Vevea, J. L. (2005). Selection method approaches. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 145-173). John Wiley & Sons.

Helier, M.P., & Finley, N.F. (1992).Variable uses of alternative conceptions: A case study in current electricity. *Journal of Research in Science Education, 29*(3), 259-276.

Heng, C.K., & Karpudewan, M. (2017). Facilitating primary school students' understanding of water cycle through guided inquiry-based learning. In Karpudewan, M., Zain, A.N., & Chandrasegaran, A. (Eds.), *Overcoming students' misconceptions in science*. (1st Ed., pp. 29-51). Springer Nature

Hewson, P.W. (1992, June). *Conceptual change ın scıence teachıng and teacher educatıon* [paper presentation]. Research and Curriculum Development in Science Teaching, Madrid, Spain.

Hewson, P. W., & Hewson, M. G. (1989). Analysis and use of a task for identifying conceptions of teaching science. *Journal of Education for Teaching, 15*(3), 191-209. https://doi.org/10.1080/0260747890150302

Higgins, J.P., & Green, S. (2019). C*ochrane handbook for systematic reviews of ınterventions*. The Cochrane Collaboration.

Higgins, J. P. T., Thompson, S. G, Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Education and debate, 327*, 557-560. https://doi.org/10.1136/bmj.327.7414.557

Horton, P., McConney, A., Gallo, M.A., Woods, A.L., Senn, G.J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education, 77*, 95-111.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*(2), 193-206.

Hunt, M. (1997), *How science takes stock: the story of meta-analysis*. Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage Publications.

Hynd, C. R., & Alvermann, D. E. (1986).The role of refutation text in overcoming difficulty with science concepts. *Journal of Reading, 29,* 440–446.

Illeris, K. (2018). An overview of the history of learning theory. *European Journal of Education, 53*, 86-101.

Ivowi, U. M. O. (1984). Misconception in physics amongst Nigerian secondary school students. *Physics Education, 19,* 279-285.

JASP Team (2022). JASP (Version 0.16.1)[Computer software]. https://jasp-stats.org/

Kaçar, T., Terzi, R. Arıkan, I., & Kırıkçı, C. A. (2021). The effect of inquiry-based learning on academic success: a meta-analysis study. *International Journal of Education & Literacy Studies, 9*(2), 15-23. https://doi.org/ 10.7575/aiac.ijels.v.9n.2p.15

Kaltakci, D., & A. Eryilmaz (2010). Sources of optics misconceptions. In contemporary science education research: learning and assessment, edited by G. Çakmakçı and M. F. Taşar, 13–16. Ankara: Pegem Akademi

Karakuş, M. & Öztürk, H. Ġ. (2016). Türkiye'de uygulanan işbirliğine dayalı öğrenme yönteminin fen bilimleri öğretiminde akademik başarı ve derse karşı tutumlar üzerindeki etkisini incelemeye yönelik bir meta-analiz çalışması. *International Journal of Active Learning (IJAL), 1* (1), 1-28.

Karpudewan, M., Zain, A. N., & Chandrasegaran, A. (2017). *Overcoming students' misconceptions in science.* Springer Nature

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling.* Taylor & Francis Group.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology, 40*(1), 1-33.

Kim, J. S., Gilbert, J., Yu, Q., & Gale, C. (2021). Measures matter: a meta-analysis of the effects of educational apps on preschool to grade-3 children's literacy and math skills. *AERA Open*. *7*(1), 1–19.

Kintz, B.L., Delprato, D.J., Mettee, D.R., Persons, C.E., & Schappe, R.H. (1965). The expermenter effect. *Psychological Bulletin, 63,* 223-32.

Koparan, T., Yıldız, C., Köğce, D., & Güven, B. (2010). The effect of conceptual change approach on 9th-grade students' achievement. *Procedia Social and Behavioral Sciences, 2*(2), 3926–3931.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of educational research, 88*(4), 547-588.

Kuhn, T. (1970). *The structure of scientific revolutions,* second edition, enlarged. University of Chicago Press.

Kulik, J.A., Bangert, R.L., & Williams, G.W. (1983). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology, 75*, 19-26.

Kulik, C.C., & Kulik, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7,* 75-94.

Kulik, J.A., Kulik, C.C., & Bangert-Drowns, R. (1985). Effectiveness of computer-based education in secondary schools. *The Journal of Computer Based Instruction, 12,* 59-68.

Kulik, J.A., Kulik, C.C., & Cohen, P.A. (1980). Effectiveness of computer-based college teaching: a meta-analysis of findings. *Review of Educational Research, 50,* 525 – 544

Kulik, J.A., Kulik, C.C., & Cohen, P.A. (1979). Research on audio-tutorial instruction: A meta-analysis of comparative studies. *Research in Higher Education, 11,* 321-341.

Lasserson T. J. , Thomas, J., & Higgins, J. P. (2019). Starting a review: Why do a systematic review? Higgins, J.P., & Green, S. (Ed.). Cochrane Handbook for Systematic Reviews of Interventions. 3-12. Oxford: The Cochrane Collaboration.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: effects of guidance. *Review of Educational Research, 86*(3), 681-718.

Leach, J., & Scott, P. (2002). Designing and evaluating science teaching sequences: an approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education, 38*, 115 - 142.

Lee, G., & Byun, T. (2012). An explanation for the difficulty of leading conceptual change using a counterintuitive demonstration: The relationship between cognitive conflict and responses. *Research in Science Education, 42*, 943-965. https://doi.org/10.1007/s11165-011-9234-5

Lee, J. (1999). Effectiveness of computer-based ınstructional simulation: A meta-analysis. *International journal of instructional media, 26*, 71-85.

Liao, Y. K. C. (1999). Hypermedıa and students' achıevement: a meta-analysıs. *EdMedia+ Innovate Learning*, 1398-1399.

Limon, M. (2001) On the cognitive conflict as an ınstructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*, 357-380. https://doi.org/10.1016/S0959-4752(00)00037-2

Linder, C. J. (1993). University physics students' conceptualizations of factors affecting the speed of sound propagation. *International Journal of Science Education, 15* (6), 655-662.

Linnenbrink, E.A., & Pintrich, P.R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading and Writing Quarterly,* 19, 119-137

Lipsey, M. W., & Wilson, D. B. (2001).*Practical meta-analysis*. Sage Publications.

Magliocca, N.R., Rudel, T.K., Verburg, P.H., Mcconnell, W.J., Mertz, O., Gerstner, K., Heinimann, A., & Ellis, E.C. (2014). Synthesis in land change science: methodological patterns, challenges, and guidelines. *Regional Environmental Change, 15*, 211 - 226.

Maier, M.A., Bartoš, F., & Wagenmakers, E. (2020). Robust bayesian meta-analysis: Addressing publication bias with model-averaging. http/doi.org/10.31234 /osf.io/u4cns

Maksimović, J. (2011). The application of meta-analysis in educational research. *Philosophy, Sociology, Psychology and History*,*10*(1), 45 – 55.

Marczyk, G.R., DeMatteo, D., & Festinger, D.S. (2005). *Essentials of research design and methodology*. John Wiley & Sons, Inc.

Martin, R., Sexton, C., & Gerlovich, J. (2002) Teaching science for all children: methods for constructing understanding.[Unpublished doctoral dissertation Boston Univesity], Boston.

Matt, G. E., & Cook T. D.,(2019). Threats to the valıdıty of generalızed ınferences from research syntheses. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.).*The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. Chapter 22. P. 490-510.

McGiverin, J., Gilman, D.A., & Tillitski, C.J. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal, 90*, 47 - 56.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247–266. https://doi.org/10.1016/0272-7358(91)90103-2

Mufit, F., Asrizal, A., & Puspitasari, R.D. (2020). Meta-analysis of the effect of cognitive conflict on physics learning. *Journal Penelitian & Pengembangan Pendidikan Fisika*. *6*(2), 267, 278. https://doi.org/10.21009/1.06213

Mullen, B., Muellerleile, P., & Bryant, B. (2001). Cumulative meta-analysis: A consideration of indicators of sufficiency and stability. *Personality and Social Psychology Bulletin, 27*(11), 1450–1462.

Mungsing, W. (1993).Students' Alternative Conceptions about Genetics and The Use of Teaching Strategies For Conceptual Change. [Unpublished doctoral dissertation, University of Alberta], Ottowa.

Nelmes, A.J. (2005). Overcoming misconceptions: using bridging analogies to cue scientific ideas.[Unpublished doctoral dissertation, Loughborough University], British.

Onuoha, O. C. (2007). Meta-analysis of the effectiveness of computer-based laboratory versus traditional hands-on laboratory in college and pre-college science instructions. [Unpublished Doctoral Dissertation, Capella University], USA.

Orwin, R. G. (1983). A fail-safe *N* for effect size. *Journal of Educational Statistics. 8*, 147–159.

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.). *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 177-207). Russell Sage Foundation.

Osborne, R. J. (1982). Science education: Where do we start? *Australian Science Teachers Journal, 28*(1), 21-30.

Özdemir, Ö. F., (2004). *The coexistence of alternative and scientific conceptions in physics*. (Publication No. 400155) [Doctoral dissertation, Ohio State University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

Özdemir, Ö. F., (2015). A qualitative inquiry about students' conceptualizations of force concept in terms of ontological categories. *Journal of Turkish Science Education*. *12*(1), 29-42.

Özdemir, G. & Clark, D., (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science and Technology Education*. 3, 351-361.

Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J.M., Grimshaw, J., Hrõbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P.F., & McKenzie, J.E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and examples for reporting systematic reviews. https://doi.org/10.1136/bmj.n160

Peters, J.L., Sutton, A., Jones, D.R., Abrams, K.R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in medicine*, *26 25*, 4544-62.

Petticrew, M. & Roberts, H. (2006). Starting the review: Refining the question and defining the boundaries. *Systematic Reviews in the Social Sciences: A practical guide*. Oxford: Blackwell Publishing.

Pfundt, H., & Duit, R. (1991). Bibliography: Students' Alternative Frameworks and Science Education. 3rd ed. Kiel, Institute for Science Education, Univerity of Kiel.

Piaget, J. (1964). Cognitive development in children: development and learning. *Journal of Research in Science Teaching*, 2, 176-186. https://doi.org/10.1002/tea.3660020306

Pigott, T.D. (2012). *Advances in meta-analysis*. Springer Verlag.

Pigott, T. D. (2019). Data diagnostic. In Cooper, H., Hedges, L.V., & Valentine, J.C. *The Handbook of research synthesis and meta-analysis*, (3th ed.) 367-430. Russell Sage Foundatıon.

Posner, G. J., Strike, K. A., Hewson, P.W., & Gertzog, W. A. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227. https://doi.org/10.1002/sce.3730660207.

Rahim, R. Noor, N & Zaid, N. M. (2015). Meta-analysis on element of cognitive conflict strategies with a focus on multimedia learning material development. *International Education Studies, 8*(13), 73-78.

Rendina-Gobioff, G. (2006). *Detecting publication bias in random-effects meta-analysis: An empirical comparison of statistical methods* [Unpublished doctoral dissertation]. University of South Florida.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44*(2), 162-169.

Roberts, I., Ker, K., Edwards, P., Beecher, D., Manno, D. & Sydenham, E. (2015). The knowledge system underpinning healthcare is not fit for purpose and must change. *Clinical research ed.* 350.

Rosenthal, R. (1979). The 'file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R. (1999). Meta-analytic procedures for social research. Vol. 6. Sage publications.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*(1), 59-82. https://doi.org/10.1146/annurev.psych.52.1.59

Rosnow, R. L., & Rosenthal, R. (2003).Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology. 57*(3), 221-237.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 1-9). John Wiley & Sons.

Schmidt, F. L., & Le, H. (2004). *Software for the Hunter Schmidt Meta-analysis Methods* [Unpublished Doctoral Dissertation]. University of Iowa.

Schroeder, N.L., & Kucera, A. (2021). Refutation text facilitates learning: A meta-analysis of between-subjects experiments. *Educational Psychology Review*, 1 - 31. https://doi.org/10.1007/s10648-021-09656-z

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316.

Shadish, W. R., & Haddock, C. K. (1994). Combing estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261-281). Russell Sage Foundation.

Shadish, W.R., Rindskopf, D., & Hedges, L.V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 188 - 196.

Sharpe, D.M. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17 8*, 881-901.

Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine, 98*, 1-11.

Sinatra, G.M., & Mason, L (2008). Beyond knowledge: Learner characteristics influencing conceptual change. In Vosniadou, S. (Eds.), *International handbook of research on conceptual change (*pp. 560-583). Routledge.

Sinatra, G.M., & Pintrich, P.R. (2003). *Intentional conceptual change*. Lawrence Erlbaum.

Slotta, J. D., Chi, M. T. H, & Joram, E. (1995). Assesing students' misclassification of physics concepts: An ontological basis for conceptual change. *Cognition and Instruction, 13*, 373-400.

Smith, M., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*,752-760.

Smith, E. L., Blakeslee, T. D., & Anderson, C.W. (1993). Teaching strategies associated with conceptual change. *Journal of Research in Science Teaching*, *30*(2), 111-126. https://doi.org/10.1002/tea.3660300202

Spiro, R.J., Feltovich, P.J., Coulson, R., & Anderson, D. (1988). Multiple analogies for complex concepts: antidotes for analogy-induced misconception in advanced knowledge acquisition. (Report no. 439). *ERIC*. https://eric.ed.gov/?id=ED301873

Stanley, T.D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, *8*, 581 - 591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Sterne, J.A., & Harbord, R.M. (2004). Funnel plots in meta-analysis. *The Stata Journal, 4*, 127 - 141.

Stock, W. A. (1994). Systematic coding for research synthesis. *The handbook of research synthesis, 236*, 125-138.

Strike, K. (1983). Misconceptions and conceptual change: Philosophical reflections on the research program. In H. Helm & J. Novak (Eds.), *Proceedings of the International Seminar on Misconceptions in Science and Mathematics*. Cornell University.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435-452). Russell Sage Foundation.

Swanson, E.A., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L.V. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education, 47*, 13 - 3. https://doi.org/10.1177/0022466911419516

Tabachnick, B.G. & Fidell. L.S. (2012). *Using multivariate statistics (6th ed.)*. Harper Collins College Publishers

Tebala, G. D. (2015). What is the future of biomedical research? *Medical Hypothesis, 85*, 488-490.

Teichert, M.A., Tien, L.T., Anthony, S., & Rickey, D.V. (2008). Effects of context on students' molecular-level ıdeas. *International Journal of Science Education, 30*, 1095 - 1114. https://doi.org/10.1080/09500690701355301

Thornton, A.J., & Lee, P.P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of clinical epidemiology, 53 2*, 207-16.

Tillema, H., H. (1997). Promoting conceptual change in learning to teach. *Asia-Pacific Journal of Teacher Education, 25*(1).

Turner, R.M., & Higgins, J.P. (2019). Bayesıan Meta-Analysıs. Cooper, H., Hedges, L. V., & Valentine, J. C., (Eds.).*The handbook of research synthesis and meta-analysis* (3rd ed.), 300-314. Russell Sage Foundation.

Ueno T., & Fastrich, G. M. (2016). Meta-analysis to integrate effect sizes within an article: Possible misuse and type 1 error inflation. *Journal of Experimental Psychology: General, 145*(5), 643–654.

Üstün, U. (2012). *To what extent is problem-based learning effective as compared to traditional teaching in science education? A meta-analysis study.* (Publication no. 318903) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

Üstün, U., & Eryılmaz, A. (2014). A research methodology to conduct effective research syntheses: Meta-analysis. *Education and Science, 39* (174), 1-32. https://doi.org/10.15390/EB.2014.3379

Valentine, J. C. (2019). Incorporatıng judgments about study quality into research syntheses. In h. Cooper, l. V. Hedges & Valentine, J. C. (eds.) *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of ıntervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods, 13,* 130-149.

Vevea, J. L., Zelinsky, N. A. M., & Orwin, R. G. (2019). Evaluatıng coding decisions. In Cooper, H., Hedges, l. V., & Valentine, J. C. (Eds.). *The handbook of research synthesis and meta-analysis*. Chapter 10, 173-201. Russell Sage Foundation

Vidak, A., Odžak, S., & Mešić, V. (2019). Teaching about thermal expansion: Investigating the effectiveness of a cognitive bridging approach. *Research in Science & Technological Education, 37*(3), 324-345. https://doi.org/10.1080/02635143.2018.1551200

Vosnaidou, S. (1988). *Knowledge restructuring and science instruction* [paper presentation] American Educational Research Association, New Orleans, LA.

Vosniadou, S. (1989). Analogical reasoning and knowledge acquisition: A developmental perspective. In Vosniadou, S., & Ortony, A. (Eds.). *Similarity and analogical reasoning* (pp. 413–422). Cambridge University Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change [special issue]. *Learning and Instruction, 4*, 45-69.

Vosniadou, S. (2008). *International handbook of research on conceptual change*. Routledge.

Vosniadou, S. (2019). The Development of Students' Understanding of Science. *Frontiers in Education*. https://doi.org/10.3389/feduc.2019.00032

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I., (2008). The framework theory approach to the problem of conceptual change. In Vosniadou, S. (Eds.), *International handbook of research on conceptual change* (pp. 3-34). Taylor & Francis.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Eds.), *Handbook of research on science teaching and learning*, 177-210.

Windschitl, M. A. (1995). Using computer simulations to enhance conceptual change: the roles of constructivist instruction and student epistemological beliefs.[Unpublished doctoral dissertation, Iowa State University], Iowa.

White, W.A. (1988). A meta-analysis of the effects of direct instruction in special education. *Education and Treatment of Children, 11*(4), 364-374. https://www.jstor.org/stable/42899084

Wilson, D. B. (2019). Systematic coding for research synthesis. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. 154-171.

Wise, K.C., & Okey, J.R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching, 20*, 419-435.

Yan, X., & Su, X., G. (2009). *Linear regression analysis: Theory and computing* . World Scientific Publishing

Yu, Z. (2021). A meta-analysis of effects of blended learning on performance, attitude, achievement, and engagement in different countries. *Research Square*. 1-23. https://doi.org/10.21203/rs.3.rs-536691/v1

References of Primary Studies

*Acar, B., & Tarhan, L. (2008). Effects of cooperative learning on students' understanding of metallic bonding. *Research in Science Education, 38*, 401-420. https://doi.org/10.1007/s11165-007-9054-9

*Adesope, O.O., Cavagnetto, A., Hunsu, N.J., Anguiano, C.J., & Lloyd, J. (2017). Comparative effects of computer-based concept maps, refutational texts, and expository texts on science learning. *Journal of Educational Computing Research, 55*, 46 - 69. https://doi.org/10.1177/0735633116654163

*Akbaş, Y. (2008). *Ortaöğretim 9. sınıf öğrencilerinin iklim konusundaki kavram yanılgılarının giderilmesinde kavramsal değişim yaklaşımının etkisi.*(Publication No. 235834) [Doctoral dissertation, Atatürk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Akbulut, İ., H., Çiğdem, Ş. Ç., & Salih, Ç. (2011). Effect of using different teaching methods and techniques embedded within the 5e instructional model on removing students' alternative conceptions: fluid pressure. *Social and Educational Studies. 4.* 2403-2414.

*Akgül, P. (2010). *The effect of conceptual change texts enriched with meta conceptual processes on preservice science teachers' conceptual understanding about heat and temperature*. (Publication No. 277999) [Master thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Aksu, Ş. (2010). *Ortaöğretim kimya programında mol konusundaki kavram yanılgılarının önlenmesinde aktif öğrenme yönteminin etkisi*. (Publication No. 265528) [Doctoral dissertation, Dokuz Eylül University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Alemisoğlu, Ö. K. (2014). *İlköğretim 7. sınıf öğrencilerinin karışımlar konusundaki kavram yanılgılarının belirlenmesi ve giderilmesinde kavram değişim metinlerinin etkisi.* (Publication No. 363165) [Master Thesis, 19 Mayıs University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Al Khawaldeh, S.A. (2007). Facilitating conceptual change in ninth-grade students' understanding of human circulatory system concepts. *Research in Science & Technological Education, 25,* 371 - 385.

*Al Khawaldeh, S.A. (2012). Enhancing ninth-grade students' understanding of human circulatory system concepts through conceptual change approach. *The European Journal of Social & Behavioral Sciences, 2.* 201-222.

*Al Khawaldeh, S.A. (2013). Prediction/discussion-based learning cycle versus conceptual change text: comparative effects on students' understanding of genetics. *Research in Science & Technological Education, 31,* 168-183.

*Al Khawaldeh, S.A., & Olaimat, A.M. (2010). The contribution of conceptual change texts accompanied by concept mapping to eleventh-grade students' understanding of cellular respiration concepts. *Journal of Science Education and Technology, 19,* 115-125. https://link.springer.com/article/10.1007/s10956-009-9185-z

*Allen, M., & Coole, H. (2012). Experimenter confirmation bias and the correction of science misconceptions. *Journal of Science Teacher Education, 23*, 387-405. https://doi.org/10.1007/s10972-012-9277-0

*Alparslan, C., Tekkaya, C., & Geban, Ö. (2003). Using the conceptual change instruction to improve learning. *Journal of Biological Education, 37*, 133 - 137. https://doi.org/10.1080/00219266.2003.9655868

*Amponsah, K. D. & Ochonogor, C. E. (2016). Impact of a constructivist approach to learning on high achieving students' comprehension of electrochemistry concepts. *International Conference on Education in Mathematics, Science & Technology (ICEMST),* May 19 - 22, Bodrum / Turkey.

*Anyanwu, R. (2009). *The implementation and evaluation of a constructivist intervention in secondary school science teaching in Seychelles*. [Unpublished doctoral dissertation], Unıversıty of South Afrıca.

*Arslan, H.O., Geban, Ö., & Sağlam, N. (2015). Learning cycle model to foster conceptual understanding in cell division and reproduction concepts. *Journal of Baltic Science Education*. 14(5), 670-684. https://doi.org/10.33225/jbse/15.14.670

*Asana, Y. Y. (2020). *Astronomi konularında uygulanan kavramsal değişim metinlerinin fen bilimleri öğretmen adaylarinin astronomiye yönelik başarı, tutum ve ilgi düzeylerine etkisi*. (Publication No. 637114) [Master thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Aslan, A.E., & Demircioglu, G. (2014). The effect of video-assisted conceptual change texts on 12th-grade students' alternative conceptions: the gas concept. *Procedia - Social and Behavioral Sciences, 116*, 3115-3119. https://doi.org/10.1016/j.sbspro.2014.01.718

*Atasoy, B., Akkus, H., & Kadayifci, H. (2009). The effect of a conceptual change approach on understanding of students' chemical equilibrium concepts. *Research in Science & Technological Education, 27*, 267 - 282. https://doi.org/10.1080/02635140903162587

*Atılğanlar, N. (2014). *Kavram karikatürlerinin ilköğretim yedinci sınıf öğrencilerinin basit elektrik devreleri konusundaki kavram yanılgıları üzerindeki etkisi*.(Publication No. 378513) [Master thesis, Hacettepe University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Aydın, G. (2011). *Öğrencilerin "hücre bölünmesi ve kalıtım" konularındaki kavram yanılgılarının giderilmesinde ve zihinsel modeller üzerinde yapilandirmaci yaklasimin etkisi*. (Publication No. 286527) [Doctoral dissertation, Dokuz Eylül University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Aydın, Z. (2007). *Isı ve sıcaklık konusunda rastlanan kavram yanılgıları ve bu kavram yanılgılarının gđderđlmesđnde kavram harđtalarinin kullanilmasi*.(Publication No. 200943) [Master thesis, Yüzüncü Yıl University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Ayhan, A. (2004). *Effect of conceptual change orıented ınstructıon accompanıed wıth cooperatıve group work on understandıng of acıd-base concepts*. (Publication No.153159) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Aykutlu, I. & Şen, A. İ. (2011). Using analogies in determining and overcoming high school students' misconceptions about electric current. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED) 5*(2), 221-250.

*Azizoğlu, N. (2004). *Conceptual change-oriented instruction and students' misconceptions in gases*. (Publication No. 153194) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Balcı, C. (2006). *Conceptual change text-oriented instruction to facilitate conceptual change in rate of reaction concepts*. (Publication No. 199356) [Master thesis, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Barthlow, M. J. (2011). *The effectıveness of process orıented guıded ınquıry learnıng to reduce alternate conceptıons in secondary chemıstry.* [Unpublished doctoral dissertation]. Liberty University

*Başer, M. (2006a). Fostering conceptual change by cognitive conflict based ınstruction on students' understanding of heat and temperature concepts. *Eurasia journal of mathematics, science and technology education, 2*, 96-114. https://doi.org/10.12973/ejmste/75458

*Başer, M. (2006b). Effects of conceptual change and traditional confirmatory simulations on pre-service teachers' understanding of direct current circuits. *Journal of Science Education and Technology, 15*, 367-381.

*Başer, M., & Çataloğlu, E. (2005). Effect of conceptual change orıented ınstructıon on remediation of students' mısconceptıons related to heat and temperature concepts. *Hacettepe University Journal of Education, 29, 43-52.*

*Başer, M., & Geban, Ö. (2007). Effect of instruction based on conceptual change activities on students' understanding of static electricity concepts. *Research in Science & Technological Education, 25*, 243 - 267.

*Başer, M., & Geban, Ö. (2007). Effectiveness of conceptual change instruction on understanding of heat and temperature concepts. *Research in Science & Technological Education, 25*, 115 - 133.

*Bawaneh, A. K. A., Zain, A. N., & Saleh, S. (2010). Radical conceptual change through teaching method based on constructivism theory for eighth-grade Jordanian students. *Uluslararası Sosyal Araştırmalar Dergisi 3*(14). 131-147.

*Bayar, D. (2009). *Kavramsal değişim yaklaşımının ilköğretim 8.sınıf öğrencilerinin fotosentez ve bitkilerde solunum konusunu anlamalarına etkisi.* (Publication No. 255470) [Master thesis, Sakarya University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Belge Can, H., & Boz, Y. (2016). Structuring cooperative learning for motivation and conceptual change in the concepts of mixtures. *International Journal of Science and Mathematics Education, 14*, 635-657. https://doi.org/10.1007/s10763-014-9602-5

*Berber, N. C. & Sarı, M. (2009). Kavramsal değişim metinlerinin iş, güç, enerji konusunu anlamaya etkisi. *Selçuk Üniversitesi Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 27*, 159 -172.

*Bilgin, I., & Geban, Ö. (2001). Benzeşim (analoji) yöntemi kullanarak lise 2. sınıf öğrencilerinin kimyasal denge konusundaki kavram yanılgılarının giderilmesi. *Hacettepe University Journal of Education, 20*(20).

*Bilgin, I., & Geban, Ö. (2006). The effect of cooperative learning approach based on conceptual change condition on students' understanding of chemical equilibrium concepts. *Journal of Science Education and Technology, 15*, 31-46.

*Bozkoyun, Y. (2004). *Facılıtatıng conceptual change ın learnıng rate of reactıon concept.* (Publication No. 153458) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Broughton, S.H., Sinatra, G.M., & Reynolds, R.E. (2010). The nature of the refutation text effect: an ınvestigation of attention allocation. *The Journal of Educational Research, 103*, 407 - 423. https://doi.org/10.1080/00220670903383101

*Budiman, Z.B., Halim, L., Mohd Meerah, S., & Osman, K. (2014). The effects of cognitive conflict management on cognitive development and science achievement. *International Journal of Science and Mathematics Education, 12*, 1169-1195. https://link.springer.com/article/10.1007/s10763-013-9460-6

*Can, B.T., Yaşadi, G., Sönmezer, D., & Kesercioğlu, T.İ. (2006). Fen öğretiminde kavram haritaları ve senaryolar kavram yanılgılarını giderebilir mi? *Hacettepe University Journal of Education, 31*, 133-146.

*Canpolat, N., Pinarbaşi, T., Bayrakçeken, S., & Geban, O. (2006). The conceptual change approach to teaching chemical equilibrium. *Research in Science & Technological Education, 24*, 217 - 235. https://doi.org/10.1080/02635140600811619

*Carlsen, D.D. (1989). *Overcoming student preconceptions about simple series circuits: promoting conceptual change with text manipulations and a microcomputer simulation*. [Unpublished doctoral dissertation]. Iowa State University. Semantic scholar. https://doi.org/10.31274/RTD-180813-11283

*Ceylan, E. (2004). *Effect of instruction using conceptual change strategies on student conceptions of chemical reactions and energy*. (Publication No. 153141) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Ceylan, E., & Geban, Ö. (2009). Facilitating conceptual change in understanding state of matter and solubility concepts by using 5E learning cycle model. *Hacettepe University Journal of Education, 36*, 41-50.

*Chang, Y. & Baruffaldi, J. P. (1999). The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks, *International Journal of Science Education*, 21:4, 373-388, https/doi.org/ 10.1080/095006999290606

*Charles, E. S. A. (2003). *An ontological approach to conceptual change: the role that complex systems thinking may play in providing the explanatory framework needed for studying contemporary sciences*. [Unpublished doctoral dissertation]. Concordia University.

*Chen, C., & She, H. (2012). The impact of recurrent online synchronous scientific argumentation on students' argumentation and conceptual change. *J. Educ. Technol. Soc., 15*, 197-210.

*Chen, Y., Pan, P., Sung, Y., & Chang, K. (2013). Correcting misconceptions on electronics: effects of a simulation-based learning environment backed by a conceptual change model. *J. Educ. Technol. Soc., 16*, 212-227.

*Chiu, M., & Lin, J. (2005). Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. *Journal of Research in Science Teaching, 42*, 429-464. https://doi.org/10.1002/tea.20062

*Clement, J. (1993). Using bridging analogies and anchoring intentions to deal with student preconceptions in physics. *Journal of research in science teaching, 30*(10), 1241–1257. https://doi.org/10.1002/tea.3660301007

*Coetzee, A., & Imenda, S.N. (2012). Effects of outcomes-based education and traditional lecture approach in overcoming alternative conceptions in physics. *African Journal of Research in Mathematics, Science and Technology Education, 16*, 145 - 157. https://doi.org/10.1080/10288457.2012.10740736

*Çakir, Ö.S., Geban, Ö., & Yürük, N. (2002). Effectiveness of conceptual change text-oriented instruction on students' understanding of cellular respiration concepts. *Biochemistry and Molecular Biology Education, 30*(4), 239-243. https://doi.org/10.1002/bmb.2002.494030040095

*Çakmak, G. (2016). *Bilgisayar destekli kavramsal değişim metinlerinin fen bilimleri dersinde öğrencilerin başarılarına ve tutumlarına etkisi*. (Publication No. 452038) [Doctoral dissertation, Fırat University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Çalık, M., Kolomuç, A., & Karagölge, Z. (2010). The effect of conceptual change pedagogy on students’ conceptions of rate of reaction. *Journal of Science Education and Technology, 19*, 422-433.

*Çaycı, B. (2007). Kavram değiştirme metinlerinin kavram öğrenimi üzerindeki etkisinin incelenmesi. *Gazi Eğitim Fakültesi Dergisi, 27*(1), 87-102.

*Çaycı, B. (2018). The impacts of conceptual change text-based concept teaching on various variables. *Universal Journal of Educational Research*. *Universal Journal of Educational Research 6*(11), 2543-2551.

*Çelebi, Ö. (2004). *Effect of conceptual change-oriented instruction on removing misconceptions about phase changes* (Publication No. 153142) [Master thesis, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Çelik, G. (2013). *Sınıf öğretmenliği öğrencilerinin gazlar konusundaki kavram yanılgılarına tahmin-gözlem-açıklama tekniğinin etkisi. (Publication No.345319)* [Master thesis, Bülent Ecevit University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Çelikten, O., Ipekcioglu, S., Ertepınar, H., & Geban, Ö. (2012). The effect of the conceptual change-oriented instruction through cooperative learning on 4th grade students' understanding of earth and sky concepts. *Science education international, 23*, 84-96.

*Çetin, G. *(2003). The effect of conceptual change ınstructıon on understandıng of ecology concepts* (Publication No.143632*)* [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Çetin, G., Ertepınar, H., & Geban, Ö. (2015). Effects of conceptual change text-based instruction on ecology, attitudes toward biology and environment. *Educational Research Review, 10*, 259-273.

*Çetin, P. S. (2009). *Effects of conceptual change orıented ınstructıon on understandıng of gasses concepts* (Publication No. 255254) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Çetin, P.S., Kaya, E., & Geban, Ö. (2009). Facilitating conceptual change in gasses concepts. *Journal of Science Education and Technology, 18*, 130-137.

*Çetingül, I. P. (2006). *Effect of conceptual change texts accompanıed wıth analogıes on promotıng conceptual change ın acıd and base concepts* (Publication No.181013) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Çetingül, I. P., & Geban, Ö. (2005). Understanding of acıd-base concept by using conceptual change approach. *H. U. Journal of Education*. 29: 69-74.

*Çetingül, I. P., & Geban, Ö. (2011). Using conceptual change texts with analogies for misconceptions in acids and bases. *Hacettepe Universitesi Egitim Fakultesi Dergisi-hacettepe University Journal of Education, 41*, 112-123.

*Çıbık, A.S. (2011). *Elektrik akımı konusunda yanlış kavramlar ve bunların giderilmesinde analojilerle desteklenmiş proje tabanlı öğrenme yönteminin* *etkisi* (Publication No. 298512) [Doctoral dissertation, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Çıbık, A.S., Diken, E.H., & Darçın, E.S. (2008). The effect of group works and demonstrative experiments based on conceptual change approach: photosynthesis and respiration. *Asia-Pacific Forum on Science Learning and Teaching, 9*(2), 1-22.

*Çinici, A., Sözbilir, M. & Demir, Y. (2011). Effect of cooperative and individual learning activities on students' understanding of diffusion and osmosis. *Eurasian Journal of Educational Research, 43*, 19-36.

*Çobanoglu, E. O.& Bektas, H. (2012). *Kavramsal değişim metinlerinin ilköğretim 6. Sınıf öğrencilerinin dolaşım sistemi konusundaki kavram yanılgılarının giderilmesine etkisi* [paper presantaton]. X. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Niğde, 0, 6.

*Çoruhlu, T. Ş. & Çepni, S (2015). Kavramsal değişim pedagojileri ile zenginleştirilmiş 5E modelinin öğrenci kavramsal değişimi üzerine etkisinin değerlendirilmesi: "Kuyruklu Yıldız", "Yıldız Kayması" ve "Meteor" örneği. *Journal of Educational Sciences 41*, 139-155.

*Damlı, V. (2011). *Kavramsal değişim yaklaşımına dayalı web tabanlı etkileşimli öğretimin üniversite öğrencileirnin ısı ve sıcaklık konusundaki kavram yanılgılarını gidermeye etkisi* (Publication No. 279795) [Master thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Demirci, M. P. & SArıkaya, m. (2004). Sınıf öğretmeni adaylarının ısı ve sıcaklık konusundaki kavram yanılgıları ve yanılgıların giderilmesinde yapısalcı kuramın etkisi. *XIII. Ulusal Eğitim Bilimleri Kurultayı, 6-9 Temmuz 2004 İnönü Üniversitesi, Eğitim Fakültesi, Malatya.*

*Demircioglu, G. (2009). Comparison of the effects of conceptual change texts implemented after and before ınstruction on secondary school students' understanding of acid-base concepts. *Asia-Pacific Forum on Science Learning and Teaching, 10*(2), 1-29.

*Demircioglu, G., Ayas, A., & Demircioglu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice, 6*, 36-51.

*Demircioğlu G., Aydın M. A., & Demircioğlu, H. (2013). Kavramsal değişim metninin ve üç boyutlu modelin 7. sınıf öğrencilerinin atomun yapısını anlamalarına etkisi. *Bayburt Üniversitesi Eğitim Fakültesi, 7(*2), 70-96.

*Demirel, M., & Anıl, Ö. (2018). Kavramsal değişim yaklaşımına yönelik çalışma: gazlar konusu. *Science Journal of Turkish Military Academy, 27*(2), 93-118.

*Demirer, G. M. (2015). *Kavram yanılgılarının giderilmesinde simülasyonların etkisinin incelenmesi: Işık ve ses ünitesi örneği* (Publication No. 418437) [Master thesis, Kırıkkale University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Demirezen, S., & Yağbasan, R. (2013). 7E modelinin basit elektrik devreleri konusundaki kavram yanılgıları üzerine etkisi. *Hacettepe Universitesi Egitim Fakultesi Dergisi-hacettepe University Journal of Education, 28*, 132-151.

*Diakidoy, I., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: a comparison of the effectiveness of two instructional approaches. *Learning and Instruction, 11*, 1-20.

*Diakidoy, I., Kendeou, P., & Ioannides, C. (2003). Reading about energy: the effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*, 335-356.

*Diakidoy, I., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction, 41*, 60-69.

*Dilber, R. (2006). *Fizik öğretiminde analoji kullanımının ve kavramsal değişim metinlerinin kavram yanılgılarının giderilmesine ve öğrenci başarısına etkisinin araştırılması* (Publication No. 181511) [Doctoral dissertation, Atatürk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Dilber, R. (2008). Teaching of the water waves: effectiveness of using computer simulations on student success and elimination of misconceptions. *Bulgarian Journal of Chemistry Education, 17*(1), 33-45.

*Dilber, R. (2010). Effect of conceptual change instruction on students' understanding of electricity concepts. *International Journal of Innovation and Learning, 7(4)*, 478-496.

*Dilber, R., & Düzgün, B. (2008). Effectiveness of analogy on students' success and elimination of misconceptions. *Latin-American Journal of Physics Education, 2*, 3.

*Dilber, R., Karaman, I., & Duzgun, B. (2009). High school students' understanding of projectile motion concepts. *Educational Research and Evaluation, 15(3),* 203 - 222.

*Duman, M. Ş. (2015). *8.sınıf öğrencilerinin "maddenin halleri ve ısı" ünitesinde karşılaşılan kavram yanılgılarının belirlenmesi ve giderilmesine, başarı düzeylerine ve öğrenilenlerin kalıcılığına sanal laboratuvar uygulamalarının etkisi* (Publication No. 395311) [Master thesis, Mersin University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Durmuş, J. (2009). *İlköğretim fen bilgisi dersinde kavramsal değişim metinlerinin ve deney yönteminin akademik başarıya ve kavram yanılgılarını gidermeye etkisi* (Publication No. 234852) [Master thesis, Selçuk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Erdemir, A (2006). *Effect of cooperative learning based on conceptual change conditions on seventh-grade students' understanding of classification of matter and physical and chemical changes* (Publication No. 172321) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Eryılmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching, 39,* 1001-1015.

*Eymur, G. (2014). *The collaboration of cooperative learning and conceptual change: enhancing the students' understanding of chemical bonding concepts* (Publication No. 415520) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Fan, X., Geelan, D.R., & Gillies, R.M. (2018). Evaluating a novel ınstructional sequence for conceptual change in physics using ınteractive simulations. *Education Sciences, 8,* 29-48.

*Feyzioglu Y. E., Ergin, Ö., & Kocakülah, M.S. (2012). *The effect of 5e learning model ınstruction on seventh-grade students' conceptual understanding of force and motion. International Online Journal of Educational Sciences,* 4 (3), 691-705.

*Gayeta, N. E. & Caballes, D. G. (2017). Measuring conceptual change on stoichiometry using mental models and ıllstructured problems ın a flipped classroom environment. *Asia Pacific Journal of Multidisciplinary Research, 5*(2), 104-113.

*Gedik, E., Ertepınar, H., & Geban, Ö. (2002). *Lise öğrencilerinin elektrokimya konusundaki kavramları anlamalarında kavramsal değişim yaklaşımına dayalı gösteri yönteminin etkisi.* [paper presantation]. 5. Ulusal Fen Bilimleri ve Matematik Eğitim Kongresi, Ankara.

*Gokhale, A.A. (1996). Using Analogies to Overcome Misconceptions: A Technology Course Example. *The Journal of Technology Studies, 22*, 10-14.

*Gülçiçek, N. Y. (2004). *Kavramsal değişim metinlerinin öğrencilerin manyetizma konusunu anlamalarına ve fizik tutumlarına etkisi* (Publication No. 146344) [Doctoral dissertation, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Günay, B. (2005). *Conceptual change text orıented ınstructıon to facılıtate conceptual change ın atoms and molecules* (Publication No. 167126) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Gürbüz, F. (2008). *İlköğretim 6. sınıf öğrencilerinin ısı ve sıcaklık konusundaki kavram yanılgılarının düzeltilmesinde kavramsal değişim metinlerinin etkisinin araştırılması* (Publication No. 232824) [Master, Atatürk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Gürkan, Ş. (2021). *Çürütme metni destekli hikâyeler ve dijital hikâyeler içeren kavramsal değişim yaklaşımlarının kavram yanılgılarının giderilmesine etkisi: yer kabuğu ve dünyamızın hareketleri ünitesi* (Publication No. 690741) [Master Thesis, Mehmet akif Ersoy University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Gürses, A., Doğar, Ç., Yalçın' M. & Canpolat, N. (2002). Kavramsal değişim yaklaşımının öğrencilerin gazlar konusunu anlamalarına etkisi. Academia paper. https://www.academia.edu/775426

*Hacımustafaoğlu, M. (2015). *Ortaokul 8. sınıf öğrencilerinin "Maddenin Halleri ve Isı" ünitesinde kavramsal değişim sağlamalarında farklı kavramsal değişim yöntem ve tekniklerle zenginleştirilmiş rehber materyallerin etkisi* (Publication No. 415482) [Master thesis, Hacettepe University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Hanson, R. & Seheri, N. (2019). Assessing conceptual change instruction accompanied with Concept Maps and Analogies: A case of acid-base strengths. *Journal of Turkish Science Education, 15*. 55-64.

*Harman, G. (2016). *5. sınıf "yaşamımızın vazgeçilmezi: elektrik" ünitesinde kullanılan analojinin öğrenci başarısı, tutum, zihinsel modelleme ve kavram yanılgıları üzerine etkisi 5. sınıf "yaşamımızın vazgeçilmezi: elektrik" ünitesinde kullanılan analojinin öğrenci başarısı, tutum, zihinsel modelleme ve kavram yanılgıları üzerine etkisi* (Publication No. 442976) [Doctoral dissertation, 19 Mayıs University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi

*Hırça, N., Çalık, M., Seven, S., & Fen, T. (2011). Effects of guide materials based on 5e model on students' conceptual change and their attitudes towards physics: a case for 'work, power and energy' unit. *Journal of Turkish Science Education, 8*(1), 139-152. https://www.tused.org/index.php/tused/article/view/550

*İnal, A. (2003). *Lise 1. sınıftaki öğrencilerin ısı ve sıcaklık konusundaki yanlış kavramlarının belirlenmesi* (Publication No. 133962) [Master Thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*İpek, İ. (2007). *Implementation of conceptual change-oriented instruction using hands-on activities on tenth-grade students' understanding of gasses concepts* (Publication No. 199439) [Master, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*İşcan, Y. V. (2020). *Kavramsal değişimi sağlayan materyallerle destekli 5e öğrenme modelinin fen akademik başarısına etkisi: isi ve sicaklik konusu örneği* (Publication No. 646851) [Master thesis, Cumhuriyet University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Jensen, M.S., Wilcox, K.J., Hatch, J.T., & Somdahl, C. (1996). A computer-assisted ınstruction unit on diffusion and osmosis with a conceptual change design. *The Journal of Computers in Mathematics and Science Teaching, 15*, 49-64.

*Johnson, M., & Sinatra, G.M. (2013). Use of task-value ınstructional ınductions for facilitating engagement and conceptual change. *Contemporary Educational Psychology, 38*, 51-63.

*Karakethudaoglu, N. A. (2010). *Kavramsal değişim yaklaşımına dayalı öğretimin kimyasal denge kavramlarini anlamaya ve tutuma etkisi* (Publication No. 265019) [Master thesis, Karaelmas University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Karakuyu, Y., & Tüysüz, C. (2011). Misconceptions in electricity and conceptual change strategy. *Gaziantep University Journal of Social Sciences, 10*, 867-890.

*Karamustafaoğlu, S., & Mamlok-Naaman, R. (2015). Understanding Electrochemistry Concepts Using the Predict-Observe-Explain Strategy. *Eurasia journal of mathematics, science and technology education, 11*, 923-936.

*Karamustafaoğlu, S, Ayaş, A,. & Cotu, B. (2002, September). *Sınıf öğretmeni adaylarının çözeltiler konusundaki kavram yanılgıları ve bu yanılgılarının kavram haritasi tekniği ile giderilmesi* [paper presantation]. V. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, 1, 664-671

*Karslı, F. & Ayas, A. (2013). Farklı kavramsal değişim yöntemleri ile alternatif kavramları gidermek ve bilimsel süreç becerilerini geliştirmek mümkün müdür? elektrokimyasal piller örneği. *Journal of computer and educational research. 1*(1), 1-26.

*Kasap, G., & Ültay, N. (2014). To determine the effect of the activities based on conceptual change approach on students 'conceptual understanding of floating-sinking objects. *Kastamonu Eğitim Dergisi, 22*(2), 455-472. https://www.academia.edu/6098968

*Kaya, E. (2011). *The effect of conceptual change based instruction on students' understanding of rate of reaction concepts* (Publication No. 266184) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi..

*Keleş, P. U. (2009). *Kavramsal değişim metinleri, oyun ve drama ile zenginleştirilmiş 5e modelinin etkililiğinin belirlenmesi: "canlıları sınıflandıralım" örneği* (Publication No. 244499) [Doctoral dissertation, Karadeniz Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Kılıç, D. (2007). *Analojilerle öğretim modelinin 9. sınıf öğrencilerinin kimyasal bağlar konusundaki yanlış kavramlarının giderilmesi üzerine etkisi* (Publication No.190954) [Master thesis, Gazil University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi..

*Kılıç, Y. (2016). *İşbirlikli öğrenme yönteminin 5. sınıf öğrencilerinin fen bilimleri dersi vücudumuzun bilmecesini* (Publication No. 450246) [Master thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi..

*Kıngır, S., Geban, O., & Gunel, M. (2013). Using the science writing heuristic approach to enhance student understanding in chemical change and mixture. *Research in Science Education, 43*, 1645-1663.

*Kırık, Ö.T., & Boz, Y. (2012). Cooperative learning instruction for conceptual change in the concepts of chemical kinetics. *Chemistry Education Research and Practice, 13*, 221-236.

*Kör, S. A. (2006). *İlköğretim 5. sınıf öğrencilerinde "yaşamımızdaki elektrik" ünitesinde görülen kavram yanılgılarının giderilmesinde bütünleştirici öğrenme kuramına dayalı geliştirilen materyallerin etkisi* (Publication No.182943) [Master thesis, Karadeniz Techical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Köse, S. (2004). *Fen bilgisi öğretmen adaylarında fotosentez ve bitkilerde solunum konularında görülen kavram yanılgılarının giderilmesinde kavram haritalarıyla verilen kavramdeğişim metinlerinin etkisi* (Publication No. 156154) [Doctoral dissertation, Karadeniz Techical University]. Council of Higher Education Thesis Center.https://tez.yok.gov.tr/UlusalTezMerkezi.

*Köseoğlu, F. & Bayır, E. (2012). Sorgulayici-arastirmaya dayali analitik kimya laboratuvarlarında kimya öğretmen adaylarinin kavramsal değişimlerine, bilimi ve bilim öğrenme yollarini algilamalarina etkileri . *Türk Eğitim Bilimleri Dergisi , 10* (3) , 604-626.

*Küçük, Z., & Çalık, M. (2015). Effect of enriched 5E model on grade-7 students' conceptual change levels: a case of 'electric current' subject. *Adıyaman University Journal of Educational Sciences, 5*(1), 1-28. http://dx.doi.org/10.17984/adyuebd.80603

*Launey, B.L. (1995). *Using Drawing Tasks to Communicate Ideas About Photosynthesis: A Conceptual Change Strategy for Use in the Elementary School Classroom*. [Unpublished doctoral dissertation], Louisiana State University.

*Lee, C., & She, H. (2010). Facilitating students' conceptual change and scientific reasoning involving the unit of combustion. *Research in Science Education, 40*, 479-504.

*Li, C.S. (2008). Examining the use of dynamic modeling environment to support learning and teaching of science: a quantitative analysis. *International journal of instructional media, 35*(4), 443+. http://www.ijim.info/

*Liao, Y., & She, H. (2009). Enhancing eight-grade students' scientific conceptual change and scientific reasoning through a web-based learning program. *J. Educ. Technol. Soc., 12*, 228-240.

*Lin, Y., Liu, T., & Chu, C. (2011). Implementing clickers to assist learning in science lectures: the clicker-assisted conceptual change model. *Australasian Journal of Educational Technology, 27,* 979-996.

*Liu, Q. (2008). *Conceptual change with refutational maps* [Unpublished master thesis], Qingdao University.

*Loon, M.H., Dunlosky, J., Gog, T.V., Merriënboer, J.V., & Bruin, A.D. (2015). Refutations in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology, 42*, 39-48.

*Loyens, S.M., Jones, S.H., Mikkers, J., & Gog, T.V. (2015). Problem-based learning as a facilitator of conceptual change. *Learning and Instruction, 38*, 34-42.

*Mason, L., Baldi, R., Ronco, S. , Scrimin, S., Danielson, R. W.& Sinatra, G. M. (2017). Textual and graphical refutations: effects on conceptual change learning. *Contemporary Educational Psychology,* 49, 275–288.

*Mason, L., Zoccoletti, S., Carretti, B., Scrimin, S., & Diakidoy, I. (2019). The role of ınhibition in conceptual learning from refutation and standard expository texts. *International Journal of Science and Mathematics Education, 17*, 483-501.

*Mikkilä-Erdmann, M. (2001). Improving conceptual change concerning photosynthesis through text design. *Learning and Instruction, 11*, 241-257.

*Muis, K.R., Sinatra, G.M., Pekrun, R., Winne, P.H., Trevors, G.J., Losenno, K.M., & Munzar, B. (2018). Main and moderator effects of refutation on task value, epistemic emotions, and learning strategies during conceptual change. *Contemporary Educational Psychology*. 55, 155–165.

*Niaz, M., & Chacón, E. (2003). A conceptual change teaching strategy to facilitate high school students' understanding of electrochemistry. *Journal of Science Education and Technology, 12*, 129-134.

*Nwankwo, M.C., & Madu, B.C. (2014). Effect of analogy teaching approach on students' conceptual change in physics. *Greener Journal of Educational Research, 4*, 119-125.

*Okur, M.(2009). *Kavramsal değişimi sağlayan farklı metotların karşılaştırılması: sesin yayılması konusu örneği (*Publication No. 233689) [Master thesis, Karadeniz Techical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Önder, İ. (2006). *The effect of conceptual change approach on students' understanding of solubility equilibrium concept* (Publication No. 180962) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Önder, İ. (2017). The effect of conceptual change texts supplemented ınstruction on students' achievement in electrochemistry. *International Online Journal of Educational Sciences*.

*Özkan, Ö., Tekkaya, C., & Geban, Ö. (2004). Facilitating conceptual change in students' understanding of ecological concepts. *Journal of Science Education and Technology, 13*, 95-105.

*Özkan, G. (2013). *The effects of conceptual change texts and real-life context-based learning on students' approach to learning physics and their conceptual understandings* (Publication No. 342308) [Master thesis, Dokuz Eylül University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Özmen, H. (2007). The effectiveness of conceptual change texts in remediating high school students' alternative conceptions concerning chemical equilibrium. *Asia Pacific Education Review, 8*, 413-425.

*Özmen, H. (2011). Effect of animation enhanced conceptual change texts on 6th-grade students' understanding of the particulate nature of matter and transformation during phase changes. *Comput. Educ. 57*, 1114-1126.

*Özmen, H. & Demircioğlu, G. (2003). Asitler ve bazlar konusundaki öğrenci yanlış anlamalarının giderilmesinde kavram değişim metinlerinin etkisi. *Milli Eğitim Dergisi, 159*, 111–119.

*Özmen, H., Demircioglu, H., & Demircioglu, G. (2009). The effects of conceptual change texts accompanied with animations on overcoming 11th-grade students' alternative conceptions of chemical bonding. *Comput. Educ., 52*, 681-695.

*Özmen, H., & Naseriazar, A. (2017). Effect of simulations enhanced with conceptual change texts on university students' understanding of chemical equilibrium. *Journal of The Serbian Chemical Society, 83*, 121-137.

*Pabuççu, A. (2004). *Effect of conceptual change texts accompanied with analogies on understanding of chemical bonding concepts* (Publication No. 153160) [Master thesis, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Pabuçcu, A., & Geban, Ö. (2015). 5E öğrenme döngüsüne göre düzenlenmiş uygulamaların asit-baz konusundaki kavram yanılgılarına etkisi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 15*, 191-206.

*Pekel, Y., & Hasenekoğlu, P. (2015). Dynamising conceptual change approach to teach some genetics concepts/genetik kavramlarının öğretiminde kavramsal değişim yaklaşımının etkinliğinin artırılması. *e-International Journal of Educational Research, 6,* 51-68.

*Pekmez, E. (2010). Using analogies to prevent misconceptions about chemical equilibrium. *Asia-Pacific Forum on Science Learning and Teaching* 11(2), 1-35.

*Perdana, G.P., Suma, K., & Pujani, N.M. (2018). The effect of conceptual change text structure on concept understanding and misconception reduction of dynamic electricity. SHS Web of Conferences 42. https://doi.org/10.1051/shsconf/20184200075

*Pınarbaşı, T., Canpolat, N., Bayrakçeken, S., & Geban, Ö. (2006). An ınvestigation of effectiveness of conceptual change text-oriented ınstruction on students' understanding of solution concepts. *Research in Science Education, 36*, 313-335.

*Polat, D. (2007). *Kuvvet ve hareket konusu ile ilgili öğrencilerin kavram yanılgılarının tespiti ve kavram karmaşası yöntemiyle düzeltilmesi* (Publication No.207156) [Master thesis, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Sahyar, & Nst, F.H. (2017). The Effect of Scientific Inquiry Learning Model Based on Conceptual Change on Physics Cognitive Competence and Science Process Skill (SPS) of Students at Senior High School. *Journal of Education and Practice, 8,* 120-126.

*Saigo, B. W. (1999). *A study to compare tradıtıonal and constructıvısm-based ınstructıon of a hıgh school biology unıt on bıosystematıcs.* [Unpublished Doctoral Dissertation]. The University of Iowa.

*Sanger, M.J., & Greenbowe, T.J. (2000). Addressing student misconceptions concerning electron flow in aqueous solutions with instruction including computer animations and conceptual change strategies. *International Journal* *of Science Education, 22*, 521 - 537. https://doi.org/10.1080/095006900289769

*Sarı Ay, Ö. (2011). *The effect of using conceptual change texts and students' opinions in the misconceptions identified removal in the unit of states of matter and heat? in the science and technology course of primary 8th class* (Publication No. 308425) [Master thesis, Hacettepe University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Savinainen, A., Scott, P., & Viiri, J. (2005). Using a bridging representation and social ınteractions to foster conceptual change: designing and evaluating an ınstructional sequence for newton's third law. *Science Education, 89*, 175-195.

*Sevim, S. (2007). *Preparation and application of conceptual change texts on solution and chemical bonding concepts* (Publication No. 212054) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkez

*Seyedmonir, M. (2000). *The development and validation of science learning ınventory (slı): a conceptual change framework* (Publication No. 3039331) [Doctoral dissertation, Wes Virginia University]. ProQuest Dissertation. https://www.proquest.com/docview

*She, H. C. & Lee, C. Q (2008). SCCR digital learning system for scientific conceptual change and scientific reasoning. *Computers & Education* 51, 724–742.

*Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24*(2), 261–289.

*Sota, M. (2012). *The effect of contrasting analogies on understanding of and reasoning about natural selection* (Publication No. 3519372) [Doctoral dissertation, Florida University]. ProQuest Dissertation. https://www.proquest.com/docview

*Södervik, I., Mikkilä-Erdmann, M., & Vilppu, H. (2014). Promoting the understanding of photosynthesis among elementary school student teachers through text design. *Journal of Science Teacher Education, 25*, 581-600.

*Södervik, I., Virtanen, V., & Mikkilä-Erdmann, M. (2015). Challenges in understanding photosynthesis in a university introductory biosciences class. *International Journal of Science and Mathematics Education, 13*, 733-750.

*Stavy, R.(1991). Using analogy to overcome misconceptions about conservation of matter. *J. Res. Sci. Teach. 28,* 305–313. https://doi.org/10.1002/tea.3660280404

*Sungur, S., Tekkaya, C., & Geban, Ö. (2001). The contribution of conceptual change texts accompanied by concept mapping to students' understanding of the human circulatory system. *School Science and Mathematics, 101,* 91-101.

*Şeker, A. (2012). *Conceptual change-oriented instruction and students' misconceptions in chemical bonding concepts* (Publication No. 305019) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Şendur, G., & Toprak, M. (2013). The role of conceptual change texts to improve students' understanding of alkenes. *Chemistry Education Research and Practice, 14,* 431-449.

*Şendur, G., Toprak, M., & Pekmez, E.Ş. (2008). Buharlaşma ve kaynama konularındaki kavram yanılgılarının önlenmesinde analoji yönteminin etkisi. *Ege Eğitim Dergisi, 9*(2), 37-58.

*Taşdelen, U. (2011). *The effects of computer-based interactive conceptual change texts on 11th-grade students' understanding of electrochemistry concepts and attitude toward chemistry* (Publication No. 285723) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Taştan, I., Dikmenli, M., & Çardak, O. (2008). Effectiveness of the conceptual change texts accompanied by concept maps about students' understanding of the molecules carrying genetic ınformation. *Asia-Pacific Forum on Science Learning and Teaching, 9*(1), Article 11.

*Taştan, Ö., Yalçınkaya, E., & Boz, Y. (2008). Effectiveness of conceptual change text-oriented ınstruction on students' understanding of energy in chemical reactions. *Journal of Science Education and Technology, 17,* 444-453.

*Tekkaya, C. (2003). Remediating high school students' misconceptions concerning diffusion and osmosis through concept mapping and conceptual change text. *Research in Science & Technological Education, 21,* 16 - 5.

*Tekkaya, Ö. Ö. C.,& Geban Ö. (2004). Facilitating conceptual change in students' understanding of ecological concepts. *Journal of Research Education and Technology, 13*(1), 91-101.

*Tezcan, H., & Salmaz, Ç. (2005). Atomun yapısının kavratılmasında ve yanlış kavramaların giderilmesinde bütünleştirici ve geleneksel öğretim yöntemlerinin etkileri. *Gazi Üniversitesi Eğitim Fakültesi Dergisi,* 25(1), 41 – 54.

*Tlala, B.M. (2014). Investigating grade 10 learners' achievements in photosynthesis using conceptual change models. *Journal of Baltic Science Education*. 13(2), 155-164.

*Tokatlı, F. R. (2010). *Kavramsal değişim yaklaşımı işbirlikli öğrenme ve bilgisayar destekli öğretimin öğrencilerin fen başarısına etkisi* (Publication No. 274809) [Master thesis, Sakarya University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Tokur, F., Duruk, Ü., & Akgün, A. (2014). Investigation of the effect of POE activities on remedying preservice science teachers' misconceptions in the context of growing and developing in flowery plants unit. *Route Educational & Social Science Journal. 1*(1), 1-13.

*Toros, S. (2007). *Sosyal bilgiler öğretiminde bilgisayar destekli öğretimin kavram yanılgılarını giderme üzerine etkisi* (Publication No.407642) [Master thesis, Cumhuriyet University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Trevors, G. (2011). Learner, text, and context factors on conceptual change in biology (Publication No. MR84312) [Master thesis, McGill University]. ProQuest Dissertation. https://www.proquest.com

*Tsai, C. (2003). Using a conflict map as an instructional tool to change student alternative conceptions in simple series electric circuits. *International Journal of Science Education, 25*, 307 - 327.

*Turgut, Ü., & Gürbüz, F. (2011). Effects of teaching with 5E model on students' behaviors and their conceptual changes about the subject of heat and temperature. *International Online Journal of Educational Sciences,* 3(2), 679-706

*Udogu, M., & Njelita, C.B. (2010). Effect of constructivist-based ınstructional model on students' conceptual change and retention on some difficult concepts in chemistry. *African Research Review, 4*(2*),* 219-229. https://doi.org/ 10.4314/afrrev.v4i2.58305

*Uyanık, G. & Dindar, H. (2016). The effect of the conceptual change texts on removing misconceptions in primary 4th-grade science course. *GEFAD / GUJGEF 36*(2), 349-374

*Uzun, B. (2010). F*en ve teknoloji öğretiminde kavramsal değişim stratejilerine dayalı olarak maddenin yapısı ve özellikleri konusunun öğretimi.(Publication No.265524)* [Doctoral dissertation, Dokuz Eylül University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Uzuntiryaki, E. (2003). *Effectiveness of constructivist approach on students' understanding of chemical bonding concepts* (Publication No. 143656) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Uzuntiryaki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. *Instructional Science, 33,* 311-339. https://link.springer.com

*Üce, M. (2009). Teaching the mole concept using a conceptual change method at college level. *Education 3-13, 129,* 683-691.

*Ünlü, Y. (2012). *Conceptual change texts oriented instruction in teaching solution concept* (Publication No. 368782) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Vatansever, O. (2006). *Effectiveness of conceptual change instruction on overcoming students' misconceptions of electric field, electric potential, and electric potential energy at tenth-grade level (*Publication No. 199350) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Windschitl, M., & Andre, T. (1996). Using computer simulations to enhance conceptual change: the roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching, 35*, 145-160. https://doi.org/10.1002/(SICI)1098-2736

*Woloshyn, V. E., Paivio, A., & Pressley, M. (1994). Use of elaborative interrogation to help students acquire information consistent with prior knowledge and information inconsistent with prior knowledge. *Journal of Educational Psychology, 86*(1), 79–89. https://doi.org/10.1037/0022-0663.86.1.79

*Wood, L., Ebenezer, J.V., & Boone, R.T. (2013). Effects of an intellectually caring model on urban African American alternative high school students' conceptual change and achievement in chemistry. *Chemistry Education Research and Practice, 14*, 390-407.

*Wozniak, B. M. (2012). *Effect of Process-Oriented Guided-Inquiry Learning on Non-majors Biology Students' Understanding of Biological Classification*. [Unpublished doctoral dissertation, Minnesota State University]. https://ui.adsabs.harvard.edu

*Yalçınkaya, E., & Boz, Y. (2015). The effect of case-based instruction on 10th-grade students' understanding of gas concepts. *Chemistry Education Research and Practice, 16*, 104-120.

*Yaman, İ. (2013). *Effects of instructions based on cognitive brıdgıng and cognıtıve conflıct on 9th-grade students' understandıng of force and motıon, epistemological belıefs, and self-effıcacy* (Publication No. 338278) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Yang, D., Streveler, R. A., Miller, R. L., Slotta, J. D., Matusovich, H. M., and Magana, A. J.. (2012). Using computer-based online learning modules to promote conceptual change: helping students understand difficult concepts in thermal and transport science. *International Journal of Engineering Education, 28*(3), 686-700.

*Yavuz, A. (2005). *Effectıveness of conceptual change ınstructıon accompanıed wıth demonstratıons and computer assısted concept mappıng on students' understandıng of matter concepts* (Publication No. 167430) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Yenilmez, A., & Tekkaya, C. (2006). Enhancing students' understanding of photosynthesis and respiration in plant through conceptual change approach. *Journal of Science Education and Technology, 15*, 81-87.

*Yılmaz, D. (2007). *The comparative effects of prediction/discussion-based learning cycle, conceptual change text, and traditional instructions on students' genetics understanding and self-regulated learning* (Publication No. 218053) [Doctoral dissertation, Middle East Technical University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Yılmaz, D., Tekkaya, C., & Sungur, S. (2011). The comparative effects of prediction/discussion-based learning cycle, conceptual change text, and traditional instructions on student understanding of genetics. *International Journal of Science Education, 33*, 607 - 628.

*Yılmaz, S., & Eryılmaz, A. (2010). Integrating gender and group differences into bridging strategy. *Journal of Science Education and Technology, 19*, 341-355.

*Yılmaz, S., Eryılmaz, A., & Geban, Ö. (2006). Assessing the Impact of Bridging Analogies in Mechanics. *School Science and Mathematics, 106,* 220-230.

*Yılmaz, Z. A. (2010). *Kavramsal değişim metinlerinin üniversite öğrencilerinin geometrik optik konusundaki kavram yanılgılarının düzeltilmesi ve fizik dersine karşı tutumlarına etkisinin incelenmesi* (Publication No.269347) [Doctoral dissertation, Ataturk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi.

*Yin, Y.M., Tomita, M.K., & Shavelson, R.J. (2014). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and achievement in science. *International Journal of Science Education, 36,* 531 - 552.

*Yumuşak, A., Maras, I., & Şahin, M. (2015). Effects of computer-assisted ınstruction with conceptual change texts on removing the misconceptions of radioactivity. *Journal for the Education of Gifted Young Scientists, 3,* 23-50.

*Yürük, N. (2007). The effect of supplementing ınstruction with conceptual change texts on students' conceptions of electrochemical cells. *Journal of Science Education and Technology, 16,* 515-523.

*Yürük, N., Beeth, M.E., & Andersen, C. (2009). Analyzing the effect of meta conceptual teaching practices on students' understanding of force and motion concepts. *Research in Science Education, 39,* 449-475.

*Zohar, A., & Kravetsky, S. (2005). Exploring the effects of cognitive conflict and direct teaching for students of different academic levels. *Journal of Research in Science Teaching, 42,* 829-855.

*Zohar, A., & Kravetsky, S. (2003). Cognitive conflict, direct teaching, and student's academic level. Annual meeting of the national association for research in science teaching. Philadelphia, PA, March 23-26.

APPENDICES

A. Final Draft of Coding Sheet

CODING SHEET				1. Publication Type	2. Year	3. Country:
Study No:	Title of the Study:	Author(s) Name, Lastname:	Surname, year	(Select)	(Select)	
1	Using analogy to overcome misconceptions about conservation of matterr	Ruth Stavy	Stavy, 1991	Journal Article	Published Year: 1991	Israel
1				Journal Article	Implementation Year: Received Year: Accepted Year : Published Year :	Israel
2	Using Elaborative Interrogation To Help Students Overcome Their Inaccurate Science Beliefs.	Vera E. Woloshyn; Allan Paivio; Mike Pressley	Woloshyn et al, 1992	Journal Article		CANADA
3	Effect of Enriched 5Es Model on Grade 7 Students' Conceptual Change Levels: A Case of 'Electric	Zeynel Küçük, Muammer Çalık	Kucuk and Calik, 2015	Journal Article	Published Year:2015	Turkey
4	Using Computer-based Online Learning Modules to Promote Conceptual Change: Helping Students Understand Difficult Concepts in Thermal and Transport Science	Dazhi Yang, Ruth A. Streveler , Ronald L. Miller , James D. Slotta, Holly M. Matusovich, Alejandra J. Magana	Yang et al, 2012	Journal Article	Published Year:2012	USA
4						
4						
5	Helping Students Understand Challenging Topics in Science through Ontology Training	James D. Slotta and Michelene T. H. Chi	Slotta and Chi, 2006	Journal Article	Implementation Year:2006	USA
6	Using computer-based online learning modules to promote conceptual change: helping students	Byron Levy Launey	Launey, 1995	Doctoral Dissertation	Published Year:1995	USA
7	The Effects of a Problem Solving Based Instructional Model on the Achievement and Alternative	Chun-Yen Chang,	Chang, 1995	Journal Article	Implementation Year:1995	Taiwan
8	A Computer-Assisted Instruction Unit on Diffusion and Osmosis with a Conceptual Change Design	Murray S. Jensen, Kimerly J. Wilcox, Jay T. Hatch, And Charles Somdahl	Jensen et al, 1994	Journal Article	Implementation Year:1994	USA
9	Using Analogies To Overcome Misconceptions: A Technology Course Example	Anu A. Gokhale	Gokhale, 1996	Journal Article	Published Year: 1996	USA
10	Using Computer Simulations to Enhance Conceptual Change: The Roles of Constructivist Instruction and Student Epistemological Beliefs	Mark Windschitl,;Thomas Andre	Windschitl and Andre, 1995	Journal Article	Implementation Year:1995	USA
10						
10						
10						
10						
10						

	4. Subject Area	5. Educational Stages	6. School Type	7. Demography	8. Gender Distribution	9. Sampling (select)		10. Group Design
Study No:	(Select)	(Select)	(Select)	(Select)	(Select)	Sampling Method	Sampling Method Type	(Select)
1	Chemistry	Middle (11-13)	Unspecified	Urban area(city center)	Unspecified	Nonrandom_Sampling	Convenience Sampling	Poor Experimental
1	Chemistry	Middle (11-13)	Unspecified	Urban area(city center)	Unspecified	Nonrandom_Sampling	Convenience Sampling	Poor Experimental
2	Physics	Middle (11-13)	Public school	Unspecified	Unspecified	Nonrandom_Sampling	Convenience Sampling	True Experimental
3	Physics	Middle (11-13)	Public school	Rural (town or smaller)	Unspecified	Nonrandom_Sampling	Convenience Sampling	Quasi Experimental
4	Physics	Undergraduate (18-)	Unspecified	Unspecified	Unspecified	Nonrandom_Sampling	Purposive Sampling	True Experimental
4								
4								
5	Biology	Undergraduate (18-)	Unspecified	Unspecified	Mixed	Nonrandom_Sampling	Purposive Sampling	True Experimental
6	Biology	Elementary (5-10 ages)	Public school	Urban area(city center)	Unspecified	Nonrandom_Sampling	Convenience Sampling	True Experimental
7	Earth Science	High school (14-17)	Public school	Urban area(city center)	Mixed	Nonrandom_Sampling	Convenience Sampling	Quasi Experimental
8	Biology	Undergraduate (18-)	Public school	Urban area(city center)	Mixed	Nonrandom_Sampling	Purposive Sampling	Quasi Experimental
9	Physics	Undergraduate (18-)	Public school	Urban area(city center)	Mixed	Nonrandom_Sampling	Purposive Sampling	Quasi Experimental

Study No:	11. Lenght of Treatment		12. Treatment duration (N. of course)	13. Treatment Intensity (For 1 week)	14. Teacher Training	15. Internal Validity Threats			
	Control Group	Treatment Group			Select	Researcher Effect (select)	Same Teacher Effect (select)	Confusion of Method and Medium Effect (select)	Level of Control Internal Validity (select)
1	Unspecified	Unspecified	Unspecified		unstated	Unspecified	Unspecified	Unspecified	Unspecified
1	Unspecified	Unspecified	Unspecified		unstated	Unspecified	Unspecified	Unspecified	Unspecified
2	Unspecified	Unspecified	Unspecified		unstated	Unspecified	Unspecified	Unspecified	Unspecified
3	6 hours	6 hours	6		unstated	Unspecified	Unspecified	Researcher does not control the method and medium	Unspecified
4 4 4	4-5 hours	4-5 hours	5		unstated	Unspecified	Same teacher for both control and experimental group	Researcher control the method and medium effect	Unspecified
5	Unspecified	Unspecified	Unspecified		unstated	Unspecified	Unspecified	Researcher control the method and medium effect	Unspecified
6	6x4x50 min	6x4x50 min	24	4	unstated	Researcher is not one of the teachers	Different teachers for control and experimental	Unspecified	Average
7	6 week	6 week	Unspecified		stated	Unspecified	Unspecified	Unspecified	
8	1 week	1 week	2	2	unstated	Unspecified	Unspecified	Unspecified	
9	50 min	50 min	1	1	unstated	Unspecified	Unspecified	Unspecified	Unspecified
10 10 10 10 10	7 hours (3 weeks)	7 hours (3 weeks)	7	2	unstated	Researcher is one of the teachers	Same teacher for both control and experimental group	Researcher control the method and medium effect	Unspecified

	16. Instrument Name		17. Instrument Type		18. Question Types in Diagnostic Instrument	19. Number of Question Stage	20. Application Time of Posttest	21. Delay Test	22. Treatment Verification
Study No:	Instrument 1	Instrument 2	Instrument 1 (select)	Instrument 2 (select)	(Select)	(Select)	(Select)	(Select)	(Select)
1	Unspecified		Researcher developed test		Open ended	1	Unspecified	No	Stated
1	Unspecified		Unspecified				Unspecified	No	Stated
2	proposition-generating task		Researcher developed test		Objective	1	Unspecified	No	Stated
3	Electricity concept test test		Researcher developed test		Mix	2	Unspecified	Yes	Unstated
4	Diffusion concepts assessments								
4	Microfluidics concepts assessments		Researcher developed test		Open ended	1	Unspecified	No	Stated
4	Heat transfer concepts assessments								
5	physics learning text (electricity)		Researcher developed test		Open ended	1	Unspecified	No	Stated
6	Drawing task		Researcher developed test		Open ended	1	Unspecified	Yes	Stated
7	alternative frameworks open-ended instrument		Researcher developed test		Open ended	1	Unspecified	No	Stated
8			Researcher developed test	Concept test	Open ended	2	Unspecified	No	Stated
9	Electric concept test		Researcher developed test		Objective	1	Unspecified	No	Stated
10									
10									
10	Concept Comparison Posttest	Multiple Choice Posttes	Researcher developed test	Unspecified	Mix	1		No	Stated
10									
10									

Study No:	23. Type of Conceptual Change Strategies (Select)	24. Instructional Material (Select)	25. Type of Integrated Methods to Conceptual Change Strategies? (Select)	26. Instructional Tool (Select)	27. Measuring Outcome (Select)	28. Statistical Analysis for Results (Select)	29. Sample Size	30. Average class size
1	Cognitive_bridging	Hands on	Unspecified	Analogy	Misconception Test	Chi-square	74	37
1	Cognitive_bridging	Hands on	Unspecified	Analogy		Chi-square	74	37
2	Cognitive_bridging	Text based	Student centered	Unspecified	Misconception Test	One way ANOVA	60	30
3	Cognitive_conflict	Hands on	Teacher centered	Specify: Animation, simulation, refutational text and worksheet	Misconception Test		46	23
4	Ontological_category	Computer based	Unspecified	Computer Modeling	Misconception Test		60	30
4				Computer Modeling				30
4				Computer Modeling				30
5	Ontological_category	Computer based	Student centered	Worksheets	Misconception Test		24	12
6	Cognitive_conflict	Computer based		Specify: Drawing task	Misconception Test		52	26
7	Unspecified	Text based	Student centered	Specify: Cooperative groups discussion	Misconception Test	Chi-square	172	86
8	Cognitive_conflict	Computer based			Misconception Test	MANOVA	184	20
9	Cognitive_bridging	Computer based		Analogy	Achievement Test	One way ANOVA	46	23
10	Cognitive_conflict	Computer based		Simulation	Misconception Test	MANCOVA	210	15
10								
10								
10								
10								
10								

Study No:		post test z value	post test t value	Posttets p value	post test F value	Chi square	Other Value	Effect Size cohens d	Effect Size Hedges' g	Std error	Weighted Hedges g	Weighte d Std	Weighted Varience	Weight
	31. Study Results													
	Instrument 1													
	Statistical Values (Instrument 1)													
1				0.001		29.2			1,598	0,302	1,214	0,274	5,563	10,97417206
1					17,103			0,975	0,965	0,243				16,88633756
2				0.001	8,8			0,766	0,756	0,264	0,756	0,264		
3									0,646	0,298	0,646	0,298		
4								0,560	0,544	0,260				14,83445216
4								0,600	0,579	0,260	0,373	0,458	4,187	14,76279766
4									0,011	0,255				15,39709121
5				0,0163	6,765				0,899	0,362	0,899	0,362		
6									1,278	0,301	1,278	0,301		
7				0.001		21.63		0,746	0,755	0,164	0,755	0,164		
8				0.001	12,47			0,538	0,546	0,157	0,546	0,157		
9				0.01	2,51				0,459	0,294	0,459	0,294		
10									0,371	0,195				26,38812426
10									0,320	0,227				19,35953361
10				0.05	3,99				-0,342	0,211	0,042	0,132		22,45527199

	31. Study Results													
	Instrument 1							Instrument 1						
	Control Group (Instrument 1)							Treatment Group (Instrument 1)						
Study No:	Number of Student	Pretest Mean	Pretest Standart Deviation	Posttest Mean	Posttest Standart Deviatio	Variation	Other	Number of Student	Pretest Mean	Pretest Standart Deviation	Posttest Mean	Posttest Standart Deviation	Variation	Other
1	37							37						
1	37							37						
2	30			4,6				30			5,67			
3	23	16,13	4,89	18,26	5,55			23	17,52	4,96	24,04	7,65		
4	30			13,87	2,88			30			15,40	2,67		
4	30			2,77	1,45			30			3,60	1,38		
4	30	14,03	5,61	15,93	6,24			30	14,63	5,06	16,60	6,23		
5	12							12						
6	26	14,25	2,13	23,22	3,82			26	13,33	1,52	27,98	4,87		
7	86							86						
8	63							121						
9	23			8,53	2,23			23			10,71	2,47		
10	55			0,53	0,5			51			0,71	0,46		
10	37			0,51	0,51			40			0,67	0,48		
10	42			0,45	0,5			48			0,27	0,54		

B. Coding Manual

Directions:

This coding sheet consists of 31 independent items as an excel sheet. Some items are multiple choices (open list), and some items are open-ended. For the items with multiple choices, it is required to choose an option from the open list. For open-ended items, it is expected to write short answers specified in this manual.

Near or below the items that are multiple choices, there is a sign that is written "select." The items with written "select" are expected to select one option the most appropriate for the study you are coding. Some of the items are expected to write short answers in provided spaces. If there is a colon (:) sign on the title, it means there is no open list for that column, and it is expected to write information to the provided space, for example, country, grade level, and sample size. In some of these items, it is expected to write the information in numerical form, and for some items, it is expected to write as word form. You can also find the figures showing how the coding is done for each item in this manual.

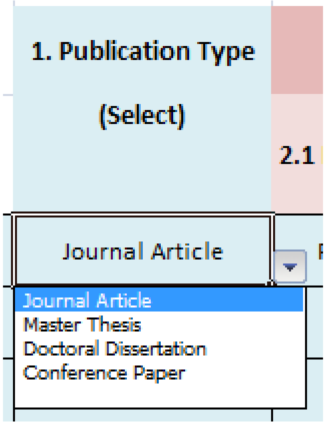
In some items (Example: item 2 and item 3), if the authors do not provide enough information about what is asked on the item, code it as "unspecified" by selecting or writing it explicitly to the provided spaces.

There is also the "Specify" option in the open list for some items. If there isn't an option in multiple choices, please explicitly write the answer to the "Specify" choice (Exp.: item 4).

The following instructions were provided to ease your coding process. It starts with a clear explanation of what you are expected to write for each item, and then (if necessary) some important points are highlighted. Please, read and try to follow the instructions as strictly as possible to establish consistency through the coding process.

1. Publication Type (Select)

In this item, you will select an option between an open list. The options are listed below with order.



Journal Article: Published articles in a journal about a specific topic.

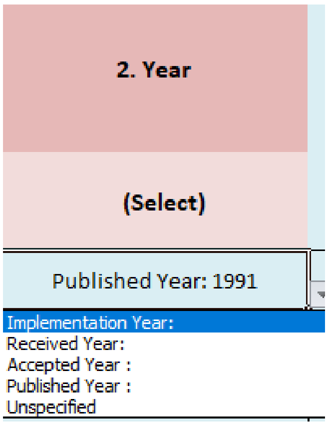
Master Thesis: The thesis written for the master degree.

Doctoral Dissertation: The thesis written for the doctoral degree.

Conference Paper: A scientific paper written for a conference.

2. Year

You will select an option between an open list with three choices in this item. Later, you will write just the year (not month or day) numerically to the item you selected from the open list. The choices are listed below in order.



Implementation Year: The implementation year is the year of treatment ends.

Received Year: This is the year the article was first submitted by the author.

Accepted Year: This is the year the article has first accepted the article.

Publication Year: This is the year the article was published by the article.

Unspecified: If it is not specified by the authors explicitly, please record as "Unspecified."

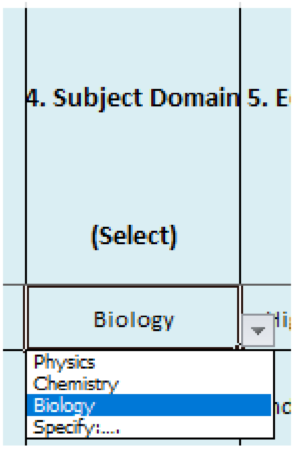
For the articles, write the received year of the study, which is the year that the author submitted article. If the received year is not indicated, record the year that article was accepted to be published in the journal. If the accepted year is not indicated, write the study's publication year. (Please follow the order).

3. Country

In this item, there is <u>no open list</u>. You will explicitly write the name of country/region to the provided space. Indicate the country where the study was implemented.

The country where the study has been published may be different from the one implemented. Be careful that, in this item, “country” refers to the one the treatment was implemented.

4. Subject Domain (Select)



You will select an option between an open list with five choices in this item. Select the appropriate subject domain for the study.

Some subjects can be expected for more than one subject domain, and researchers may not record the subject. Then, use the below list to decide.

Physics: Record the physics subject like work and energy, radiation, pressure, electromagnetism, force and motion, optic, mechanic, heat and temperature, scattering and reflection, electricity, and earth science.

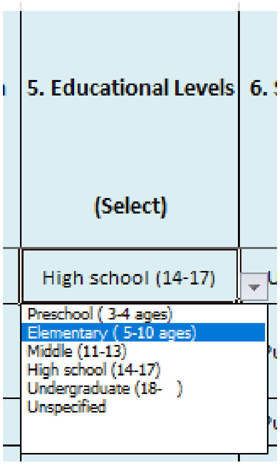
Chemistry: Record the chemistry subject like mol, periodical tables, gases, orbitals, atom models, matter, reactions, electrochemistry, physico-chemistry, acid and bases boundings, periodical table, conservation of matter, sink and floating,

Biology: Please select the biology subjects with respect to below list. Health, evolution, adaptation, heredity, living, microbiology, physicobiology, cells...

Specify: If you are unsure about the subject, please write the subject to the "specify" option.

Even if the subject is labeled as general science (or similar terms) on the paper, code this item as physics, chemistry or biology. For example, if the topic is "air pressure" label the subject as physics even if it is called as general science (or science and technology) on the paper by the authors.

5. Educational Level (Select)



In this item, you will select an option between an open list. Select the appropriate educational level. Every country has a specific educational system. Therefore, please firstly take into account the level to decide on educational levels by looking the below definitions.

Preschool (3-4 ages): This stage is also called pre-kindergarten, kindergarten, nursery, or first school. Children are 3 or 4 years old.

Elementary (Primary) (5-10 ages): This is the first stage found in formal education, beginning at about age 5 to 7 and ending at about age 10.

Middle School (11-13 ages): This stage is also called junior high school, beginning at about 11 and ending at about 13.

High School (Secondary) (14-17 ages): This stage is also called secondary school, beginning at about 14 and ending at about 17.

Undergraduate (18 - more): This stage begins at 18 and includes university, master, doctorate, or post-doctorate degrees

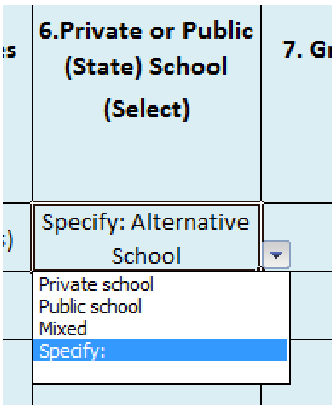
If the age range or stages are not stated, please record as unspecified.

For example, in USA educational stages 3rd grade for elementary school students is eight years old.

If the given age range includes two stages at the same time, please record the lower stage for consistency between coders

6. School Type (Select)

In this item, you will select an option between an open list. Select the appropriate school type.



Private School (Independent): These schools are founded and funded by a private organization, individuals, or religious groups (catholic schools) rather than by local, state, or national government. These schools are also known as independent schools. Some schools can also be founded by other states or missionaries, such as American schools and British schools.

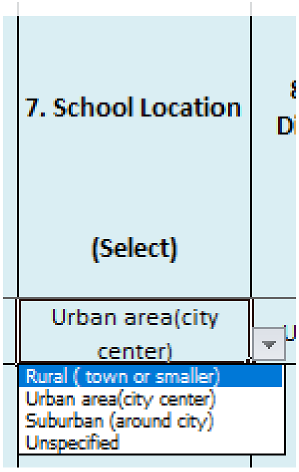
Public School (State-controlled): These schools are founded and funded totally by the government or local government (State schools, boarding schools, vocational schools)

Mixed: Some schools are founded by the private sector but funded by the State due to financial problems. Such as state-integrated schools or charter schools.

Specify: If there is more than one school type or if you are not sure.

7. School Location (Select)

In this item, there is no open list. Record the school location of the sample which is used in the research. Please use the following intervals.



Rural (town or smaller): A district called a town, village, or smaller.

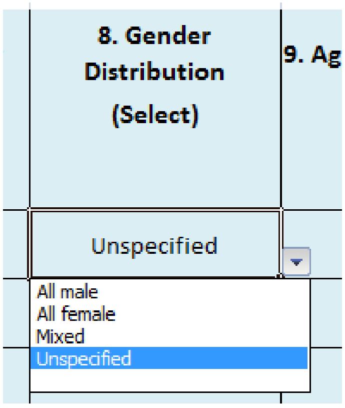
They generally have less than 50 thousand inhabitants. If it is larger than 50.000, please record as urban.

Urban Area (city center): A district called city, metropolis or larger.
They generally have more than 50 thousand inhabitants. If it is smaller and there is no information about district status, please record it as rural.

Suburban (around the city): A district that is not in the center of the city but around it.

Unspecified: If it is not specified by the authors explicitly, please record it as "Unspecified."

8. Gender Distribution (Select)



In this item, you will select an option between an open list.

Select the appropriate choice for the gender distribution of
the sample.

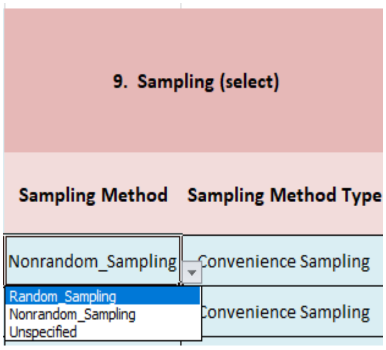
All male: All of the samples consist of males.

All Female: All of the samples consist of females.

Mixed: The sample consists of both males and females

Unspecified: The researcher is not stated the gender distribution.

9. Sampling (Select)



In this item, you will record two parts sampling method and sampling method type. At each part, there is an open list.

a) Sampling Method (Select)

In this item, you will select an option between an open list. Select the appropriate sampling method procedure for the study.

Random Sampling: "Selecting a representative sample from the population. Every member of the population presumably had an equal chance of being selected" (Fraenkel et al. 2012: p. 93, 8th ed.). Researchers try to get an accurate view of the population by selecting a representative sample from the population.

Nonrandom Sampling: "Each member of the population has no equal chance of being selected**,** some, in fact, may have no chance" (Fraenkel et al., 2012: p. 94). Each of the selected individuals must possess all the criteria defined by the researcher(s).

Unspecified: If there is no information about the sampling method of the study or if you do not think that the study is neither random sampling nor nonrandom sampling, please select this option.

Random Sampling Method Type (Select)

Simple Random Sampling: "In which each and every member of the population has an equal and independent chance of being selected" (Fraenkel, Wallen, and Hyun, 2012: p. 94).

For example:

A researcher wants to survey the academic achievement of elementary students in Ankara. The researcher should randomly select individuals from the list of students, including all students in Ankara.

Stratified Random Sampling: "This is a process in which certain subgroups, or strata, are selected for the sample in the same proportion as they exist in the population" (Fraenkel et al., 2012: p. 95). In this sampling, the researcher aims to ensure that certain characteristics of individuals in the population are represented in the same proportions.

For example:

A researcher wants to survey the academic achievement for elementary students in Ankara. Researcher creates strata as gender, age range, race, nationality, and socioeconomic status. A random sample from each stratum is taken in a number proportional to the stratum's size when compared to the population. These *subsets of the strata are then pooled to form a random sample.*

Cluster Random Sampling: "The selection of groups or clusters of subjects rather than individuals known as cluster random sampling. It is similar to simple random sampling except that groups rather than individuals are randomly selected" (Fraenkel et al., 2012: p. 96). When the population is too large to select individuals, the researcher divide the population into separate groups called clusters. Then some clusters are randomly selected from the population, not individuals.

Two-Stage Random Sampling: **"**It is often useful to combine cluster random sampling with individual random sampling. Rather than randomly selecting 100 students from a population of 3,000 ninth graders located in 100 classes, the researcher might decide to select 25 classes randomly from the population 100 classes and randomly select 4 students from each class" (Fraenkel et al., 2012: p. 97)**.**

Multistage Random Sampling: "If the sampling procedure is divided into three or more stages, this is called as multistage sampling. The sample is selected at least twice using different types of sampling techniques at each stage"(Alvi, 2016**).** This method does not require weighting, such as the stratified sampling method.

Example for multistage cluster sampling:

A researcher wants to conduct a survey for which the population is all elementary students in Turkey. It is very expensive and time-consuming to reach every student in Turkey. Instead, the researcher defines geographic region clusters as cities. Later, randomly select cities and randomly select schools from each city. Finally, the researcher can randomly select individuals from each selected school.

Non-Random Sampling Method Type (Select)

Systematic Sampling: "Every nth individual in the population list is selected for inclusion in the sample". For example, in a population list of 500 names to select a sample of 500, a researcher would select every tenth name on the list until reaching a total of 500 names "(Fraenkel, Wallen, and Hyun, 2012: p. 97).

Convenience Sampling: "This is a group of individuals who (conveniently) are available for study. A certain group of people was chosen for study because they are available"(Fraenkel et al. 2012: p. 99).

For example:

"A high school counselor interviews all the students who come to him for counseling about their career plans" (Fraenkel et al. 2012: p. 99).

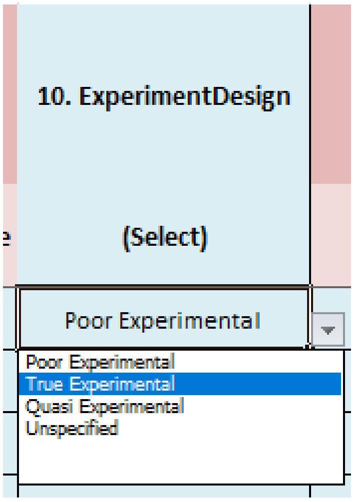
Purposive Sampling: "Purposive sampling is different from convenience sampling in that researchers do not simply study whoever is available samples but rather use their judgment to select a sample that they believe, based on prior information "(Fraenkel et al., 2012: p. 100).

For example:

The researcher asks the opinions of certain students who have specific characteristics that the researcher needs in a certain high school about attitudes about physics questions at the university entrance exam.

Unspecified: If there is no information about the sampling method of the study or if you do not think that the study is none of the above nonrandom sampling types, please select this option.

10. Experiment Design (Select)



In this item, you will select an option between an open list. Select the appropriate experimental design for the study.

Decide whether the research has been designed as poor, true, quasi-experimental, or factorial design. Make your decision based on an explanation of the details of the research. Your decision may not be the same as what the author(s) indicates about the type of research design.

Poor Experimental: "Designs that are weak do not have built-in controls for threads to internal validity. In addition to the independent variable, there are a number of other plausible explanations for any outcomes that occur" (Fraenkel et al., 2012: p.269).

Examples for Poor Experimental Designs:

- *The one-shot case study design*

X	O
treatment	observation

- *The one group pretest-posttest design*

O	X	O
observation (pre-test)	treatment	observation (post-test)

- *The static group comparison design*

X	O
treatment	observation
	O
	observation

- *The static group pretest-posttest design*

O	X	O
observation	treatment	observation
O		O
observation		observation

True Experimental: "True essential ingredient of a true experimental design is that subjects are randomly assigned to treatment groups" (Fraenkel et al., 2012: p.270).

For true experimental design, two conditions should be satisfied:

✓ There should be control and experimental groups

✓ Participants have to be randomly assigned to the sample groups

"Random assignment means that every participant in the experiment has an equal chance of being assigned to the experimental or control group while random selection means that every member of the population has an equal chance of being selected for the sample"(Fraenkel et al., 2012: p.267).

Examples for True Experimental Designs:

- The randomized posttest-only control experiment design

treatment group	R	X	O
	random	treatment	observation
control/ comparison group	R	C	O
	random	control/ comparison	observation

- The randomized pretest-posttest control group design

treatment group	R	O	X	O
	random	observation (pre-test)	treatment	observation (post-test)
control/ comparison group	R	O	C	O
	random	observation (pre-test)	control/ comparison	observation (post-test)

- The randomized Solomon four group design

Treatment group	R	O	X
	random	observation	treatment
Control group	R	O	C
	random	observation	control
Treatment group	R		X
	random		treatment
Control group	R		C
	random		control

- Random assignment with matching

a) The randomized posttest-only control group design, using match subjects

Treatment Group	M_r	X	O
	Random Matching	Treatment	Observation
Control Group	M_r	C	O
	Random Matching	Control	Observation

b) The randomized pretest-posttest control group design, using match subjects

Treatment Group	M_r	O	X	O
	Random Matching	Observation	Treatment	Observation
Control Group	M_r	O	C	O
	Random Matching	Observation	Control	Observation

Quasi-Experimental: "Designs that do not include the use of random assignment. Researchers who employ these designs rely instead on other techniques to control threads to internal validity"(Fraenkel et al., 2012, p.275).

For quasi-experimental design:

- ✓ There should be control and experimental groups
- ✓ There is no random assignment of participants but the random assignment of groups.

Examples for Quasi-Experimental Designs:

- The matching-only design

a) Matching only posttest only control group design

Treatment Group	M	X	O
	Matching	Treatment	Observation
Control Group	M	C	O
	Matching	Control	Observation

b) Matching only pretest-posttest control group design

Treatment Group	M	O	X
	Matching	Observation	Treatment
Control Group	M	O	C
	Matching	Observation	Control

- ***Counterbalanced designs***

A three-treatment counterbalanced design.

Group I	X_1	O	X_2	O	X_3	O
Group 2	X_2	O	X_3	O	X_1	O
Group 3	X_3	O	X_1	O	X_2	O

- ***Time series designs***

A basic time series design

O_1	O_2	O_3	X	O_4	O_5	O_6

- **Factorial Design:** "This is a modification of the pretest-posttest control group design. It involves one treatment and one control group and a moderator variable having two levels ($Y_{1\text{ and }}Y_2$)" (Fraenkel et al. 2012: p.277).

Treatment	R	O	X	O	Y_1	O
Control	R	O	C	O	Y_1	O
Treatment	R	O	X	O	Y_2	O
Control	R	O	C	O	Y_2	O

11. Intervention Length:

In this item, there is no open list. Record the total length of treatment for control and treatment groups separately, which is the time interval between the beginning and end of the implementation, as it is stated in the paper.

11. Intervention length	
Control Group	Treatment Group
5 weeks(5x3x50 min)	5 weeks(5x3x50 min)

Please record the order week, course hour per week, and course hour as min.

If there is just a year, month, week, day, or hour, write, as numerical form.

Exp.: 6 weeks, 4 hours, 2 months, 50 min,

If there is both week, course per week and course hour

Exp.: 6 weeks, 4 sessions per week, and one course is 50 min as 6x4x50 min.

If there is both month and week and day, write all of them by using "and" conjunction.

If there is a course hour and week or month, write the first course hour and later week in parenthesis

Exp.: 7 course hours (3 weeks), 210 min (2 months),

If there are both month, week and day write in order

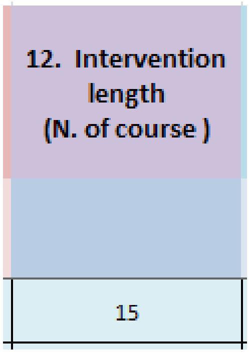
Exp.: 1 month and 2 weeks and 3 days (1x2x3 days)

If the intervention length is not certain, please record like 4-5 days, 2-3 weeks, 3-4 course hours, 40-50 min e.g.

Do not forget to specify the unit (i.e., year, month, week, day, or hour).

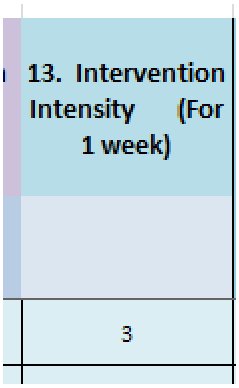
12. Intervention Length as Course Hour

In this item, there is no open list. Record total length of intervention length as a number of courses just for the treatment group, which is the time interval between the beginning and end of the implementation, as stated in the primary study.



13. Intervention Intensity

In this item, there is no open list. Please record how many courses were done for a week period. If there is no information, please record it as "unspecified" to the provided space.

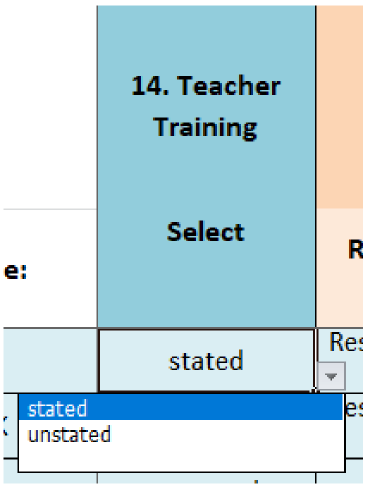


14. Teacher Training:

Please record whether the researcher stated that the implementers are trained or not?

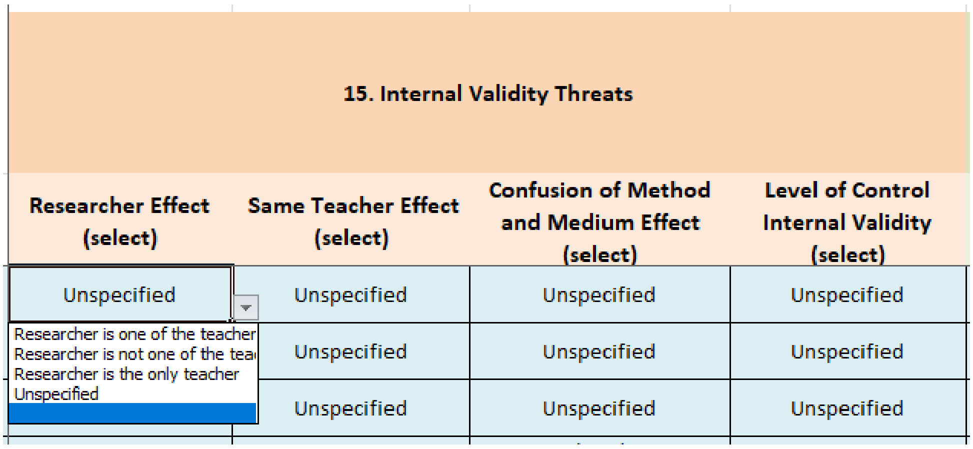
Stated: Researcher is done teacher training process and stated it.

Unstated: The researcher is not stated whether training has been done or not



15. Internal Validity Threats (select)

In this item, you will select an option between an <u>open list</u> for each item (researcher effect, same teacher effect, and method and medium confusion). "Internal validity means that observed differences on the dependent variable are directly related to the independent variable, and not due to some other unintended variable" (Fraenkel et al. 2012: p.166).

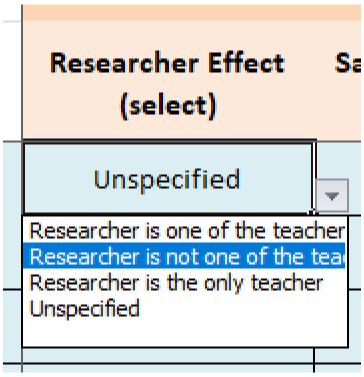


15. Internal Validity Threats			
Researcher Effect (select)	Same Teacher Effect (select)	Confusion of Method and Medium Effect (select)	Level of Control Internal Validity (select)
Unspecified	Unspecified	Unspecified	Unspecified
Researcher is one of the teacher Researcher is not one of the tea Researcher is the only teacher Unspecified	Unspecified	Unspecified	Unspecified
	Unspecified	Unspecified	Unspecified

Researcher Effect (select)

In this item, you will select an option between an <u>open list.</u>

Indicate whether researcher(s) has been involved in any of the control or experimental groups as a teacher.



Researcher is one of the teachers: If one of the researchers has been involved in the control or treatment groups, label the item as "Researcher is one of the teachers". If the regular teacher is one of the researchers, please label the item as "Researcher is one of the teachers".

Researcher is not one of the teachers: If the researchers have not been involved in the control or treatment groups, label the item as "Researcher is not one of the teachers" (if the regular teacher is not one of the researchers). However, if the researcher(s) were involved in the groups just to observe the lessons (for treatment verification or any other purpose) or to train teacher or help teacher for student orientation and for material use, and did not take part in instruction, then select “Researcher is not any of the teachers”.

Researcher is the only teacher: If the researcher is the only teacher and there is no regular teacher during the treatment label the item as "Researcher is the only teacher".

Unspecified: If there is no information, please record it as unspecified.

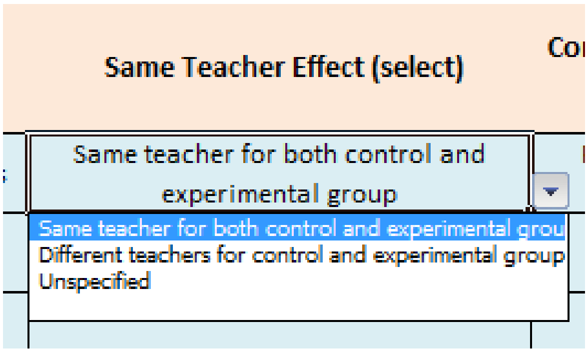
Same Teacher Effect (select)

In this item, you will select an option between an open list. Indicate whether the same teacher has instructed both control and experimental groups. This item does not aim to discriminate whether the researcher is one of the teachers.

Same teacher for both control and experimental groups: If both control and experimental groups were instructed by the same researcher, we should select “same teacher for both control and experimental group” option.

Different teachers for both control and experimental groups: We should select this item if the different researchers instruct both control and experimental groups.

Unspecified: Select this item if the researcher does not state the teachers for any group.



Confusion of Method and Medium Effect (select)

In this item, you will select an option between an <u>open list.</u>

Sometimes the effects of the method and medium are reported together. In this case, we cannot observe the effects of these two variables separately. Yet, the researchers sometimes report as if one of them is the only source of the effect.

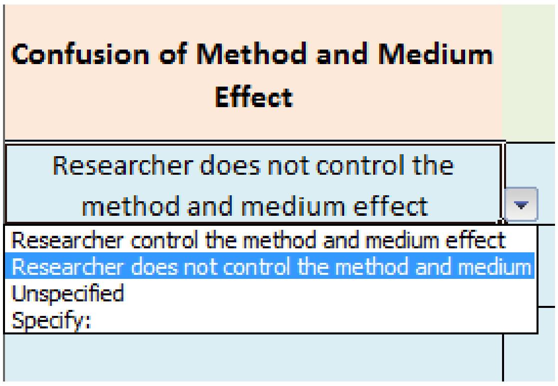
For example, an instructor may use demonstrations assisted with computer simulation. In this sense, the effect of computer simulation is not regarded, and the effect of demonstration is reported as the only source of effect or simulation may be supported with text-based material but simulation may be reported as the only source.

Researcher controls the method and medium effect: If the instructor reports the effect of both medium (computer, text, or hands-on medium) and method (conceptual change methods) separately, please record as "Researcher control the method and medium".

Researcher does not control the method and medium: If the instructor does not report the effect of both medium (computer, text, or hands-on medium) and method (conceptual change methods) separately, even though there exist both effects, please record as "Researcher does not control the method and medium".

Unspecified: If there is no information about the medium or you do not realize the medium clearly, please record it as "Unspecified."

Specify: In this option, it is expected to <u>write your answer explicitly</u>. If you think that there is no effect of medium or you have any other idea, please specify your idea for this item.



Level of Control over Threats to Internal Validity (select)

In this item, you will select an option between an open list. Decide the extent to which threats to internal validity have been controlled using the following list of possible threats and criteria:

Subject Characteristics, Loss of Subjects (Mortality), Location, Instrument Decay, Data Collector Characteristics, Data Collector Bias, Testing, Extraneous Events (History), Maturation, Attitude of Subjects, Regression.

None: None of the threats to internal validity was controlled

Poor: 1-3 of the threats to internal validity were controlled

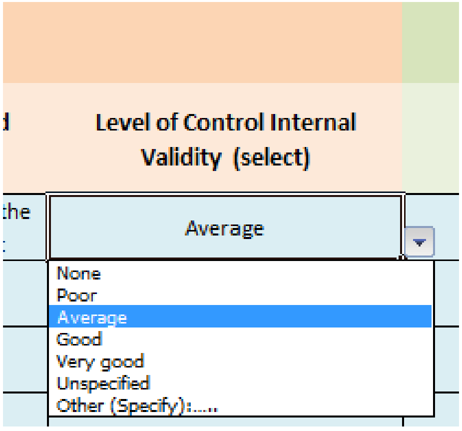
Average: 4-6 of the threats to internal validity were controlled

Good: 7-9 of the threats to internal validity were controlled

Very Good: All of the threats to internal validity were controlled

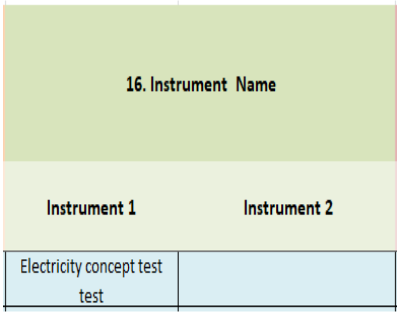
Be aware of that:

That the authors do not mention how they have controlled the possible threats to internal validity does not necessarily mean that they have not done anything for these threats. In this item, the quality of research design and reporting quality may interfere and it can be hard to make judgments about the threats to internal validity and the measures taken into control to distinguish which result provide no information about internal validity. So, to be able to get a fair judgment about the degree of internal validity, we should take care of the research design in study. Thus, any finding from this item is limited by what is *reported on the paper by the authors.*



16. Instrument Name:

In this item, there is no open list. Record the instrument name(s) used in the research. Just record the instruments providing quantitative data, do not record qualitative instruments like interviews, observations, records, reports, etc. There are three colons to record instrument names. Please write other instrument names to the instrument three colon if you need more colons.



17. Instrument Type (select)

In this item, you will select an option between an <u>open list.</u> Select the appropriate type of instrument.

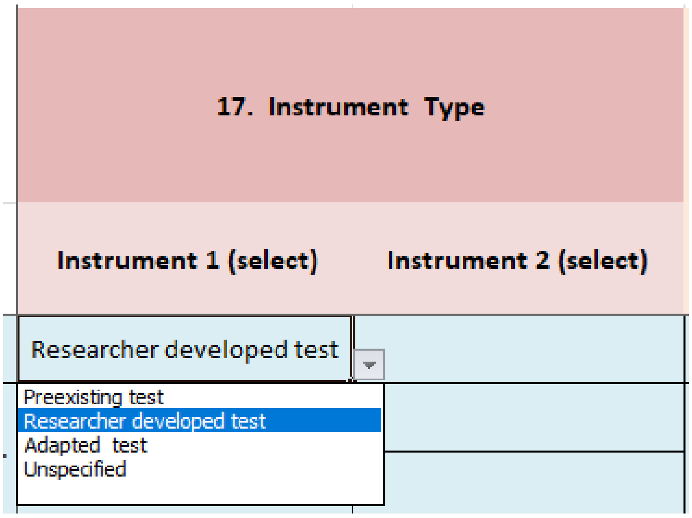
Pre-existing test: Refers to the tests that have already been developed by other researchers and available in the literature. These tests do not have to be standardized ones. Just being pre-existing is enough to label the test as a pre-existing test.

Researcher-developed test: The authors developed the test for this study. The test had not been available in the literature before this study, and it is totally original, not an adaptation of the pre-existing test.

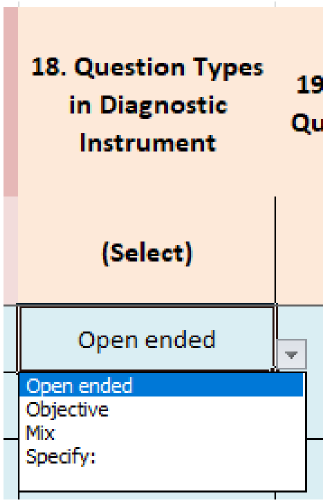
Adapted test: Refers to the tests that have been adapted from one or more pre-existing tests for this study by the authors. However, the adapted version of the pre-existing test has not been used before for another study.

Unspecified: If there is no information about the test type of the study or if you do not think that the study is not fit to the above test types, please select this option.

If there are more than three assessment instruments, please add extra columns for them to code the necessary information.



18. Question Type in Diagnostic Instrument (select)



In this item, you will select an option between an open list. Select the appropriate type of question that is used in the diagnostic instrument.

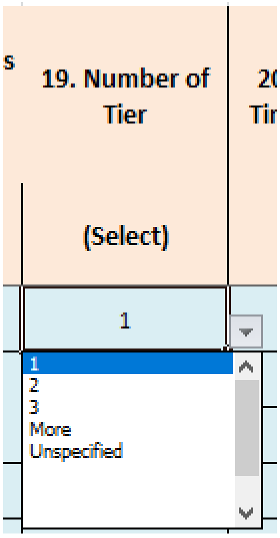
Open-ended: The respondent must elaborate on their ideas by writing without answering simple yes or no, true or false.

Objective: It requires selecting the correct answer from among one or more of several alternatives or supplying a word or two, and that demands an objective judgment when it is scored like multiple choices, scales, true-false, yes-no questions.

Mix: It requires to answer for open-ended and objective-type questions in the same question. It is common to use in multiple-tiered questions.

Specify: If you are not sure about the type, please record it here by defining your reason explicitly.

19. Number of Tiers (select)



In this item, you will select an option between an open list.

One Tier: Includes only one question root and one tier.

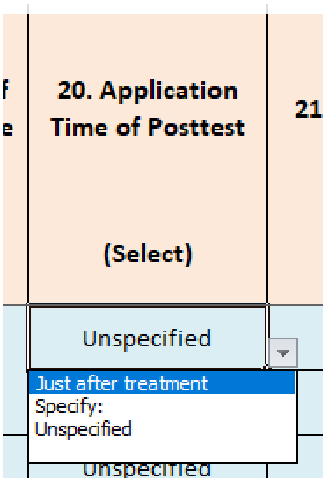
Two Tier: Includes two-tier items. It is commonly used in the format that multiple-choice at the first tier is used and on the second tier, the reasons of answer on the first tier are used.

Three Tier: Includes three tier items. It is commonly used in the format that multiple-choice at the first tier is used, on the second tier, the reasons of answer on the first tier is used, on the third tier, it is questioned on the certainty about the answer.

More: If there are more than three-tier, please record them here.

Unspecified: If there is no information about questions, please record them here.

20. Application Time of Post Test (select)



In this item, you will select an option between an <u>open list.</u>

Just After Treatment: If the diagnostic post-test is applied just after the treatment or on the same day with treatment, please record it here.

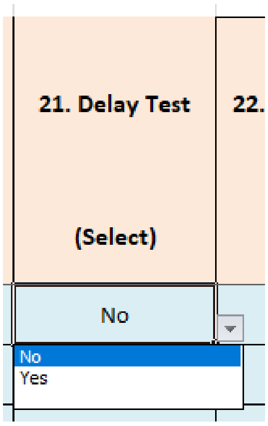
Specify: The diagnostic post-test is applied after a certain time interval, please specify the time interval.

For example 1 day, 1 week ...

Unspecified: If there is no information about the application

time of post-test, please record here.

21. Delay Test (select)

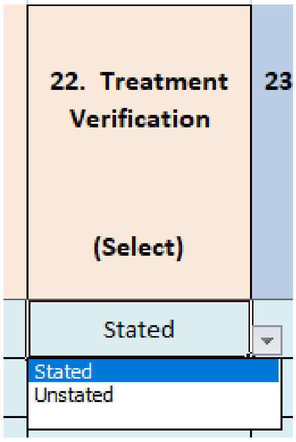


In this item, you will select an option between an <u>open list.</u>

No: If there isn't a delay test, please record it here.

Yes: If there is a delay test, please record it here.

22. Treatment Verification (select)

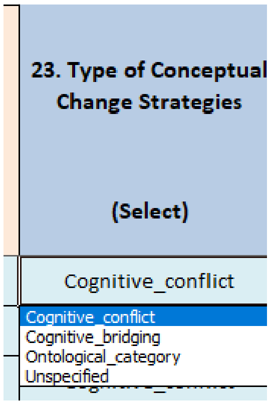


In this item, you will select an option between an open list. Please record whether the researcher is stating that the treatment verification is done or not?

Stated: Researcher did treatment verification and stated it.

Unstated: Researcher does not state whether treatment verification has been done or not.

18. Type of Conceptual Change Strategies (select)



In this item, you will select an option between an open list. Please read the following explanations carefully to code type of conceptual change strategy.

Conceptual change strategies include any instructional methods with an explicit aim of changing students' specific misconceptions and helping them construct scientific conceptions with a well-defined instructional sequence.

To identify an instructional method as a "conceptual change strategy":

- ✓ The researcher(s) should explicitly address some specific misconceptions
- ✓ There should be an explicit intention of changing misconceptions with well-defined instructional steps.
- ✓ The researchers should follow one of the conceptual change strategy steps.

- ➢ **Cognitive Conflict** is based on the classical conceptual change model developed by Posner, Strike, Hewson and Gertzog (1982). Focuses on students' misconceptions. The major aim of the instruction is to locate students' misconceptions, and falsification of these misconceptions and then help students' gain the scientific conception.

- **Cognitive Bridging:** It is based on the theoretical arguments of Smith (1992) and diSessa's (2002) theoretical arguments on the element knowledge perspective. Focuses on students' productive conceptions. The major aim of the instruction is to locate students' productive pre-conceptions and then use them to help students gain scientific conception.
- **Ontological Category Shift:** It is based on the theoretical arguments developed by Chi (1993; 1994; 1995; 2002). Focuses on the ontological nature of students' conceptions. The major aim of the instruction is to identify the ontological category of students' conceptions and then helping students shift their conceptions into an appropriate ontological category.
- **Unspecified:** If there is no information about the study method of the study or if you do not think that the method does not fit the above types, please select this option.

Cognitive Conflict:

According to Piaget (1964), there are two major phases of conceptual change assimilation and accommodation. Firstly, it is critical to assimilate the current knowledge. But the current knowledge may not be appropriate or adequate to grasp scientific knowledge. Therefore students should replace or reorganize their prior knowledge. Posner et al. (1982) assert the idea that the radical phase of the restructuring process to achieve conceptual change is called accommodation. The accommodation process includes well-defined and in order steps that should be fulfilled in order to bring about conceptual change. Therefore the following four conditions should be satisfied to achieve accommodation that is likely to take place. The first condition of the cognitive conflict process is dissatisfaction with existing knowledge. Hewson (1992) put forward the notion that dissatisfaction is a reason for changing status of prior knowledge. When individuals find their current concepts unreasonable to reorganize or to replace with a new one, individuals are not satisfied with their prior knowledge and may not want to retain their current concepts. This process prompts learners to question the effectiveness of their prior knowledge. Consequently, cognitive conflict implies the inoperativeness of prior knowledge so that it should be extended or exchanged with new knowledge.

The properties of new knowledge are critical to activating conceptual change. The second condition proposes that new knowledge should be intelligible (the learner should know what the new knowledge is). Posner et al. (1982) stated that intelligibility requires understanding concepts, terms, symbols, or identifying representations of what the functions and theories are saying.

Thirdly, new knowledge should be plausible. Posner et al. (1982) define it as the new knowledge that should be consistent with current scientific knowledge. In other words, it can be defined as the capacity of presented knowledge to solve problems. Plausibility enables us to enhance explanatory power with regard to the difficulties faced by students attempting to learn concepts.

Fourthly, new knowledge should be fruitful, which can suggest new insights and discoveries when encountering new situations. When new knowledge is both intelligible and plausible, students may interpret new experiences to resolve problems. Therefore, fruitful new knowledge provides the accommodation process more persuasive and permanent for students.

Cognitive Bridging:

Although there is no precise definition for cognitive bridging strategy, it can be shortly defined as using productive prior knowledge to construct and impose scientific knowledge without focusing on conflicting processes. Yaman (2013) stated that cognitive bridging is an instructional strategy that uses students' existing knowledge elements. The main assumption is that students come to class with lots of resources gained from daily life experiences. These resources may provide better acquisition of new knowledge. A number of experimental studies also sign the effectiveness of this strategy on the conceptual change process.

diSessa (1998) stated that, with respect to the effectiveness of the cognitive bridging perspective, some instructional implications should be taken into consideration. Firstly, adequate time is needed for better conceptual understanding and achieving deep results from instruction. The cognitive bridging strategy accepts conceptual change as a longer-term process in contrast to the conflict perspective.

Secondly, the richness of conceptual resources should be used productively rather than dissatisfied with them. The bridging perspective implies a link between existing knowledge and new knowledge to achieve conceptual change. With respect to instructional implications, it is critical to attend to students carefully in a classroom environment by using their experiences which are relevant. Clement (1993) focuses that analogies effectively trigger relevant experiences in the learning process even if there are naïve concepts in some contexts causing misconceptions. The critical argument is that the learner should activate his prior knowledge to modify, displace, replace or suppress it. Otherwise, developing new conceptions may not be possible.

Thirdly, one of the main concerns of bridging perspective is that coaching meta-conceptual awareness enables us to develop scientific knowledge by constructing prior knowledge. In this way, the learner can differentiate productive prior knowledge pieces. An effective coaching process provides more healthy learning for conceptual change.

Finally, assessing learning outcomes should be different from the classical conceptual change perspective. Coherent and unfragmented prior knowledge structure can be monitored with classical assessment tools. On the other hand, the element type of prior knowledge should be considered with its context and need to reveal the relations. Therefore, more comprehensive and process-oriented measuring tools enable observing learning outcomes. diSessa (1998) proposes that more formative assessment tools should be used to consider and monitor the effects of contextual elements. It provides more valid and informative results also.

Ontological Category Shift

As a different perspective from classical conceptual change, the process proposes that misconceptions arise from the incorrect assignment of concepts in a lateral category. In this sense, the ontological category adopts the conceptual change process markedly different from classical conceptual change than direct falsification for prior conceptions. Chi and Slotta (1993) indicate that creating conflict may not necessarily provide conceptual change. They also address the idea that conceptual change is possible by removing misconception, which is the mis-categorization of knowledge in the absence of the correct lateral category. This is a progressive and gradual process rather than a direct accommodation**.** Chi and Roscue (2002) define conceptual change as the shift of mis-categorized knowledge from one ontological category (mis-category) to a workable (true) ontological category.

The major point of view for the ontological category process is creating radical conceptual change. Such a change requires transformation between ontological categories. Learners should change their knowledge of mis-category into a scientifically true one. Therefore, instructional implications for triggering conceptual change processes should be more structured. The instruction should;

i) begin with describing the attributes of the existing ontological category

ii) Secondly, the attributes of the new ontological category should be defined

iii) Thirdly, learners should understand the meaning of individual concepts in a new ontological category to advance new conceptual understanding.

iv) Finally, learners should reassign old ontological categories to a new ontological category to assimilate the new knowledge (Chi and Slotta,1993).

24. Instructional Material (select)

In this item, you will select an option between an <u>open list.</u> Please read the following explanations carefully for code-type instructional material.

Text-based: Record this choice when the instructor uses text-based materials(text, maps or drawing type materials) to facilitate conceptual change.

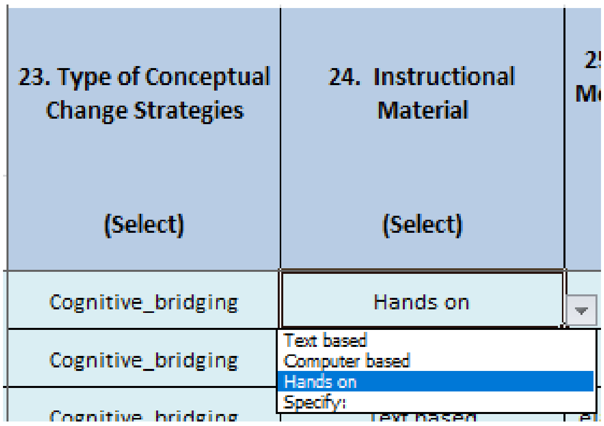
Computer-based: When the instructor uses computer-based materials (demos, simulations, animations, films) to facilitate conceptual change, record this choice.

Be aware of that:

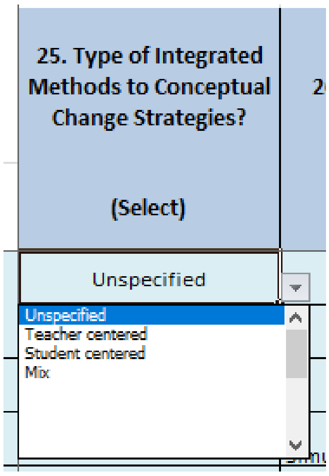
The instructor may use text, maps, or drawings on a computer program, please record in this item.

Hands-on: The instructor may not use text-based materials or computer-based materials. Laboratory activities, workshops, and outdoor applications are hands-on materials. (Example: demonstrations, inquiry-oriented applications, workshops, etc. studies in a classroom, laboratory, or outside)

Specify: if you do not think that the medium is not fit to above types, please explicitly write the material to this option



25. Type of Integrated Method to Conceptual Change? (select)



In this item, you will select an option between an <u>open list.</u> Please read the following explanations carefully to code of integrated methods for conceptual change.

Unspecified: If there is no information about the integrated methods or if you do not think that the study is not fit the below types, please select this option.

Student-Centered Methods

This kind of method enables learners to reach new knowledge by making inferences. In this method, the student is expected to collect data, analyze, synthesize and make inferences. Following methods are some examples of student-centered methods:

- ✓ Inquiry learning methods
- ✓ Problem-solving method
- ✓ 3E-5E- 7E learning methods
- ✓ Predict-observe-explanation method
- ✓ Elaboration learning methods
- ✓ Project-based learning

Teacher-Centered Methods

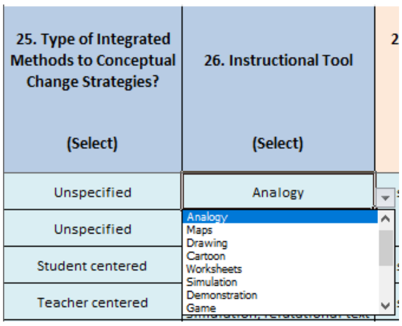
These methods are also called "traditional methods", "conventional methods", "lecture-based methods," or "expository methods" in literature. In this method, the student is not expected to collect, synthesize or analyze data. Knowledge is transferred directly to the student by means of direct instruction, discussion, simulation, demonstration, etc. Following methods are some examples of traditionally designed methods:

- ✓ Direct instruction (lecturing)

- ✓ Meaningful learning
- ✓ Kinesthetic learning
- ✓ Classroom discussions
- ✓ Simulation-based learning
- ✓ Flipped classroom
- ➢ Mix Methods

Researchers use both student-centered and teacher-centered methods simultaneously or orderly during the instruction on the same sample. The order or the emphasis on methods is not important.

26. Instructional Tool (Select)



In this item, you will select an option between an open list. In many conceptual change instructions, supporting instructional tools like analogies, maps, cartoons, games… are used to assist the conceptual change process. If there are more than one tool please record each of them to the provided spaces by using comma "," between tools. The following materials are some of the examples of instructional tools:

Analogy: A tool to compare two things to make new information intelligible to learners by comparing it to the information that is already familiar to them.

Maps: Concept maps, conflict maps, or cognitive maps to facilitate conceptual change.

Drawing: Any drawing that assists conceptual change method.

Cartoon: Caricatures or cartoons that assist conceptual change method.

Worksheet: Any worksheet that assists conceptual change method.

Simulation: Any interactive or expository computer program that assists conceptual change method.

Demonstration: Any demonstration that assists conceptual change method.

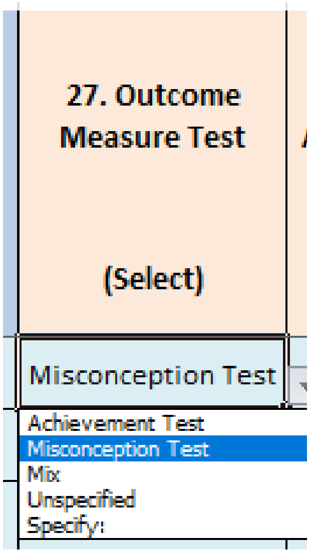
Game: Computer games or hands-on games that assist conceptual change method.

Models: Computer models or hands-on models that assist conceptual change method.

Animation: Any computer animation that assists conceptual change method.

- ✓ **Specify:** if you do not think that the tool is not fit the above types, please explicitly write the tool to this option.
- ✓ **Unspecified:** If there is no information about the integrated methods or if you do not think that the study is not fit the above types, please select this option.

27. Outcome Measure



Achievement Test: The diagnostic test aims to measure student general achievement rather than conceptual understanding or the change in misconception.

Misconception Test: The diagnostic test aims to measure change in misconceptions or the degree of conceptual understanding of the subject where students have misconceptions.

Mix: The diagnostic test that includes both general achievement and change in misconceptions.

Unspecified: If there is no information about the content of the diagnostic test, please record it as unspecified.

Specify: If you do not think that the diagnostic tool's content does not fit the above types, please explicitly write your idea to this option.

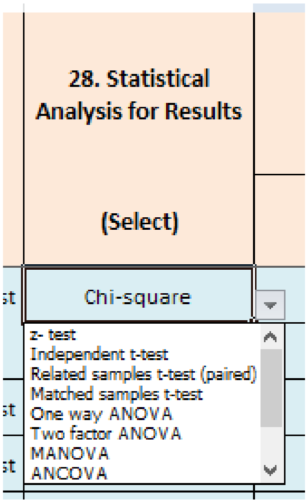
28. Statistical Analyzes (Select)

In this item, you will select an option between an <u>open list.</u>

You are expected to decide which type of statistical analyses

Were used to evaluate data.

If the researcher(s) is/are used <u>more than one analysis</u> please record each of the analyses to the "other" option.



z-Test: "Tells the distance between the score and the mean in terms of the number of standard deviations. The sign tells whether the score is located above (+) or below (-) the mean." (Gravetter and Wallnau, 2009: p. 141.) The overall z-formula;

$$z = \frac{X - \mu}{\sigma}$$

X= Score within a distribution

μ= Mean value for distribution

σ= Standard deviation

Independent t-test: "A research design that uses a separate sample for each treatment condition (or for each population) or the difference between two sample means to evaluate the difference between two population means". (Gravetter and Wallnau, 2009: p. 309-311.)

The overall t-formula;

$$t = \frac{samplemean - populationmean}{estimatedstandarderror}$$

Repeated samples t-test: "This is one in which a single sample of individuals is measured more than once on the same dependent variable." (Gravetter & Wallnau, 2009: p. 341.)

Matched Samples t-test: "Each individual in one sample is matched with an individual in the other sample." (Gravetter & Wallnau, 2009: p. 341.)

One Way ANOVA: "This is a hypothesis testing procedure that is used to evaluate mean differences between two or more treatments(or populations)." (Gravetter & Wallnau, 2009: p. 394.)**.** There is one independent variable and one dependent variable.

	Before	After	6 months after
Group 1	Scores group 1	Scores group 1	Scores group 1

Two Factor ANOVA: The researcher wants to investigate the combined effect of two independent variables (factors) on one dependent variable. There are more than one independent variable and one dependent variable in this design.

Treatment Time: 3 levels

Age: 2 levels

	Before	**After**	**6 months after**
Young	Young- Before therapy score	Young- After therapy score	Young- 6 months after therapy score
Elder	Elder- Before therapy score	Elder- After therapy score	Elder - 6 months after therapy score

MANOVA:"Multivariate analysis of variance (MANOVA) is simply an ANOVA with several dependent variables. That is to say, ANOVA tests for the difference in means between two or more groups, while MANOVA tests for the difference in two or more vectors of means." (French, 2008)

ANCOVA: "One-way analysis of covariance is designed to assess group differences on a single DV after the effects of one or more covariates are statistically removed. For example, age and degree of reading disability are usually related to the outcome of a program of educational therapy (the DV)." (Tabachnick and Fidell, 2012 p.19)

MANCOVA: "In addition to dealing with multiple DVs, multivariate analysis of variance can be applied to problems when there are one or more covariates. In this case, MANOVA becomes a multivariate analysis of covariance—MANCOVA." (Tabachnick and Fidell, 2012 p.21)

Chi-square: "Analysis of variance examines the relationship between a discrete variable (the IV) and a continuous variable (the DV); correlation and regression examine the relationship between two continuous variables, and the chi-square (x2) test of independence is used to examine the relationship between two discrete variables." (Tabachnick and Fidell, 2012 p.49)

Statistical Techniques for Ordinal Data: When the data is ordinal, parametric tests cannot be used instead an alternative tests use.

- **"Mann-Whitney Test** uses data from two separate samples to evaluate the difference between two treatment conditions or two populations. This can be viewed as an alternative to the independent measures t hypothesis test." (Gravetter and Wallnau, 2009: p.666.)
- "**Wilcoxon Test** uses data from a repeated measures design to evaluate the difference between two treatment conditions. This can be viewed as an alternative to the repeated measures t hypothesis test." (Gravetter and Wallnau, 2009: p.666.)
- "**Kruskal Wallis Test** uses data from three or more separate samples to evaluate the difference between three or more treatment conditions. This can be viewed as an alternative to the single factor, independent measures ANOVA test." (Gravetter and Wallnau, 2009: p.666.)
- "**Friedman Test** uses data from a repeated measures design to compare the difference between three or more treatment conditions. This test is an alternative to the repeated measures ANOVA test." (Gravetter and Wallnau,

2009: p.666.)

29. Sample Size:

In this item, there is <u>no open-list.</u> Please record the total

number of students that include in treatment group and control group.

29. Sample Size
74

30. Average Class Size:

In this item, there is <u>no open-list.</u> Please record the average

class size just for the treatment group.

30. Average class size
37

31. Study Results:

There is a brief explanation for coding. It is expected you record the data that enable the computation of effect size value if no data is given, leave the item as blank.

Note: It is critical to record treatment-control group sizes, posttest and pretest mean and standard deviation scores. If there is no descriptive data, please record other inferential scores (F, Chi-square, t, z, gain, U test and so on). Please record the effect size value if it is written.

Statistical Values

Record each quantitative result for treatment at instrument 1. It is expected to record values providing effect size for treatment effect. Do not record the interaction scores, covariate scores, error scores, or other quantitative scores.

Table-2: ANCOVA Summary (Group vs. Achievement)

Source	Sum of squares	df	Mean square	F	p
Covariate (Pre TCE)	117.79	1	117.79	24.11	0.00*
Treatment	472.62	1	472.62	96.72	0.00*
Gender	6.77	1	6.77	1.39	0.243
Treatment * Gender	30.61	1	30.61	6.26	0.014*
Error	376.27	77	4.89		

* $p < 0.05$

Figure 33. Recording part for ANCOVA results

Outcome Type: There is an open list in this item. Record the dependent variable (achievement, attitude, motivation or skill scores) investigated by the researcher. If there is more than one dependent variable, please record by using "-" between variables and the scores in order. This is an example of coding;

Table 5
Follow-up Univariate ANCOVAs of Dependent Variables.

Source	Dependent variable	*df*	MS	Univariate *F*	*p*
PRESCA	POSTSCA	1.00	372.09	37.48	.00
	ATC		25.38	0.09	.76
Group	POSTSCA	1.00	375.75	37.85	.00
	ATC		24.88	0.41	.52

PRESCA: Pre-solution concept achievement; POSTSCA: Post-solution concept achievement; ATC: Attitude Towards Chemistry; MS: Mean square; N = 87.

Record z-value, t value, p-value, F-value, and other given values. If the effect size is given, you should record the value details like Cohen's d= …, Glass Δ=…., Hedges g=..... , Eta square= ……, chi square=…… and others. Please use the dot "." to write decimal numbers.

Instrument 1													
Statistical Values (Instrument 1)													
post test z value	post test t value	Posttets p value	post test F value	Chi square	Other Value	Effect Size cohens d	Effect Size Hedges' g	Std error	Weighted Hedges g	Weighted Std Error	Weighted Varience	Weight
		0.001		29.2			1,598	0,302	1,214	0,274	5,563	10,97417
			17,103			0,975	0,965	0,243				16,88634
		0.001	8,8			0,766	0,756	0,264	0,756	0,264		

If there is more than one instrument for quantitative results, please record other items (instrument 2 and instrument 3). If there are more, please add additional colons. If there is a different value to compute an effect size, please record the "Other Value" item.

Record each item for each instrument and groups (control and treatment) separately. You should record the sample size, average class size, number of students, post-test and pretest means for the groups (control and treatment), and post-test and pretest standard deviations for the groups (control and treatment).

Instrument 1								
		Control Group (Instrument 1)						
Sample Size	Average class size	Number of Student	Pretest Mean	Pretest Standart Deviation	Posttest Mean	Posttest Standart Deviatio	Variation	Other
74	37	37						
74	37	37						

Other Value: If there is a value that is not stated in the coding sheet, record the value in a given space.

C. List of Effect Sizes Revealed From Primary Studies

No	Surname, Year	Effect Size (Hedges' g)	Std. Deviation	Publication Type
1	Stavy, 1991	1.214	0.274	Journal Article
2	Woloshyn, Paivio, & Pressley, 1994	0.756	0.264	Journal Article
3	Kucuk & Calik, 2015	0.646	0.298	Journal Article
4	Yang, Streveler, Mıller, Slotta, Matusovıch, & Magana, 2012	0.373	0.458	Journal Article
5	Slotta & Chi, 2006	0.899	0.362	Journal Article
6	Launey, 1995	1.278	0.301	Doctoral Dissertation
7	Chang & Barufaldi, 1999	0.755	0.164	Journal Article
8	Jensen, Wilcox, Hatch, & Somdahl, 1996	0.546	0.157	Journal Article
9	Gokhale, 1996	0.459	0.294	Journal Article
10	Windschitl & Andre, 1996	0.042	0.132	Journal Article
11	Saigo, 1999	0.190	0.214	Doctoral Dissertation
12	Seyedmonir, 2000	0.030	0.371	Doctoral Dissertation
13	Sanger & Greenbowe, 2000	0.934	0.246	Journal Article
14	Diakidoy & Kendeou, 2001	1.383	0.300	Journal Article
15	Mikkila-Erdmann, 2001	0.233	0.142	Journal Article
16	Sungur, Tekkaya, & Geban, 2001	0.953	0.298	Journal Article
17	Çakır, Geban, & Yürük, 2002	0.734	0.224	Journal Article
18	Eryilmaz, 2002	0.169	0.174	Journal Article
19	Uzuntiyaki, 2003	2.457	0.405	Doctoral Dissertation
20	Tekkaya, 2003	0.586	0.304	Journal Article
21	Alparslan, Tekkaya, & Geban, 2003	0.807	0.250	Journal Article
22	Niaz and Chac´o, 2003	1.090	0.449	Journal Article
23	Çetin, 2003	0.700	0.227	Doctoral Dissertation
24	Tsai, 2003	-0.317	0.145	Journal Article
25	Charles, 2003	0.036	0.569	Doctoral Dissertation
26	Bozkoyun, 2004	1.310	0.291	Doctoral Dissertation
27	Ayhan, 2004	1.384	0.380	Doctoral Dissertation
28	Özkan, Tekkaya, & Geban, 2004	0.780	0.269	Journal Article
29	Azizoğlu, 2004	1.462	0.224	Doctoral Dissertation
30	Çelebi, 2004	0.599	0.307	Master Thesis
31	Ceylan, 2004	3.939	0.437	Doctoral Dissertation
32	Uzuntiyaki & Geban, 2005	1.013	0.263	Journal Article
33	Çetingul, 2006	2.016	0.354	Journal Article
34	Gulcicek, 2004	1.144	0.314	Master Thesis
35	Çetingul & Geban, 2011	1.267	0.306	Journal Article
36	Yavuz, 2005	0.885	0.240	Doctoral Dissertation
37	Zohar & Kravetsky, 2005	-0.115	0.217	Journal Article
38	Günay, 2005	1.411	0.329	Doctoral Dissertation

39	Demircioğlu, Ayas, & Demircioğlu, 2005	1.060	0.226	Journal Article
40	Savinainen, 2005	0.527	0.298	Journal Article
41	Chiu & Lin, 2005	1.358	0.937	Journal Article
42	Bilgin & Geban, 2006	2.250	0.273	Journal Article
43	Bilgin & Geban, 2001	1.020	0.339	Journal Article
44	Yilmaz, Eryılmaz, & Geban, 2006	2.297	0.236	Journal Article
45	Yilmaz & Eryılmaz, 2010	0.965	0.121	Journal Article
46	Baser, 2006a	0.795	0.219	Journal Article
47	Baser, 2006b	2.012	0.269	Journal Article
48	Baser & Geban, 2007	1.781	0.302	Journal Article
49	Baser & Cataloglu, 2005	0.775	0.239	Journal Article
50	Baser & Geban, 2007b	1.267	0.256	Journal Article
51	Pabuccu, 2004	0.772	0.318	Doctoral Dissertation
52	Canpolat,Pınarbaşı,Bayrakçeken & Geban, 2006	1.718	0.252	Journal Article
53	Önder, 2005	2.269	0.229	Doctoral Dissertation
54	Onder, 2017	0.580	0.300	Journal Article
55	Balci, 2006	0.441	0.307	Master Thesis
56	Vatansever, 2005	1.430	0.362	Master Thesis
57	Yenilmez &Tekkaya, 2006	0.737	0.135	Journal Article
58	Pinarbasi, Canpolat, Eken, & Geban, 2006	1.303	0.235	Journal Article
59	Erdemir, 2006	0.406	0.199	Doctoral Dissertation
60	Yilmaz, 2007	0.141	0.276	Master Thesis
61	Al Khawaldeh, 2007	0.935	0.244	Journal Article
62	Yürük, 2007	2.245	0.317	Journal Article
63	Taştan, Dikmenli, & Çardak 2008	0.267	0.280	Journal Article
64	Sevim, 2007	1.696	0.377	Doctoral Dissertation
65	Cibik, Diken, & Darçın, 2008	0.516	0.228	Journal Article
66	Liu, 2008	0.562	0.227	Master Thesis
67	Taştan, Yalcınkaya, & Boz, 2008	1.917	0.309	Journal Article
68	Dilber, 2008	1.705	0.302	Journal Article
69	Li, 2008	0.904	0.326	Journal Article
70	Acar & Tarhan, 2008	3.330	0.407	Journal Article
71	She & Lee, 2008	0.864	0.269	Journal Article
72	Anyanvu, 2008	1.645	0.298	Doctoral Dissertation
73	Demircioglu, 2009	1.030	0.437	Journal Article
74	Atasoy, Akkus, & Kadayifci, 2009	1.760	0.351	Journal Article
75	Özmen, Demircioğlu, & Demircioğlu, 2009	0.484	0.263	Journal Article
76	Berber & Sari, 2009	0.769	0.286	Journal Article
77	Ceylan & Geban, 2009	2.800	0.257	Journal Article
78	Uce, 2009	2.017	0.242	Journal Article
79	Çetin, 2009	1.281	0.266	Doctoral Dissertation
80	Cetin, Kaya, & Geban 2009	2.060	0.286	Journal Article
81	Dilber, Karaman, & Düzgün, 2009	1.588	0.252	Journal Article

82	Liao & She, 2009	0.637	0.245	Journal Article
83	Yuruk, Beeth, & Andersen, 2008	0.751	0.303	Journal Article
84	Bawaneh, Zain, & Saleh, 2010	1.437	0.251	Journal Article
85	Çalik, Kolomuc, & Karagölge, 2010	1.527	0.266	Journal Article
86	Turgut & Gürbüz, 2011	1.390	0.360	Journal Article
87	Pekmez, 2010	1.438	0.314	Journal Article
88	Broughton, Sinatra, & Reynolds 2010	0.369	0.313	Journal Article
89	Udogu & Njelita, 2010	2.521	0.406	Journal Article
90	Akgül, 2010	1.146	0.350	Master Thesis
91	Dilber, 2010	1.183	0.257	Journal Article
92	Aykutlu & Şen,2011	0.724	0.293	Journal Article
93	Barthlow, 2011	0.672	0.115	Doctoral Dissertation
94	Yilmaz, Tekkaya, & Sungur, 2011	0.817	0.287	Journal Article
95	Cetingul, 2006	1.268	0.300	Doctoral Dissertation
96	Özmen, 2011	2.163	0.349	Journal Article
97	Taşdelen, 2011	1.559	0.279	Doctoral Dissertation
98	Karslı & Ayaş, 2013	1.696	0.329	Journal Article
99	Cinici, Sözbilir, & Demir, 2011	0.769	0.383	Journal Article
100	Hirca, Çalık & Seven, 2011	1.016	0.322	Journal Article
101	Karakuyu &Tüysüz, 2011	1.433	0.273	Journal Article
102	Lin, Liu, & Chu, 2011	0.444	0.157	Journal Article
103	Akbas, 2008	2.294	0.270	Journal Article
104	Akbulut, Şahin, & Çepni, 2011	0.970	0.301	Journal Article
105	Nwankwo & Madu, 2014	0.718	0.196	Journal Article
106	Wozniak, 2012	3.601	0.567	Doctoral Dissertation
107	Kaya, 2009	1.286	0.262	Doctoral Dissertation
108	Seker, 2012	1.725	0.328	Doctoral Dissertation
109	Sota, 2012	0.021	0.341	Doctoral Dissertation
110	Çelikten, İpekçioğlu, Ertepınar, & Geban,2012	0.648	0.271	Journal Article
111	Feyzioglu, Ergin, & Kocakülah, 2012	0.762	0.283	Journal Article
112	Kıngır, Geban, & Günel , 2013	0.627	0.184	Journal Article
113	Köseoğlu & Bayır, 2012	1.226	0.347	Journal Article
114	Allen & Coole, 2012	0.94	0.351	Journal Article
115	Chen & She, 2012	0.358	0.164	Journal Article
116	Yaman, 2013	2.010	0.217	Doctoral Dissertation
117	Kasap &Ültay, 2014	1.665	0.322	Journal Article
118	Özkan, 2013	1.195	0.277	Master Thesis
119	Can & Boz, 2016	1.761	0.218	Journal Article
120	Chen, Pan, Sung, & Chang, 2013	0.452	0.339	Journal Article
121	Demirezen & Yağbasan, 2013	1.240	0.244	Journal Article
122	Sendur & Toprak, 2013	0.698	0.257	Journal Article
123	Johnson & Sinatra, 2013	1.507	0.233	Journal Article
124	Wood, Ebenezer, & Boonea, 2013	1.001	0.336	Journal Article

125	Södervik, Erdmann, & Vilppu, 2014	0.294	0.209	Journal Article
126	Tlala, Kibirige, & Osodo, 2014	0.870	0.235	Journal Article
127	Aslan & Demircioglu, 2014	1.372	0.342	Conference Paper
128	Ünlü, 2012	1.762	0.226	Doctoral Dissertation
129	Budiman, Halim, Meerah, & Osman, 2014	1.547	0.235	Journal Article
130	Yin, Tomita, & Shavelson, 2014	0.826	0.288	Journal Article
131	Arslan, Geban, & Sağlam, 2012	0.722	0.136	Journal Article
132	Sarı Ay, 2011	2.169	0.394	Journal Article
133	Loyens, Jones, Mikkers, & Gog, 2015	0.584	0.279	Journal Article
134	Cetin, Ertepınar, & Geban, 2015	0.645	0.226	Journal Article
135	Karamustafaoglu & Naaman, 2015	1.493	0.352	Journal Article
136	Yumusak, Maraş, & Şahin, 2015	2.508	0.470	Journal Article
137	Hacimustafaoglu, 2015	0.613	0.317	Master Thesis
138	Pekel & Hasenekoğlu, 2015	1.955	0.334	Journal Article
139	Çoruhlu & Çepni, 2015	0.993	0.248	Journal Article
140	Södervik, Virtanen, & Erdmann, 2015	-0.033	0.152	Journal Article
141	Yalcınkaya & Boz, 2015	1.240	0.192	Journal Article
142	Loon, Dunlosky, Gog, Merriënboer, & Bruin, 2015	0.545	0.190	Journal Article
143	Diakidoy, Mouskounti, Fella & İonides 2016	0.089	0.240	Journal Article
144	Eymur, 2014	2.844	0.333	Doctoral Dissertation
145	Mason, Baldi, Ronco, Scrimin, Danielson, & Sinatra, 2017	0.677	0.288	Journal Article
146	Özmen & Naseriaza, 2017	1.878	0.214	Journal Article
147	Xinxin, Geelan & Gillies, 2018	0.934	0.194	Journal Article
148	Mason, Zaccoletti, Carretti, Scrimin, & Diakidoy, 2019	0.329	0.216	Journal Article
149	Muisa, Sinatra, Pekrunc, Winne, Trevors, Losenno, & Munzar, 2018	0.941	0.191	Journal Article
150	Adesope, Cavagnetto, Hunsu, Anguiano, & Lloyd, 2017	0.748	0.310	Journal Article
151	Gayeta & Caballes, 2017	-0.135	0.284	Journal Article
152	Sahhyar & Hastini, 2017	0.867	0.263	Journal Article
153	Alkha6waldeh, 2012	0.864	0.241	Journal Article
154	Dilber & Duzgun, 2008	3.013	0.329	Journal Article
155	Sendur, Toprak ,& Pekmez, 2008	1.404	0.319	Journal Article
156	Karakethudaoglu, 2010	0.655	0.327	Master Thesis
157	Demirci & Sarikaya, 2003	1.063	0.273	Conference Paper
158	Gedik, Ertepınar, & Geban, 2001	0.767	0.301	Conference Paper
159	Tokur, Duruk, & Akgün, 2010	0.479	0.225	Journal Article
160	Demircioglu, Aydın, & Demircioğlu, 2013	1.480	0.379	Journal Article
161	Demirel & Anıl, 2018	3.137	0.325	Journal Article
162	Kırık & Boz, 2012	1.375	0.303	Journal Article
163	Seker & Geban, 2014	0.766	0.245	Journal Article
164	Köse, 2004	2.165	0.251	Doctoral Dissertation
165	Yilmaz, 2010	0.668	0.184	Doctoral Dissertation

166	Damli ,2011	2.037	0.416	Master Thesis
167	Uyanik & Dindar, 2016	1.632	0.323	Conference Paper
168	Alkhawaldeh & Olaimat, 2010	1.433	0.266	Journal Article
169	Cobanoglu & Bektas, 2012	1.976	0.352	Conference Paper
170	Diakidoy, Kendeou, & Ioannides, 2003	0.746	0.166	Journal Article
171	Gürses, Doğar, Yalçın, & Canpolat, 2002	0.642	0.256	Conference Paper
172	Alkhawaldeh, 2012	1.296	0.299	Journal Article
173	Durmus, 2009	1.865	0.242	Master Thesis
174	Tokatlı ,2010	1.747	0.311	Master Thesis
175	Özmen, 2007	0.817	0.234	Journal Article
176	Özmen & Demircioğlu, 2003	2.175	0.323	Journal Article
177	Polat, 2007	1.558	0.294	Master Thesis
178	Keleş, 2008	1.484	0.301	Doctoral Dissertation
179	Gürbüz, 2008	1.352	0.307	Master Thesis
180	Lee & She, 2010	1.297	0.287	Journal Article
181	İpek, 2007	0.940	0.281	Master Thesis
182	Zohar & Kravetsky, 2003	0.110	0.277	Conference Paper
183	Trevors, 2011	0.774	0.279	Master Thesis
184	Clement, 1993	1.271	0.323	Journal Article
185	Çakmak, 2016	1.062	0.253	Doctoral Dissertation
186	Harman, 2016	0.716	0.207	Doctoral Dissertation
187	Aydın, 2011	1.253	0.222	Doctoral Dissertation
188	Alemisoğlu, 2014	0.452	0.314	Master Thesis
189	Demirer, 2015	1.800	0.567	Master Thesis
190	Duman, 2015	2.218	0.450	Master Thesis
191	Atılğanlar, 2014	0.401	0.331	Master Thesis
192	Aksu, 2010	1.192	0.397	Doctoral Dissertation
193	Kılıç, 2016	1.541	0.282	Master Thesis
194	Coetzee & Imenda, 2012	0.158	0.174	Journal Article
195	Karamustafaoglu, Ayaş, & Çoştu, 2002	1.048	0.236	Conference Paper
196	Can, Yaşadı, Sönmezer, & Kesercioğlu, 2006	1.349	0.249	Journal Article
197	Tezcan & Salmaz, 2005	0.654	0.279	Journal Article
198	Kör, 2006	0.744	0.264	Master Thesis
199	Aydın, 2007	0.532	0.268	Master Thesis
200	İnal,2003	1.865	0.329	Master Thesis
201	Çaycı, 2007	0.441	0.285	Journal Article
202	Toros, 2015	0.902	0.205	Master Thesis
203	Çelik, 2013	1.410	0.288	Master Thesis
204	Carlsen, 1989	0.480	0.222	Doctoral Dissertation
205	Amponsah & Ochongor,2016	0.891	0.291	Conference Paper
206	Pabuçcu & Geban, 2015	1.292	0.192	Journal Article
207	Çaycı, 2018	0.972	0.200	Journal Article
208	Asana, 2020	0.769	0.285	Master Thesis

209	Hanson & Jele, 2018	0.485	0.238	Journal Article
210	İşcan, 2020	1.291	0.339	Master Thesis
211	Özmen & Naseriazer, 2017	1.896	0.220	Journal Article
212	Perdana, Suma, & Pujani, 2018	0.646	0.216	Conference Paper
213	Çıbık, 2011	0.758	0.229	Doctoral Dissertation
214	Kılıç, 2007	0.453	0.296	Master Thesis
215	Okur,2009	0.614	0.317	Master Thesis
216	Uzun, 2010	2.014	0.351	Doctoral Dissertation
217	Dilber, 2006	1.116	0.219	Doctoral Dissertation
218	Bayar, 2009	0.209	0.271	Master Thesis

D. List of Researchers Providıng Feedback for CCS Type

No	Name of Researcher	Study Name	Country	Feedback
1	Okşan Çelikten	The effect of the conceptual change oriented instruction through cooperative learning on 4th-grade students' understanding of earth and sky concepts	Turkey	Agree
2	Hatice Belge Can	Structuring Cooperative Learning for Motivation and Conceptual Change in the Concepts of Mixtures	Turkey	Agree
3	Ayhan Çinici	Effect of cooperative and individual learning activities on students' understanding diffusion and osmosis	Turkey	Agree
4	Harika Özge Arslan	Learning cycle model to Foster conCeptual undeRstanding in Cell division and Reproduction concepts	Turkey	Agree
5	Gülcan Çetin	Effects of conceptuaş change texts based instruction on ecology, attitudes toward biology and environment	Turkey	Agree
6	Özgecan Taştan	Cooperative learning instruction for conceptual change in the concepts of chemical kinetics	Turkey	Agree
7	Özgecan Taştan	Effectiveness of Conceptual Change Text-oriented Instruction on Students' Understanding of Energy in Chemical Reactions	Turkey	Agree
8	Gökhan Demircioğlu	Comparison of the effects of conceptual change texts implemented after and before instruction on secondary school students' understanding of acid-base concepts	Turkey	Agree
9	Gökhan Demircioğlu	Asitler ve Bazlar Konusundaki Öğrenci Yanlış Anlamalarının Değerlendirilmesinde Kavramsal Değişim Metinlerinin Etkisi	Turkey	Agree
10	Gökhan Demircioğlu	Conceptual change achieved through a new teaching program on acids and bases	Turkey	Agree
11	Ceren Tekkaya	Facilitating Conceptual Change in Students' Understanding of Ecological Concept	Turkey	Agree
12	Ceren Tekkaya	Remediating High School Students' Misconceptions Concerning Diffusion and Osmosis through Concept Mapping and Conceptual Change Text	Turkey	Agree
13	Zeynel Abidin Yılmaz	Kavramsal değişim metinlerinin üniversite öğrencilerinin geometrik optik konusundaki kavram yanılgılarının düzeltilmesi ve fizik dersine karşı tutumlarına etkisinin incelenmesi	Turkey	Agree
14	Ebru Kaya	Reaksiyon Hızı Konusunda Kavramsal Değişime Dayalı Öğretim Metodu ile Kavramsal Değişimin Oluşturulması	Turkey	Agree
15	Muammer Çalık	A Comparison of Different Conceptual Change Pedagogies Employed Within the Topic of Sound Propagation	Turkey	Agree
16	Muammer Çalık	Analogical reasoning for understanding solution rates: students' conceptual change and chemical explanations	Turkey	Agree
17	Muammer Çalık	Effect of Enriched 5Es Model on Grade 7 Students' Conceptual Change Levels: A Case of 'Electric Current' Subject	Turkey	Agree

18	Gamze Arık Dolu	Identification and Elimination of Several Misconceptions of University Level Students Regarding the Misconceptions in Science Course	Turkey	Agree
19	Haluk Özmen	Effect of animation enhanced conceptual change texts on 6th students' understanding of the particle nature and transformation during phase changes	Turkey	Agree
20	Haluk Özmen	The Effectiveness of Conceptual Change Texts in Remediating High School Students' Alternative Conceptions Concerning Chemical Equilibrium	Turkey	Agree
21	Haluk Özmen	The effects of conceptual change texts accompanied with animations on overcoming 11th-grade students' alternative conceptions of chemical bonding	Turkey	Agree
22	Nilüfer Cerit Berber	Kavramsal değişim metinlerinin iş, güç, enerji konusunu anlamaya etkisi	Turkey	Agree
23	Nursen Azizoğlu	Conceptual change oriented instruction and students' misconceptions in gases	Turkey	Agree
24	Eylem Yıldız Feyzioğlu	The Effect of 5E Learning Model Instruction on Seventh Grade Students' Conceptual Understanding of Force and Motion	Turkey	Disagree
25	Aybüke Pabuçcu	Kimyasal bağlarla ilgili kavram yanılgılarının kavramsal degişim metinleri kullanılarak düzeltilmesi	Turkey	Agree
26	Sevgi Kıngır	Using the Science Writing Heuristic Approach to Enhance Student Understanding in Chemical Change and Mixture	Turkey	Agree
27	Gülüzar Eymur	The Collaboration of Cooperative Learning and Conceptual Change: Enhancing the Students' Understanding of Chemical Bonding Concepts	Turkey	Agree
28	Ali Eryılmaz	Effects of Conceptual Assignments and Conceptual Change Discussions on Students' Misconceptions and Achievement Regarding Force and Motion	Turkey	Agree
29	Ali Eryılmaz	Integrating Gender and Group Differences into Bridging Strategy	Turkey	Agree
30	Ali Eryılmaz	Assessing the Impact of Bridging Analogies in Mechanics	Turkey	Agree
31	Yunus Karakuyu	Elektrik Konusunda Kavram Yanılgıları ve Kavramsal Değişim Yaklaşımı	Turkey	Agree
32	Gülten Şendur	The role of conceptual change texts to improve students' understanding of alkene	Turkey	Agree
33	Gülten Şendur	Buharlaşma ve Kaynama Konularındaki Kavram Yanılgılarının Önlenmesinde Analoji Yönteminin Etkisi	Turkey	Agree
34	Ayşe Yenilmez Türk	Enhancing Students' Understanding of Photosynthesis and Respiration in Plant Through Conceptual Change Approach	Turkey	Agree
35	Gonca Kasap	To determine the effect of the activities based on conceptual change approach on students 'conceptual understanding of floating-sinking objects	Turkey	Agree
36	Tacettin Pınarbaşı	The conceptual change approach to teaching chemical equilibrium	Turkey	Agree
37	İsmail Önder	The effect of conceptual change approach on students' understanding of solubility equilibrium concept	Turkey	Agree
38	İsmail Önder	The Effect of Conceptual Change Texts Supplemented Instruction on Students' Achievement in Electrochemistry	Turkey	Agree

39	Murat Demirel	Kavramsal değişim yaklaşımına yönelik çalışma: gazlar konusu	Turkey	Agree
40	Tülay Şenel Çoruhlu	Evaluation of the effects of the 5E model enriched with conceptual change pedagogy on students' conceptual change: "Comet", "Star Drift" and "Meteor"	Turkey	Agree
41	Gökhan Uyanık	The Effect of the Conceptual Change Texts on Removing Misconceptions in Primary 4th Grade Science Cours	Turkey	Agree
42	Ahmet Gürses	Kavramsal değişim yaklaşımının öğrencilerin gazlar konusunu anlamalarına etkisi	Turkey	Agree
43	Refik Dilber	Effect of conceptual change instruction on students' understanding of electricity concepts	Turkey	Agree
44	Refik Dilber	High school students' understanding of projectile motion concepts	Turkey	Agree
45	Refik Dilber	Effectiveness of Analogy on Students' Success and Elimination of Misconceptions	Turkey	Agree
46	Refik Dilber	Teaching of the water waves: Effectiveness of computer simulations on student success and elimination of misconceptions	Turkey	Agree
47	Mutlu Pınar Demirci Güler	Sınıf öğretmeni adaylarının ısı ve sıcaklık konusundaki kavram yanılgıları ve yanılgıların giderilmesinde yapısalcı kuramın etkisi	Turkey	Agree
48	Aygül Aslan	The effect of video-assisted conceptual change texts on 12th grade students' alternative conceptions: The gas concept	Turkey	Agree
49	Barış Çaycı	The Impacts of Conceptual Change on Text-based Concept Teaching on Various Variables	Turkey	Agree
50	Barış Çaycı	Kavram Değiştirme Metinlerinin Kavram Öğrenimi Üzerindeki Etkisinin İncelenmesi	Turkey	Agree
51	Nejla Yürük	Effectiveness of Conceptual Change Text-oriented Instruction on Students' Understanding of Cellular Respiration Concepts	Turkey	Agree
52	Nejla Yürük	The Effect of Supplementing Instruction with Conceptual Change Texts on Students' Conceptions of Electrochemical Cells	Turkey	Agree
53	Nejla Yürük	The effect of conceptual change texts enriched with meta conceptual processes on preservice science teachers' conceptual understanding of heat and temperature	Turkey	Neutral
54	Abuzer Akgün	TGA etkinliklerinin fen bilgisi öğretmen adaylarının çiçekli bitkilerin büyüme ve gelişmesi ile ilgili sahip olduğu kavram yanılgılarının giderilmesine etkisi	Turkey	Agree
55	Alipaşa Ayaş	Farklı Kavramsal Değişim Yöntemleri ile Alternatif Kavramları Gidermek ve Bilimsel Süreç Becerilerini Geliştirmek Mümkün müdür? Elektrokimyasal Piller Örneği	Turkey	Agree
56	Çiğdem Şahin Çakır	Effect of using different teaching methods and techniques embedded within the 5e instructional model on removing students' alternative conceptions: Fluid pressure	Turkey	Agree
57	Nurtaç Canpolat	The conceptual change approach to teaching chemical equilibrium	Turkey	Agree

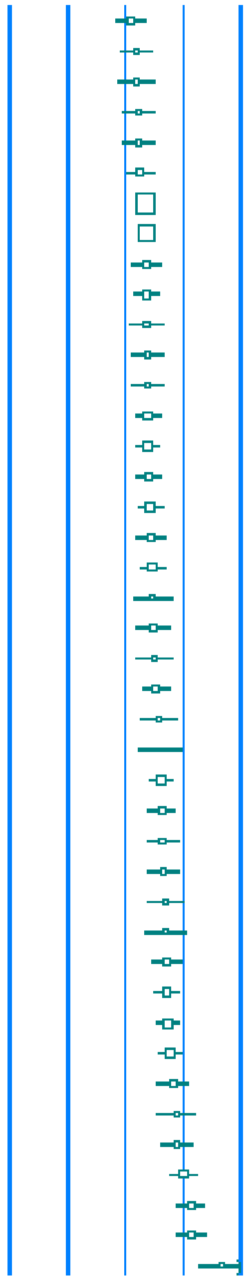
58	Hüseyin Akkuş	The effect of a conceptual change approach on understanding of students' chemical equilibrium concept	Turkey	Agree
59	Eylem Bayır	Sorgulayıcı-araştırmaya dayalı analitik kimya laboratuvarlarının kimya öğretmen adaylarının kavramsal değişimlerine, bilimi ve bilim öğrenme yollarını algılamalarına etkileri	Turkey	Agree
60	Eylem Yalçınkaya Önder	The effect of case-based instruction on 10th grade students' understanding of gas concepts	Turkey	Agree
61	Feyzi Osman Pekel	Dynamising Conceptual Change Approach to Teach Some Genetics Concepts	Turkey	Agree
62	Norrie Gayeta	Measuring Conceptual Change on Stoichiometry Using Mental Models and IllStructured Problems In a Flipped Classroom Environment	Philippines	Agree
63	Patrice Potvin	Experimental Evidence of the Superiority of the Prevalence Model of Conceptual Change Over the Classical Models and Repetition	Canada	Agree
64	Lucia Mason	Textual and graphical refutations: Effects on conceptual change learning	Italy	Neutral
65	Michelene Chi	Repairing Student Misconceptions Using Ontology Training: A Study with Junior and Senior Undergraduate Engineering Students	USA	Agree
66	Michelene Chi	Helping Students Understand Challenging Topics in Science through Ontology Training	USA	Agree
67	Mariëtte van Loon	Refutations in science texts lead to hypercorrection of misconceptions held with high confidence	Holland	Agree
68	Mary G.Nwankwo	Effect of Analogy Teaching Approach on Students' Conceptual Change in Physics	Nigeria	Agree
69	Ruth Stavy	Using Analogy to Overcome Misconceptions About Conservation of Matter	Israel	Agree
70	Trevors Gregory	Learner, Text, and Context Factors on Conceptual Change in Biology	Canada	Agree
71	Christine Howe	Peer Collaboration and Conceptual Growth in Physics: Task Influences on Children's Understanding of Heating and Cooling	Scotland	Agree
72	Ilona Södervik	Promoting the Understanding of Photosynthesis Among Elementary School Students, Teachers Through Text Design	Finland	Agree
73	Yue Yin	Using Formal Embedded Formative Assessments Aligned with a ShortTerm Learning Progression to Promote Conceptual Change and Achievement in Science	USA	Agree
74	Marcus lee johnson	Use of task value instructional induction for facilitating engagement and conceptual change	USA	Agree
75	Ali Bawaneh	Radıcal conceptual change through teachıng method based on constructıvısm theory for eıght grade jordanıan student	Jordan	Agree
76	Hsiao Ching She	SCCR digital learning system for scientific conceptual change and scientific reasoning	Taiwan	Agree
77	Kuo-En Chang	Correcting Misconceptions on Electronics: Effects of a simulation-based learning environment backed by a conceptual change model	Taiwan	Agree
78	Mark Windschitl	Using Computer Simulations to Enhance Conceptual Change: The Roles of Constructivist Instruction and Student Epistemological Beliefs	USA	Agree

79	Tomas Andree	Using Computer Simulations to Enhance Conceptual Change: The Roles of Constructivist Instruction and Student Epistemological Beliefs	USA	Agree
80	Murray S. Jensen	A Computer-Assisted Instruction Unit on Diffusion and Osmosis with a Conceptual Change Design	USA	Agree
81	Zainol Badli Budiman	The effects of cognitive conflict management on cognitive development and science achievement	Malasia	Agree
82	Mei-Hung Chiu	Promoting Fourth Graders' Conceptual Change of Their Understanding of Electric Current via Multiple Analogies	China	Disagree
83	C.B. Njelita	Effect of Constructivist-Based Instructional Model on Students' Conceptual Change and Retention on Some Difficult Concepts in Chemistry	Nigeria	Agree

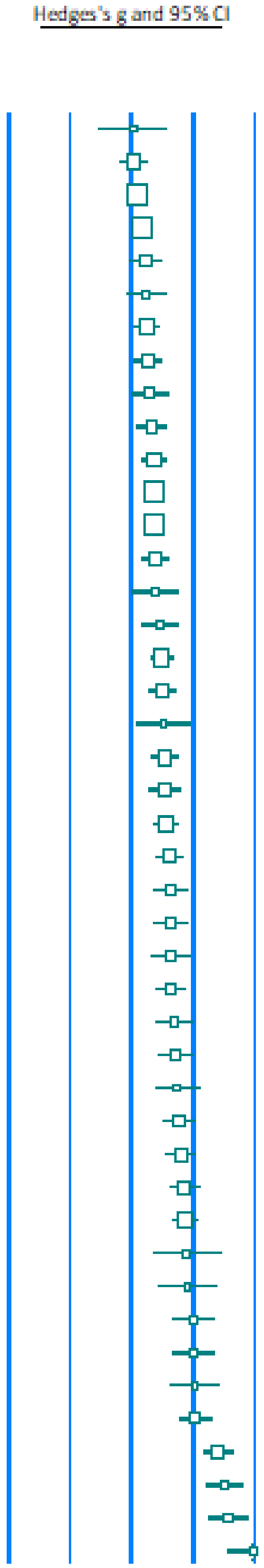
E. Forest Plots for Conceptual Change Strategy

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Statistics for each study						
Bayar 2009	0,209	0,271	0,073	-0,322	0,740	0,771	0,441
Broughton, 2010	0,369	0,313	0,098	-0,244	0,982	1,179	0,238
Atilganlar, 2014	0,401	0,331	0,110	-0,248	1,050	1,211	0,226
Balci, 2006	0,441	0,307	0,094	-0,161	1,043	1,436	0,151
Alemisoglu, 2014	0,452	0,314	0,099	-0,163	1,067	1,439	0,150
Aydýn, 2007	0,532	0,268	0,072	0,007	1,057	1,985	0,047
Barthlow, 2011	0,672	0,115	0,013	0,447	0,897	5,843	0,000
Arslan et al., 2012	0,722	0,136	0,018	0,455	0,989	5,309	0,000
Aykutlu al.,2009	0,724	0,293	0,086	0,150	1,298	2,471	0,013
Cakir, 2002	0,734	0,224	0,050	0,295	1,173	3,277	0,001
Adesope al., 2017	0,748	0,310	0,096	0,140	1,356	2,413	0,016
Berber et al,2007	0,769	0,286	0,082	0,208	1,330	2,689	0,007
Asana, 2020	0,769	0,285	0,081	0,210	1,328	2,698	0,007
Baser et al, 2005	0,775	0,239	0,057	0,307	1,243	3,243	0,001
Baser, 2006	0,795	0,219	0,048	0,366	1,224	3,630	0,000
Alparslan, 2003	0,807	0,250	0,063	0,317	1,297	3,228	0,001
Alkhawaldeh, 2011	0,864	0,241	0,058	0,392	1,336	3,585	0,000
Amponsah al,2015	0,891	0,291	0,085	0,321	1,461	3,062	0,002
Alkhawaldeh,2005	0,935	0,244	0,060	0,457	1,413	3,832	0,000
Allen al,2012	0,940	0,351	0,123	0,252	1,628	2,678	0,007
Akbulut al, 2011	0,970	0,301	0,091	0,380	1,560	3,223	0,001
Bilgin et al., 2001	1,020	0,339	0,115	0,356	1,684	3,009	0,003
Cakmak, 2016	1,062	0,253	0,064	0,566	1,558	4,198	0,000
Akgul, 2009	1,146	0,350	0,123	0,460	1,832	3,274	0,001
Aksu, 2010	1,192	0,397	0,158	0,414	1,970	3,003	0,003
Aydin, 2011	1,253	0,222	0,049	0,818	1,688	5,644	0,000
Baser et al.,2007b	1,267	0,256	0,066	0,765	1,769	4,949	0,000
Alkhawaldeh, 2012	1,296	0,299	0,089	0,710	1,882	4,334	0,000
Bozkoyun, 2004	1,310	0,291	0,085	0,740	1,880	4,502	0,000
Aslan et al.,2013	1,372	0,342	0,117	0,702	2,042	4,012	0,000
Ayhan, 2004	1,384	0,380	0,144	0,639	2,129	3,642	0,000
Alkhawaldeh al,2009	1,433	0,266	0,071	0,912	1,954	5,387	0,000
Bawaneh, 2009	1,437	0,251	0,063	0,945	1,929	5,725	0,000
Azizoðlu, 2004	1,462	0,224	0,050	1,023	1,901	6,527	0,000
Budiman al,2012	1,547	0,235	0,055	1,086	2,008	6,583	0,000
Anyanvu, 2008	1,645	0,298	0,089	1,061	2,229	5,520	0,000
Atasoy et al., 2009	1,760	0,351	0,123	1,072	2,448	5,014	0,000
Baser et al.,2007	1,781	0,302	0,091	1,189	2,373	5,897	0,000
Baser, 2006 b	2,012	0,269	0,072	1,485	2,539	7,480	0,000
Bilgin&Geban, 2002	2,250	0,273	0,075	1,715	2,785	8,242	0,000
Akbas, 2008	2,294	0,270	0,073	1,765	2,823	8,496	0,000
Acar et al.,2007	3,330	0,407	0,166	2,532	4,128	8,182	0,000

Hedges's g and 95% CI

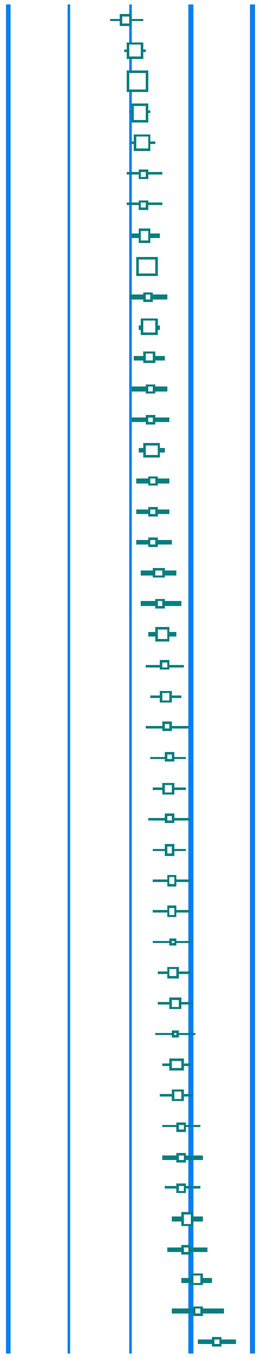


Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Charles, 2003	0,036	0,568	0,323	-1,077	1,149	0,063	0,949
Diakidoy, 2015	0,089	0,240	0,058	-0,381	0,559	0,371	0,711
Coetzee, 2012	0,158	0,174	0,030	-0,183	0,499	0,908	0,364
Chen al, 2012	0,358	0,164	0,027	0,037	0,679	2,183	0,029
Çaycý, 2007	0,441	0,285	0,081	-0,118	1,000	1,547	0,122
Chen et al., 2012	0,452	0,339	0,115	-0,212	1,116	1,333	0,182
Carlsen, 1989	0,480	0,222	0,049	0,045	0,915	2,162	0,031
Cibik et al.,2008	0,516	0,228	0,052	0,069	0,963	2,263	0,024
Celebi, 2004	0,599	0,307	0,094	-0,003	1,201	1,951	0,051
Celikten al. ,2012	0,648	0,271	0,073	0,117	1,179	2,391	0,017
Cetin, 2002	0,700	0,227	0,052	0,255	1,145	3,084	0,002
Diakidoy et al., 2003	0,746	0,166	0,028	0,421	1,071	4,494	0,000
Chang, 1995	0,755	0,164	0,027	0,434	1,076	4,604	0,000
Çibik, 2011	0,758	0,229	0,053	0,308	1,208	3,304	0,001
Cinici et al., 2011	0,769	0,383	0,147	0,018	1,520	2,008	0,045
Cheung,2008	0,904	0,326	0,106	0,265	1,543	2,773	0,006
Çaycý, 2018	0,972	0,200	0,040	0,580	1,364	4,860	0,000
Coruhlu et al.,2015	0,993	0,248	0,062	0,507	1,479	4,004	0,000
Demircioglu, 2009	1,030	0,437	0,191	0,173	1,887	2,357	0,018
Demircioglu al,2004	1,060	0,226	0,051	0,617	1,503	4,690	0,000
Demirci et al,2003	1,063	0,273	0,075	0,528	1,598	3,894	0,000
Dilber 2006	1,116	0,219	0,048	0,686	1,546	5,093	0,000
Demirezen al,2013	1,240	0,244	0,060	0,762	1,718	5,082	0,000
Cetingul al.,2011	1,267	0,306	0,094	0,667	1,867	4,141	0,000
Cetingul, 2006	1,268	0,300	0,090	0,680	1,856	4,227	0,000
Clement, 1993	1,271	0,323	0,104	0,638	1,904	3,935	0,000
Cetin, 2008	1,281	0,266	0,071	0,760	1,802	4,816	0,000
Diakidoy al.,2001	1,383	0,300	0,090	0,795	1,971	4,610	0,000
Çelik, 2013	1,410	0,288	0,083	0,846	1,974	4,896	0,000
Demircioglu al,2015	1,480	0,379	0,144	0,737	2,223	3,905	0,000
Calik etal., 2010	1,527	0,266	0,071	1,006	2,048	5,741	0,000
Dilber et al., 2008	1,588	0,252	0,064	1,094	2,082	6,302	0,000
Canpolat al.,2006	1,718	0,252	0,064	1,224	2,212	6,817	0,000
Can and Boz,2012	1,761	0,218	0,048	1,334	2,188	8,078	0,000
Demirer, 2015	1,800	0,567	0,321	0,689	2,911	3,175	0,002
Chiu and Lin, 2003	1,808	0,493	0,243	0,842	2,774	3,667	0,000
Cobanoglu et al,2012	1,976	0,352	0,124	1,286	2,666	5,614	0,000
Cetingul al.,2005	2,016	0,354	0,125	1,322	2,710	5,695	0,000
Damli ,2010	2,037	0,416	0,173	1,222	2,852	4,897	0,000
Cetin et al., 2009	2,060	0,286	0,082	1,499	2,621	7,203	0,000
Ceylan et al.,2009	2,800	0,257	0,066	2,296	3,304	10,895	0,000
Dilber et al,2008	3,013	0,329	0,108	2,368	3,658	9,158	0,000
Demirel etal,,2014	3,137	0,325	0,106	2,500	3,774	9,652	0,000
Ceylan, 2004	3,939	0,437	0,191	3,082	4,796	9,014	0,000



Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Gayeta et al,2017	-0,135	0,284	0,081	-0,692	0,422	-0,475	0,635
Eryilmaz, 2002	0,169	0,174	0,030	-0,172	0,510	0,971	0,331
Erdmann, 2000	0,233	0,142	0,020	-0,045	0,511	1,641	0,101
Gürkan, 2021	0,324	0,168	0,028	-0,005	0,653	1,929	0,054
Erdemir, 2005	0,406	0,199	0,040	0,016	0,796	2,040	0,041
Kiliç 2007	0,453	0,296	0,087	-0,126	1,032	1,533	0,125
Gokhale, 1996	0,459	0,294	0,086	-0,117	1,035	1,561	0,118
Hanson al,2018	0,485	0,238	0,057	0,019	0,951	2,038	0,042
Jensen al,1994	0,546	0,157	0,025	0,238	0,854	3,478	0,001
Hacimustafaoglu,2015	0,613	0,317	0,100	-0,008	1,234	1,934	0,053
Kingir et al, 2012	0,627	0,184	0,034	0,266	0,988	3,408	0,001
Gürses et al., 2002	0,642	0,256	0,066	0,140	1,144	2,508	0,012
Kucuk al, 2015	0,646	0,298	0,089	0,062	1,230	2,168	0,030
Karakethudaoglu,2008	0,655	0,327	0,107	0,014	1,296	2,003	0,045
Harman,2016	0,716	0,207	0,043	0,310	1,122	3,459	0,001
Kör, 2006	0,744	0,264	0,070	0,227	1,261	2,818	0,005
Feyzioglu al, 2008	0,762	0,283	0,080	0,207	1,317	2,693	0,007
Gedik et al., 2001	0,767	0,301	0,091	0,177	1,357	2,548	0,011
Ýpek, 2007	0,940	0,281	0,079	0,389	1,491	3,345	0,001
Hirca et al., 2009	1,016	0,322	0,104	0,385	1,647	3,155	0,002
Karamustafaoglu,2002	1,048	0,236	0,056	0,585	1,511	4,441	0,000
Gulcicek, 2003	1,144	0,314	0,099	0,529	1,759	3,643	0,000
Dilber, 2010	1,183	0,257	0,066	0,679	1,687	4,603	0,000
Koseoglu al.,2012	1,226	0,347	0,120	0,546	1,906	3,533	0,000
Launey, 1995	1,278	0,301	0,091	0,688	1,868	4,246	0,000
Kaya, 2009	1,286	0,262	0,069	0,772	1,800	4,908	0,000
Ýþcan, 2020	1,291	0,339	0,115	0,627	1,955	3,808	0,000
Lee and She, 2009	1,297	0,287	0,082	0,734	1,860	4,519	0,000
Gurbuz, 2008	1,352	0,307	0,094	0,750	1,954	4,404	0,000
Kýrýk and Boz,2009	1,375	0,303	0,092	0,781	1,969	4,538	0,000
Gunay, 2005	1,411	0,329	0,108	0,766	2,056	4,289	0,000
Karakuyu al,2011	1,433	0,273	0,075	0,898	1,968	5,249	0,000
Keles, 2009	1,484	0,301	0,091	0,894	2,074	4,930	0,000
Karamustafaoglu,2015	1,493	0,352	0,124	0,803	2,183	4,241	0,000
johnsonet al.,2013	1,507	0,233	0,054	1,050	1,964	6,468	0,000
Kiliç, 2016	1,541	0,282	0,080	0,988	2,094	5,465	0,000
Kasap al,2013	1,665	0,322	0,104	1,034	2,296	5,171	0,000
Karsli et al.,2013	1,696	0,329	0,108	1,051	2,341	5,155	0,000
Dilber, 2008	1,705	0,302	0,091	1,113	2,297	5,646	0,000
Durmus, 2009	1,865	0,242	0,059	1,391	2,339	7,707	0,000
Ýnal,2003	1,865	0,329	0,108	1,220	2,510	5,669	0,000
Köse, 2004	2,165	0,251	0,063	1,673	2,657	8,625	0,000
Duman,2015	2,218	0,450	0,203	1,336	3,100	4,929	0,000
Eymur,2014	2,844	0,333	0,111	2,191	3,497	8,541	0,000

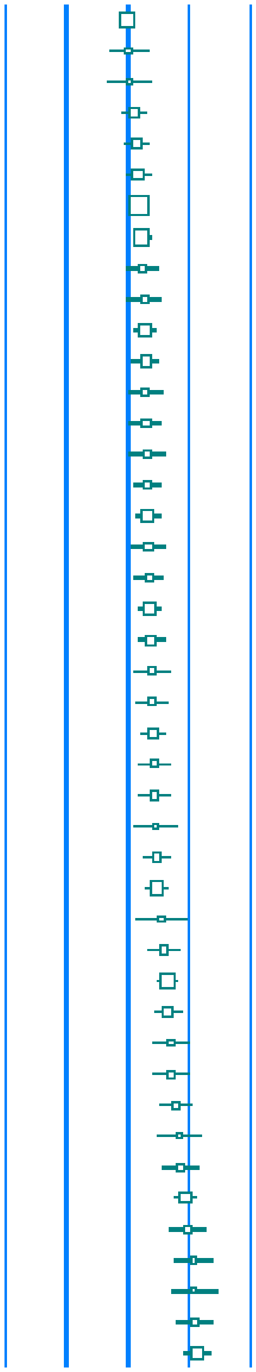
Statistics for each study



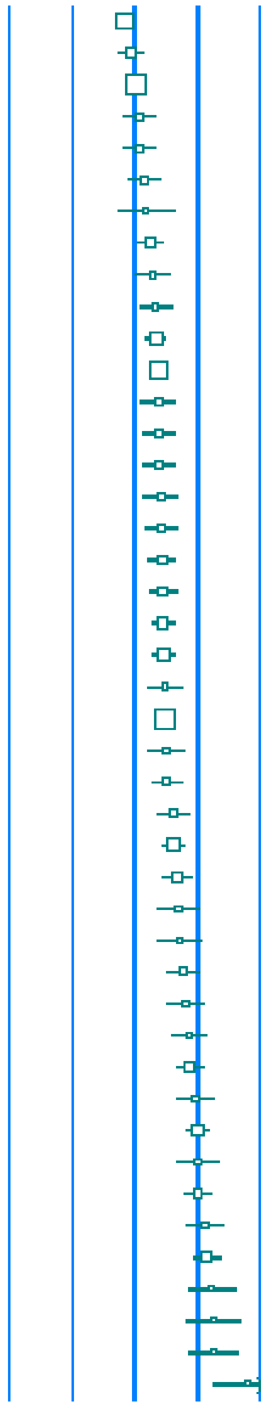
Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Sodervik et al., 2015	-0,033	0,152	0,023	-0,331	0,265	-0,217	0,828
Sota, 2012	0,021	0,341	0,116	-0,647	0,689	0,062	0,951
Seyedmonir, 2000	0,030	0,371	0,138	-0,697	0,757	0,081	0,936
Saigo, 1999	0,190	0,214	0,046	-0,229	0,609	0,888	0,375
Sodervik al., 2013	0,294	0,209	0,044	-0,116	0,704	1,407	0,160
Mason al., 2019	0,329	0,216	0,047	-0,094	0,752	1,523	0,128
Madu and Orji, 2015	0,368	0,129	0,017	0,115	0,621	2,853	0,004
Lin et al., 2011	0,444	0,157	0,025	0,136	0,752	2,828	0,005
Ozmen et al., 2008	0,484	0,263	0,069	-0,031	0,999	1,840	0,066
Savinainen, 2003	0,527	0,298	0,089	-0,057	1,111	1,768	0,077
Loon et al., 2015	0,545	0,190	0,036	0,173	0,917	2,868	0,004
Liu, 2008	0,562	0,227	0,052	0,117	1,007	2,476	0,013
Onder, 2017	0,580	0,300	0,090	-0,008	1,168	1,933	0,053
Loyens et al., 2014	0,584	0,279	0,078	0,037	1,131	2,093	0,036
Okur, 2009	0,614	0,317	0,101	-0,008	1,236	1,934	0,053
Liao al,2009	0,637	0,245	0,060	0,157	1,117	2,600	0,009
Perdana al,2018	0,646	0,216	0,047	0,223	1,069	2,991	0,003
Mason et al, 2017	0,677	0,288	0,083	0,113	1,241	2,351	0,019
Sendur et al.,2013	0,698	0,257	0,066	0,194	1,202	2,716	0,007
Nwankwo, et al,2014	0,718	0,196	0,039	0,333	1,103	3,658	0,000
Seker et al,2005	0,766	0,245	0,060	0,286	1,246	3,127	0,002
Pabuccu, 2004	0,772	0,318	0,101	0,149	1,395	2,428	0,015
Ozkan, 2004	0,780	0,269	0,072	0,253	1,307	2,900	0,004
Ozmen, 2007	0,817	0,234	0,055	0,358	1,276	3,491	0,000
She and Lee,2008	0,864	0,269	0,072	0,337	1,391	3,212	0,001
Sahhyar al,2017	0,867	0,263	0,069	0,352	1,382	3,297	0,001
Slotta al,2006	0,899	0,362	0,131	0,189	1,609	2,483	0,013
Sanger al.,2000	0,934	0,246	0,061	0,452	1,416	3,797	0,000
Muisa et al., 2018	0,941	0,191	0,036	0,567	1,315	4,927	0,000
Niaz al,2003	1,090	0,449	0,202	0,210	1,970	2,428	0,015
Ozkan, 2013	1,195	0,277	0,077	0,652	1,738	4,314	0,000
Pabuçcu al,2015	1,292	0,192	0,037	0,916	1,668	6,729	0,000
Pinarbasi, 2006	1,303	0,235	0,055	0,842	1,764	5,545	0,000
Sendur, et al., 2008	1,404	0,319	0,102	0,779	2,029	4,401	0,000
Pekmez, 2008	1,438	0,314	0,099	0,823	2,053	4,580	0,000
Polat, 2007	1,558	0,294	0,086	0,982	2,134	5,299	0,000
Sevim, 2007	1,696	0,377	0,142	0,957	2,435	4,499	0,000
Seker, 2011	1,725	0,328	0,108	1,082	2,368	5,259	0,000
Ozmen et al,2016	1,878	0,214	0,046	1,459	2,297	8,776	0,000
Pekel and Kilis, 2015	1,955	0,334	0,112	1,300	2,610	5,853	0,000
Ozmen, 2010	2,163	0,349	0,122	1,479	2,847	6,198	0,000
Sarý Ay, 2011	2,169	0,394	0,155	1,397	2,941	5,505	0,000
Ozmen et al,2003	2,175	0,323	0,104	1,542	2,808	6,734	0,000
Onder, 2005	2,269	0,229	0,052	1,820	2,718	9,908	0,000

Statistics for each study

Hedges's g and 95% CI

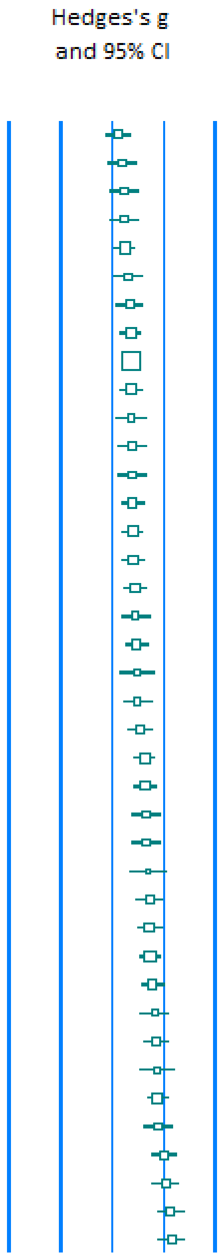


Study name	Statistics for each study							Hedges's g and 95% CI
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	
Tsai, 2003	-0,317	0,145	0,021	-0,602	-0,032	-2,180	0,029	
Zohar al,2005	-0,115	0,217	0,047	-0,540	0,310	-0,530	0,596	
Windschitl al,1995	0,042	0,132	0,017	-0,217	0,301	0,318	0,750	
Zohar et al,2003	0,115	0,277	0,077	-0,428	0,658	0,415	0,678	
Yilmaz, 2006	0,141	0,276	0,076	-0,400	0,682	0,511	0,609	
Tastan, et al 2004	0,267	0,280	0,078	-0,282	0,816	0,954	0,340	
Yang al, 2012	0,373	0,458	0,210	-0,525	1,271	0,814	0,415	
Tokur et al., 2010	0,479	0,225	0,051	0,038	0,920	2,129	0,033	
Tekkaya, 2001	0,586	0,304	0,092	-0,010	1,182	1,928	0,054	
Tezcan et al,2005	0,654	0,279	0,078	0,107	1,201	2,344	0,019	
Yilmaz, 2007	0,668	0,184	0,034	0,307	1,029	3,630	0,000	
Yenilmez al,2006	0,737	0,135	0,018	0,472	1,002	5,459	0,000	
Yuruk et al., 2008	0,751	0,303	0,092	0,157	1,345	2,479	0,013	
Woloshyn al,1992	0,756	0,264	0,070	0,239	1,273	2,864	0,004	
Trevors, 2011	0,774	0,279	0,078	0,227	1,321	2,774	0,006	
Yilmaz et al. ,2011	0,817	0,287	0,082	0,254	1,380	2,847	0,004	
Yin et al., 2014	0,826	0,288	0,083	0,262	1,390	2,868	0,004	
Tlala et al., 2014	0,870	0,235	0,055	0,409	1,331	3,702	0,000	
Yavuz, 2004	0,885	0,240	0,058	0,415	1,355	3,688	0,000	
Toros, 2015	0,902	0,205	0,042	0,500	1,304	4,400	0,000	
Xinxin et al., 2018	0,934	0,194	0,038	0,554	1,314	4,814	0,000	
Sungur al.,2001	0,953	0,298	0,089	0,369	1,537	3,198	0,001	
Yilmaz al,2006b	0,965	0,121	0,015	0,728	1,201	7,997	0,000	
Wood et al., 2011	1,001	0,336	0,113	0,342	1,660	2,979	0,003	
Uzuntiyaki al,2002	1,013	0,263	0,069	0,498	1,528	3,852	0,000	
Stavy, 1991	1,214	0,274	0,075	0,677	1,751	4,431	0,000	
Yalcýnkayaa al,2014	1,240	0,192	0,037	0,864	1,616	6,458	0,000	
Taþkýn, 2006	1,349	0,249	0,062	0,861	1,837	5,418	0,000	
Turgut al., 2009	1,390	0,360	0,130	0,684	2,096	3,861	0,000	
Vatansever, 2005	1,430	0,362	0,131	0,720	2,140	3,950	0,000	
Tasdelen, 2009	1,559	0,279	0,078	1,012	2,106	5,588	0,000	
Uyanik et al,2013	1,632	0,323	0,104	0,999	2,265	5,053	0,000	
Tokatli ,2009	1,747	0,311	0,097	1,137	2,357	5,617	0,000	
Unlu, 2011	1,762	0,226	0,051	1,319	2,205	7,796	0,000	
Tastan et al,2008	1,917	0,309	0,095	1,311	2,523	6,204	0,000	
Yaman, 2010	2,010	0,217	0,047	1,585	2,435	9,263	0,000	
Uzun, 2010	2,014	0,351	0,123	1,326	2,702	5,738	0,000	
Uce, 2007	2,017	0,242	0,059	1,543	2,491	8,335	0,000	
Yuruk, 2009	2,245	0,317	0,100	1,624	2,866	7,082	0,000	
Yilmaz al,2006	2,297	0,236	0,056	1,834	2,760	9,733	0,000	
Uzuntiyaki, 2003	2,457	0,405	0,164	1,663	3,251	6,067	0,000	
Yumusak, 2011	2,508	0,470	0,221	1,587	3,429	5,336	0,000	
Udogu et al.,2010	2,521	0,406	0,165	1,725	3,317	6,209	0,000	
Wozniak, 2012	3,601	0,567	0,321	2,490	4,712	6,351	0,000	

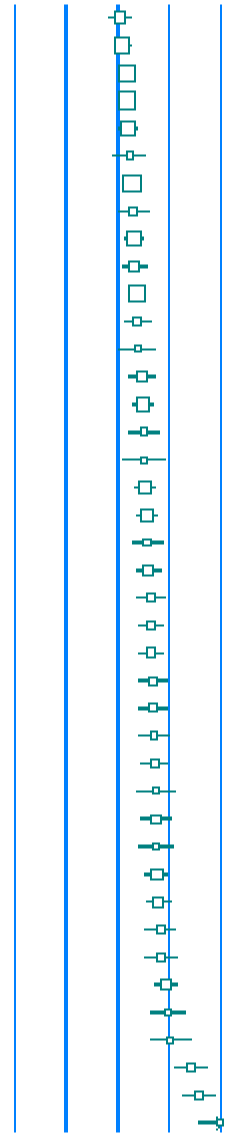


F. Forest Plots for Cognitive Conflict Studies

Study name	Type of CCM	Statistics for each study	
		Hedges's g	Standard error
Bayar 2009	Conflict	0,209	0,271
Broughton, 2010	Conflict	0,369	0,313
Balci, 2006	Conflict	0,441	0,307
Alemisoglu, 2014	Conflict	0,452	0,314
Carlsen, 1989	Conflict	0,480	0,222
Celebi, 2004	Conflict	0,599	0,307
Celikten al. ,2012	Conflict	0,648	0,271
Cetin, 2002	Conflict	0,700	0,227
Arslan et al., 2012	Conflict	0,722	0,136
Cakir, 2002	Conflict	0,734	0,224
Adesope al., 2017	Conflict	0,748	0,310
Berber et al,2007	Conflict	0,769	0,286
Asana, 2020	Conflict	0,769	0,285
Baser et al, 2005	Conflict	0,775	0,239
Baser, 2006	Conflict	0,795	0,219
Alparslan, 2003	Conflict	0,807	0,250
Alkhawaldeh, 2011	Conflict	0,864	0,241
Amponsah al,2015	Conflict	0,891	0,291
Alkhawaldeh,2005	Conflict	0,935	0,244
Allen al,2012	Conflict	0,940	0,351
Akbulut al, 2011	Conflict	0,970	0,301
Cakmak, 2016	Conflict	1,062	0,253
Aydin, 2011	Conflict	1,253	0,222
Baser et al.,2007b	Conflict	1,267	0,256
Alkhawaldeh, 2012	Conflict	1,296	0,299
Bozkoyun, 2004	Conflict	1,310	0,291
Ayhan, 2004	Conflict	1,384	0,380
Alkhawaldeh al,2009	Conflict	1,433	0,266
Bawaneh, 2009	Conflict	1,437	0,251
Azizoğlu, 2004	Conflict	1,462	0,224
Budiman al,2012	Conflict	1,547	0,235
Anyanvu, 2008	Conflict	1,645	0,298
Canpolat al.,2006	Conflict	1,718	0,252
Atasoy et al., 2009	Conflict	1,760	0,351
Can and Boz,2012	Conflict	1,761	0,218
Baser et al.,2007	Conflict	1,781	0,302
Baser, 2006 b	Conflict	2,012	0,269
Cetin et al., 2009	Conflict	2,060	0,286
Bilgin&Geban, 2002	Conflict	2,250	0,273
Akbas, 2008	Conflict	2,294	0,270

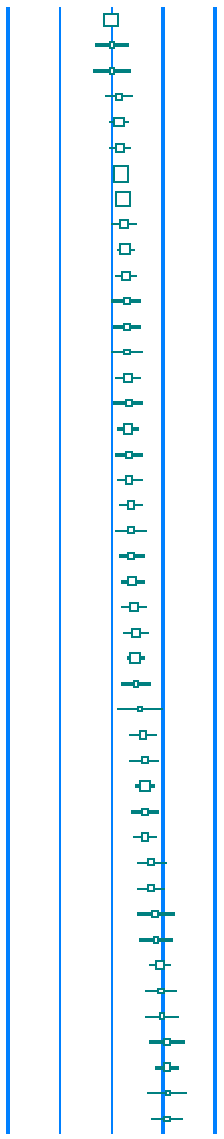


Study name	Type of CCM	Statistics for each study		Hedges's g and 95% CI
		Hedges's g	**Standard error**	
Diakidoy, 2015	Conflict	0,089	0,240	
Eryilmaz, 2002	Conflict	0,169	0,174	
Gürkan, 2021	Conflict	0,324	0,168	
Chen al, 2012	Conflict	0,358	0,164	
Erdemir, 2005	Conflict	0,406	0,199	
Chen et al., 2012	Conflict	0,452	0,339	
Jensen al,1994	Conflict	0,546	0,157	
Hacimustafaoglu,2015	Conflict	0,613	0,317	
Kingir et al, 2012	Conflict	0,627	0,184	
Gürses et al., 2002	Conflict	0,642	0,256	
Diakidoy et al., 2003	Conflict	0,746	0,166	
Feyzioglu al, 2008	Conflict	0,762	0,283	
Cinici et al., 2011	Conflict	0,769	0,383	
Ýpek, 2007	Conflict	0,940	0,281	
Çaycý, 2018	Conflict	0,972	0,200	
Hirca et al., 2009	Conflict	1,016	0,322	
Demircioglu, 2009	Conflict	1,030	0,437	
Demircioglu al,2004	Conflict	1,060	0,226	
Dilber 2006	Conflict	1,116	0,219	
Gulcicek, 2003	Conflict	1,144	0,314	
Dilber, 2010	Conflict	1,183	0,257	
Cetingul, 2006	Conflict	1,268	0,300	
Cetin, 2008	Conflict	1,281	0,266	
Kaya, 2009	Conflict	1,286	0,262	
Gurbuz, 2008	Conflict	1,352	0,307	
Kýrýk and Boz,2009	Conflict	1,375	0,303	
Gunay, 2005	Conflict	1,411	0,329	
Karakuyu al,2011	Conflict	1,433	0,273	
Demircioglu al,2015	Conflict	1,480	0,379	
Keles, 2009	Conflict	1,484	0,301	
Karamustafaoglu,2015	Conflict	1,493	0,352	
johnsonet al.,2013	Conflict	1,507	0,233	
Dilber et al., 2008	Conflict	1,588	0,252	
Kasap al,2013	Conflict	1,665	0,322	
Karsli et al.,2013	Conflict	1,696	0,329	
Durmus, 2009	Conflict	1,865	0,242	
Cobanoglu et al,2012	Conflict	1,976	0,352	
Damli ,2010	Conflict	2,037	0,416	
Eymur,2014	Conflict	2,844	0,333	
Demirel etal,,2014	Conflict	3,137	0,325	
Ceylan, 2004	Conflict	3,939	0,437	

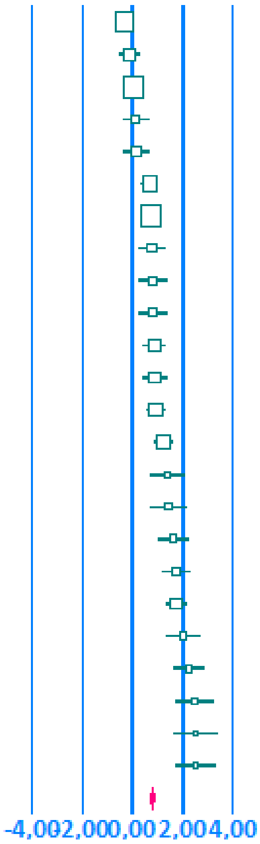


Study name	Type of CCM	Statistics for each study	
		Hedges's g	Standard error
Sodervik et al., 2015	Conflict	-0,033	0,152
Sota, 2012	Conflict	0,021	0,341
Seyedmonir, 2000	Conflict	0,030	0,371
Tastan, et al 2004	Conflict	0,267	0,280
Sodervik al., 2013	Conflict	0,294	0,209
Mason al., 2019	Conflict	0,329	0,216
Madu and Orji, 2015	Conflict	0,368	0,129
Lin et al., 2011	Conflict	0,444	0,157
Ozmen et al., 2008	Conflict	0,484	0,263
Loon et al., 2015	Conflict	0,545	0,190
Liu, 2008	Conflict	0,562	0,227
Onder, 2017	Conflict	0,580	0,300
Loyens et al., 2014	Conflict	0,584	0,279
Tekkaya, 2001	Conflict	0,586	0,304
Liao al,2009	Conflict	0,637	0,245
Kucuk al, 2015	Conflict	0,646	0,298
Perdana al,2018	Conflict	0,646	0,216
Mason et al, 2017	Conflict	0,677	0,288
Sendur et al.,2013	Conflict	0,698	0,257
Seker et al,2005	Conflict	0,766	0,245
Pabuccu, 2004	Conflict	0,772	0,318
Ozkan, 2004	Conflict	0,780	0,269
Ozmen, 2007	Conflict	0,817	0,234
She and Lee,2008	Conflict	0,864	0,269
Sanger al.,2000	Conflict	0,934	0,246
Muisa et al., 2018	Conflict	0,941	0,191
Sungur al.,2001	Conflict	0,953	0,298
Niaz al,2003	Conflict	1,090	0,449
Ozkan, 2013	Conflict	1,195	0,277
Launey, 1995	Conflict	1,278	0,301
Pabuçcu al,2015	Conflict	1,292	0,192
Lee and She, 2009	Conflict	1,297	0,287
Pinarbasi, 2006	Conflict	1,303	0,235
Polat, 2007	Conflict	1,558	0,294
Tasdelen, 2009	Conflict	1,559	0,279
Sevim, 2007	Conflict	1,696	0,377
Seker, 2011	Conflict	1,725	0,328
Ozmen et al,2016	Conflict	1,878	0,214
Tastan et al,2008	Conflict	1,917	0,309
Pekel and Kilis, 2015	Conflict	1,955	0,334
Ozmen, 2010	Conflict	2,163	0,349
Köse, 2004	Conflict	2,165	0,251
Sarý Ay, 2011	Conflict	2,169	0,394
Ozmen et al,2003	Conflict	2,175	0,323

Hedges's g and 95% CI



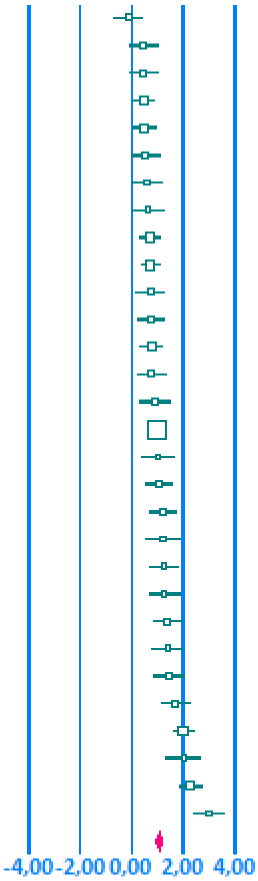
Study name	Type of CCM	Statistics for each study		Hedges's g and 95% CI
		Hedges's g	Standard error	
Tsai, 2003	Conflict	-0,317	0,145	
Zohar al,2005	Conflict	-0,115	0,217	
Windschitl al,1995	Conflict	0,042	0,132	
Zohar et al,2003	Conflict	0,115	0,277	
Yilmaz, 2006	Conflict	0,141	0,276	
Yilmaz, 2007	Conflict	0,668	0,184	
Yenilmez al,2006	Conflict	0,737	0,135	
Trevors, 2011	Conflict	0,774	0,279	
Yilmaz et al. ,2011	Conflict	0,817	0,287	
Yin et al., 2014	Conflict	0,826	0,288	
Tlala et al., 2014	Conflict	0,870	0,235	
Yavuz, 2004	Conflict	0,885	0,240	
Xinxin et al., 2018	Conflict	0,934	0,194	
Yalcýnkayaa al,2014	Conflict	1,240	0,192	
Turgut al., 2009	Conflict	1,390	0,360	
Vatansever, 2005	Conflict	1,430	0,362	
Uyanik et al,2013	Conflict	1,632	0,323	
Tokatli ,2009	Conflict	1,747	0,311	
Unlu, 2011	Conflict	1,762	0,226	
Uzun, 2010	Conflict	2,014	0,351	
Yuruk, 2009	Conflict	2,245	0,317	
Uzuntiyaki, 2003	Conflict	2,457	0,405	
Yumusak, 2011	Conflict	2,508	0,470	
Udogu et al.,2010	Conflict	2,521	0,406	
		0,747	0,047	



Precision for Studies and Efffect Size Value Relation

G. Forest Plots for Cognitive Bridging Studies

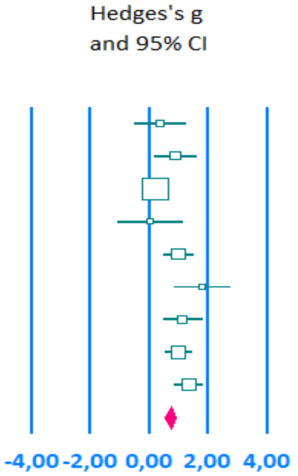
Study name	Type of CCM	Statistics for each study		Hedges's g and 95% CI
		Hedges's g	Standard error	
Gayeta et al,2017	Bridging	-0,135	0,284	
Kiliç 2007	Bridging	0,453	0,296	
Gokhale, 1996	Bridging	0,459	0,294	
Tokur et al., 2010	Bridging	0,479	0,225	
Hanson al,2018	Bridging	0,485	0,238	
Savinainen, 2003	Bridging	0,527	0,298	
Okur, 2009	Bridging	0,614	0,317	
Karakethudaoglu,2008	Bridging	0,655	0,327	
Harman,2016	Bridging	0,716	0,207	
Nwankwo, et al,2014	Bridging	0,718	0,196	
Aykutlu al.,2009	Bridging	0,724	0,293	
Woloshyn al,1992	Bridging	0,756	0,264	
Çibik, 2011	Bridging	0,758	0,229	
Gedik et al., 2001	Bridging	0,767	0,301	
Cheung,2008	Bridging	0,904	0,326	
Yilmaz al,2006b	Bridging	0,965	0,121	
Bilgin et al., 2001	Bridging	1,020	0,339	
Demirci et al,2003	Bridging	1,063	0,273	
Stavy, 1991	Bridging	1,214	0,274	
Koseoglu al.,2012	Bridging	1,226	0,347	
Cetingul al.,2011	Bridging	1,267	0,306	
Clement, 1993	Bridging	1,271	0,323	
Diakidoy al.,2001	Bridging	1,383	0,300	
Sendur, et al., 2008	Bridging	1,404	0,319	
Pekmez, 2008	Bridging	1,438	0,314	
Dilber, 2008	Bridging	1,705	0,302	
Yaman, 2010	Bridging	2,010	0,217	
Cetingul al.,2005	Bridging	2,016	0,354	
Yilmaz al,2006	Bridging	2,297	0,236	
Dilber et al,2008	Bridging	3,013	0,329	
		1,030	0,047	



Precision for Studies and Efffect Size Value Relation

H. Forest Plots for Ontological Category Shift Studies

Study name	Type of CCM	Hedges's g	Standard error
Yang al, 2012	Ontology	0,373	0,458
Slotta al,2006	Ontology	0,899	0,362
Erdmann, 2000	Ontology	0,233	0,142
Charles, 2003	Ontology	0,036	0,568
Uzuntiyaki al,2002	Ontology	1,013	0,263
Chiu and Lin, 2003	Ontology	1,808	0,493
Akgul, 2009	Ontology	1,146	0,350
Coruhlu et al.,2015	Ontology	0,993	0,248
Taþkýn, 2006	Ontology	1,349	0,249
		0,722	0,090



Precision for Studies and Efffect Size Value Relation

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Paçacı, Çağatay

Nationality: Turkish (TC)

Marital Status:

e-mail:

EDUCATION

Degree	Institution	Graduation Year
Ph.D.	METU, Mathematics and Science Education	2022
B.S.	METU, Physics Education	2011
High School,	Erzincan Milliyet Anatolian High School	2004

WORK EXPERIENCE

Year	Place	Enrollment
2016- 2022	Ministry of Education	Physics Teacher
2015-2016	Military Academy	Physics Lecturer
2013-2015	Ministry of Education	Physics Teacher

FOREIGN LANGUAGES

English TOEFL IBT: 82, YDS:87.5

PUBLICATIONS

No publication

PRIZES

Advisor, TÜBİTAK science competition between high school students.