Socio-scientific Argumentation in Science Education: A Systematic and Critical Review of Research

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)

am Munich Center of the Learning Sciences

der Ludwig-Maximilians-Universität

München

Vorgelegt von

Olga Ioannidou

17.12.2019

1st Supervisor / Erstgutachter: Prof. Dr. Tina Seidel

2nd Supervisor / Zweitgutachter: Prof. Dr. Frank Fischer

Tag der mündlichen Prüfung: 18.02.2020

Acknowledgements

The research presented in this thesis was supported by Elite Network of Bavaria (Project number: K-GS-2012-209). I would like to express my gratitude to ENB for providing me with the opportunity to present my work (e.g. in national and international conferences) and network with other researchers in the area of argumentation. With regard to the two studies presented in the thesis, Study 1 has been submitted to the International Journal of Science Education and Study 2 has been prepared for submission.

Although the following dissertation is an individual work, the completion of this thesis would not be possible without the help and support of a lot of people.

Firstly, I would like to thank my supervisors Prof. Tina Seidel and Prof. Frank Fischer, who guided me through my PhD studies. They were always positive to my ideas and they have provided the space and time for my skills and ideas to grow. I would also like to thank Dr. Andreas Hetmanek, who was my mentor for these three years. I would like to thank him for his insightful comments, his support and for inducting me into the world of research. I would also like to thank my international -and current- supervisor Prof. Sibel Erduran for her valuable feedback, but also her encouragement and her understanding that allowed me to finalize my PhD in a stress-free manner. Lastly, I would like to thank Prof. Sabine Pankofer for kindly accepting my invitation to be a member of the examination committee. Thank you!

Secondly, I would like to thank my family, my mother Eleni, my father Giorgos, and my brother Dimitris. And, of course, my partner, Ale, for his support and understanding. Although I live far away from them, knowing that they are always there for me gives me strength and hope to move forward. Thank you for existing!

Thirdly, I would like to thank my REASON colleagues. I would like to thank them for creating a pleasant and cheerful working space. I feel grateful for all my REASON friends and for the practical and psychological support that I have given and received. In particular, I would like to express my gratitude to my friend Despoina for always being my “deus ex machina” in any situation. She has been my long travel companion in my study years and I would like to thank her for always being so patient and caring. Also, I would like to thank my office mate and friend, Anastasia, for brightening up the cloudy days with her optimism and support.

I would like to extend my gratitude to all my friends in Munich -and worldwide- and especially Adriana, Nikoletta and Ioanna. Thank you!

Table of Contents

Abstract

During the last years, teaching and learning of argumentation have become an important educational goal. At the same time, argumentation and socio-scientific argumentation, in particular, are considered especially important in science education and have been introduced in science classrooms. As a result, within the last ten years, socio-scientific argumentation (SSA) has become a topic of research interest in science education. However, despite its importance, there is no consensus among researchers regarding the terms, definitions and measurement approaches currently used to address the topic. To address this issue and shed some light on the concept of socio-scientific argumentation, two studies were conducted to examine the various conceptualizations of SSA, as well as, the methodological approaches that have been used to assess the construct within the period 1999-2017. The first study aimed to outline the literature in the area of SSA with regard to the terms, definitions, argument assessment criteria and models used to assess SSA. At the same time, the first study aimed to give an overview of publication trends in the same period. The second study aimed to investigate the operationalizations of SSA in science education research. More specifically, the study examined the theoretical assumptions concerning the distinctiveness of SSA as a construct. In doing so, the second study analysed the selected studies in terms of construct and issue-specificity. Furthermore, special attention was given to potential misalignments between the mentioned theoretical assumptions and the applied methodology. A coding scheme was developed through inductive and deductive coding, based on which 67 studies were analysed following a mixed-method approach. The results of the first study suggest that there was an increase in publications on SSA since 2011, while the publications reached a peak in 2014 and in 2017. In terms of conceptualizations, the first study concluded that SSA has been mostly defined with respect to three themes, namely: (1)

argumentation regarding socio-scientific issues (SSI) (2) negotiation and decision-making to solve SSI and (3) informal reasoning regarding SSI. Moreover, the moral aspects of SSA have been considered of high importance. In terms of measurement, the first study showed that SSA has been analysed mostly qualitatively and eight argument assessment criteria were identified: (1) structure, (2) use of pre-defined criteria, (3) complexity, (4) multidimensionality, (5) use of scientific evidence, (6) accuracy, (7) coherence and (8) argument moves. Yet, researchers have mostly focused on structural criteria and models, such as Toulmin's Argumentation Pattern (TAP). The results for the second study showed that only a small number of studies explicitly mentioned that SSA is a distinctive construct, while 6 studies claimed that SSA is issue-dependent and 5 that SSA is issue-independent. In addition, the second study illustrated the heterogeneity with regard to the applied methodological approaches. Moreover, one of the findings in the second study was that only 9 out of 67 studies justified the applied argument assessment criteria. The main contribution of this study is the development of three themes describing the given definitions for SSA, as well as the presentation of an interpretative framework for the argument assessment criteria applied in the selected studies. Furthermore, the two studies raised questions regarding the processes of operationalization followed in the field of SSA. Ultimately, the studies aimed to contribute to future research on SSA by proposing a new procedural framework for the operationalization of SSA in science education research.

1. GENERAL INTRODUCTION

Our era has been often characterized as "the era of information explosion" (Kohle, 2018). Citizens receive a plethora of information from various sources. Yet, the ways in which citizens integrate the available information into their everyday argumentation and decision-making remains unclear (Doksæter Sivle, & Kolstø, 2016). These decision-making processes become more challenging when the debated topics derive from science and are presented as controversial in public discussions.

Examples of such issues that are currently discussed by the media are the topics of climate change, the consumption of genetically modified food and the use of vaccines. Such topics are frequently described as controversial (e.g. Lin, Lin & Tsai, 2014), as citizens are often exposed to a variety of contradicting arguments by a range of scientific, as well as non-scientific, sources. Thus, the question that arises is: How do individuals make decisions about these topics? What sources of evidence do they use to support their arguments?

In Europe there were more than 90,000 reported measles cases in 2019 (World Health Organization, 2019). Although there is no scientific evidence supporting that vaccines involve serious health risks, the recent measles outbreaks have raised questions about citizens' confidence in vaccines. In a recent report, the European Commission has associated the decrease of citizens' confidence in vaccines to the presence of "anti-vaccine groups in mainstream and social media" and their influence on "politics and political elections" (Larson, de Figueiredo, Karafillakis, & Rawal, 2018, p.47). In addition, in a study examining public confidence in vaccines, Larson et al. (2011) argued that: "Public decision-making related to vaccine acceptance is neither driven by scientific nor economic evidence alone, but is also driven by a mix of psychological, sociocultural, and political factors, all of which need to be understood and taken into account by policy and other decision makers" (Larson, Cooper, Eskola, Katz, & Ratzan, 2011, p.1). Thus, it becomes apparent that the reasoning and argumentation processes regarding controversial scientific issues are complex and multidimensional.

To address the complexity of these issues and prepare future citizens, science curricula in several countries have integrated the concept of argumentation regarding controversial scientific issues (e.g. Next Generation Science Standards [NGSS], 2013). At the same time, research in science education has termed them as “socio-scientific issues” (SSI) (Sadler, 2004) and, subsequently, the form of argumentation that is evolved in such contexts as “socio-scientific argumentation” (SSA) (Grooms, Sampson, & Golden, 2014). Nevertheless, despite the importance of these issues and recent attempts to integrate them in science education classrooms, there is a variety of theoretical and methodological approaches with regard to “what makes a good argument” in SSI contexts.

This thesis aims to tackle this problem and shed some light on the topic of socio-scientific argumentation in science education by examining the theoretical assumptions, as well as, the measurement approaches that have been applied to define and measure socio-scientific argumentation in science education. To that end, Study 1 systematically reviewed the conceptualizations and measurements that have been applied in science education research, since the appearance of the term “socio-scientific”. Followingly, Study 2 analyzed in more detail the operationalizations of “argumentation regarding SSI” by comparing and contrasting the theoretical and the methodological approaches used in the selected studies.

Before the description of the two studies, a General Introduction will present the background information that theoretically frames the two studies. The concepts and processes that are presented in this section will also help the interpretation and discussion of the results presented in Study 1 and Study 2. The chapter begins with Section 1.1, which describes some general educational goals in science education and the concepts of scientific literacy and scientific reasoning as educational goals are discussed.

In Section 1.2, the role of argumentation as an educational goal for science education is discussed. In addition, the conceptualization of scientific argumentation as an important part of science education is presented. Next, in Section 1.3 new approaches to the conceptualization of argumentation in science education are examined. This section discusses the impact of the introduction of concepts, such as the nature of science and socio-scientific issues, on the nature of argumentation in science classrooms. Subsequently, Section 1.4 is an introduction of socio-scientific argumentation as a new concept in science education. This section includes a description of the reasoning processes that take place during socio-scientific argumentation. The General Introduction continues with Section 1.5 and the introduction of measurement approaches that have been applied to measure argumentation skills in science education, as well as, the applied measurement tools for socio-scientific argumentation.

Finally, Section 1.6 of the General Introduction addresses the issue of operationalization in educational research. As Study 2 argues, the conceptualization and measurement of concepts in educational research should be studied in relation to each other. Thus, this section raises the question: “What are the operationalization processes followed in studies examining SSA?”. While the section begins with general issues of operationalization in educational research, science education and argumentative competencies, it closes with the discussion of more explicit issues in operationalizations of argumentative competencies in science education.

1.1 Educational Goals in Science Education

In order to highlight the reasons and contexts in which socio-scientific argumentation has become a topic of discussion in research, as well as teaching in science education, the first section will present an overview of the progression with regard to educational goals within science education. This progression has been described as a shift from “educating future scientists” to “educating future citizens” (Osborne, 2007), as a result of changes undergone in the last decades in light of new educational standards imposed by societal, as well as technological changes. Since scientific literacy is the term that is mainly used to represent this kind of goals for science education (DeBoer, 2000), the changes in the conceptualization of scientific literacy will be utilized as a starting point to discuss general shifts in the aims of science education. After a historical overview of conceptualizations of scientific literacy, an introduction of the concept of scientific reasoning as an educational goal in science in science education will be presented.

1.1.1 Scientific literacy

In the last decades, scientific literacy (SL) (DeBoer, 2011) has been internationally recognized as one of the principal educational goals for science education and, therefore, it has been empirically and theoretically examined by a plethora of studies. Although in everyday language, the term “scientific literacy” (SL) generally describes “what the general public ought to know about science” (Durant, 1992, p. 129), the construct has been conceptualized in science education research in various ways (Laugksch, 2000). It can be stated that the changes in SL conceptualizations mirror general educational goals in science education, as well as, research trends regarding other relevant constructs (e.g. the development of literature on Nature of Science views). In this section, a summary of the conceptualizations of SL will be presented, in order to highlight the differentiated foci and transformation of the meaning of the term within science education research over time.

In a first attempt to conceptualize SL, Pella et al. (Pella, O'hearn, & Gale, 1966) summarized the characteristics of a scientifically literate individual as: the ability of understanding interrelationships of science and society; ethics that influence the scientist in his work; nature of science; difference between science and technology; key concepts in science; and relationships of science and humanities.

In a later definition by Shen (1975) there are three categories of scientific literacy: practical, civic, and cultural scientific literacy. Shen claimed that from all categories the most important in industrialized countries is the civic scientific literacy. As explained, in societies with developed technology and science, 'average' citizens should be aware of science and science-related public issues. According to the study, only in that way citizens would be able to participate in decision-making and influence policymaking related to public issues, such as health, environment, or natural resources (Shen, 1975).

Some years later, Branscomb (1981) attempted to expand the notion of scientific literacy by suggesting eight categories: methodological science literacy; professional science literacy; universal science literacy; technological science literacy; amateur science literacy; journalistic science literacy; science policy literacy; and public science policy literacy. One could claim that Branscomb theoretically divided Shen's civic literacy into universal and science policy literacy. While Shen (1975) assumed that decision-making in democratic societies depends on all average citizens, Branscomb (1981) suggested a more specialized form of scientific literacy directed only to political representatives that make public decisions- science policy literacy.

Furthermore, Miller (1993) identified three major dimensions for conceptualizing SL: the understanding of scientific norms and methods; the understanding of key scientific terms and concepts (in the form of science content knowledge); and the ability to realize and understand the effect of science and technology on societies. In contrast, almost a decade later, Shamos (1995) gave a different perspective, supposing that there are three hierarchical levels of scientific literature. In his hierarchical model, 'cultural scientific literacy' describes the background knowledge of individuals of mainstream culture in areas such as history, language or science. This level is presented as a prerequisite for understanding information presented by media, elected representatives and public debates. In the second level, Shamos proposed the 'functional scientific literacy' which requires the ability to understand and communicate scientific concepts and findings. Finally, Shamos pointed out 'true scientific literacy' as the ultimate level of SL. He claimed that this level is achievable by a small part of society as it requires "highly specialized knowledge" (Shamos, 1995).

Visions for scientific literacy. While it is evident that the term "scientific literacy" has been conceptualized in various ways, Laugksch (2000) summarized the definitions and models that were used to describe SL by means of three different interpretations of the word "literate"; "literate as learned, literate as competent and literate as able to function minimally in society" (Laugksch, 2000, p. 82). These interpretations are mainly linked to the very nature of science; while "literate as learned" and "competent", assumes a de facto knowledge that can be learned, "literate as able to function minimally in society" may assume a relative definition of science, imposed by the needs of the society in which individuals live. Thus, this approach suggests that there is no universal body of scientific knowledge, as the latter heavily depends on the societies and the economies in which it is used (Laugksch, 2000; Miller, 1993). Roberts (2007) further described these definitions, claiming that there is a shift towards skills and competencies that reflect the needs of contemporary societies. In doing so, he presented the three visions of SL: *Vision I and Vision II and Vision III.*

According to Roberts (2007) Vision I emphasizes the aim of science education to teach concepts and processes related to science. This vision includes the educational standards referred to factual knowledge, learned skills and processes "within science" (Teig, & Scherer, 2016). It can be also concluded that such an approach would assume an absolute nature of science. Examples of standards containing goals allied to Vision I are contained in Science for All Americans (AAAS, 1990) and Benchmarks for Science Literacy (AAAS, 1993). Contrary to Vision I, Vision II of SL incorporates science literacy in citizenship education (Roberts, 2007). Under this scope, SL does no longer mean abstract “science for science” knowledge, but it also involves real-life situations and should be adapted to the needs of society. This conceptualization has been considered to be a part of progressive science education, as suggested by DeBoer (1991), as they are in line with the idea of being “literate as able to minimally function in society” (Laugksch, 2000).

According to Sjöström and Eilks (2017) our societies have become more complex due to the rapid developments in science and technology in the last decades. Thus, students should be able to make decisions regarding complex societal issues considering scientific, as well as social or ethical parameters (Sadler & Zeidler, 2005). Thus, the goal for science education should target not only technical and practical skills relevant to the needs of societies (as in Vision II), but also complex higher-order skills. Vision III has been introduced to address these challenges and emphasize the aspect of ethics and value-driven arguments in science.

1.1.2 Scientific reasoning

The concept of scientific literacy has been used as a broad term to describe scientifically thinking skills or scientific reasoning skills (Laugksch, 2000). Similarly, “thinking scientifically” or “working scientifically” are concepts frequently used in science education to represent “how science works” by presenting the practices commonly used in science (e.g. Smith, 2016). As such, they have been described as significant goals of science education. In what follows, an overview of the conceptualizations of scientific reasoning will be presented in order to highlight the shift from linear models of scientific reasoning (e.g. “the scientific method”) to rather descriptive models with regard to “how science works”. In addition, the scientific reasoning models that are presented in this section will enable the conceptual comparisons between scientific and socio-scientific reasoning models, which will be presented in the next sections.

Scientific reasoning has been generally defined as a complex thinking skill that involves a series of epistemic processes, such as constructing a hypothesis or evaluating evidence (Okada & Simon, 1997; Osborne, 2010). This skill is frequently highlighted as a strategy to draw conclusions from scientific evidence (Lee & She, 2010). The conceptualization of scientific reasoning in educational research follows two strands. Firstly, from a developmental point of view, theories suggest that scientific reasoning is developed in several stages following the natural development of human thinking (Inhelder & Piaget, 1958). According to these theories, the highest stage of scientific reasoning is termed “formal operational reasoning” and it is described as the ability to evaluate hypothesis with the use of evidence. Secondly, later conceptualizations of scientific reasoning were influenced by the Scientific Discovery Search (SDDS) model (Klahr & Dunbar, 1988). The SDDS model describes scientific reasoning as a problem-solving activity, which involves the construction of a hypothesis as well as, analysis of evidence.

In science education, scientific reasoning skills are generally conceptualised as the necessary skills that are required to acquire scientific knowledge (Hartmann, zu Belzen, Krüger, & Pant, 2015). According to a structural model for defining scientific reasoning, there are four sub-skills within the construct, namely: formulating research questions, generating hypotheses, planning investigations, and analysing and interpreting data (Liu, 2010). These skills have also been termed as scientific inquiry (Koslowski, 1996) or scientific thinking skills (Kuhn & Dunbar,1988). In this context, the construct oftentimes refers to the ability to design experiments, analyse scientific evidence and understand the concepts of complex theories of science (Zimmerman, 2005). Common scientific inquiry techniques applied in science classrooms include observation, comparison, experimentation and modelling (Crawford & Cullin, 2005; Grosslight, Unger, Jay, & Smith, 1991; Justi & Gilbert, 2003)

Although research in science education has adapted theories from cognitive psychology to describe scientific thinking as a psychological construct, curricula and educational standards operationalize scientific thinking as the thinking processes that take place when students follow the "scientific method". This approach manifests that students are expected to learn how to "think like a scientist" and use scientific practices that are commonly used in scientific communities. The scientific method, as often taught in science classrooms, is often taught a step-wise process including specific activities. It aims to demonstrate to students how to find patterns in natural phenomena by testing hypotheses, while models and theories are presented as the last step of the process.

The idea of a single scientific method has been universally applied as a way to teach scientific practices to students, in order to increase their interest in science and prepare "future scientists" (Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013; Windschitl, Thompson, & Braaten, 2008). However, the "scientific method" approach has received criticism by educational researchers who suggest that it oversimplifies the steps and practices that take place in scientific communities, while it emphasizes specific methods (e.g. hypothesis testing, experimentation) as being "more scientific" over others (Kind & Osborne, 2017). In support of the criticism, a number of studies have shown that instead of following a linear, step-by-step approach, scientists in different subdisciplines use a variety of methods to answer their questions and argue with evidence (e.g. Bauer, 1994).

In an attempt to recognize the pluralism of scientific methods, latest work in scientific reasoning in science education proposed the presentation of epistemic features in scientific reasoning, as well as the inclusion of the various scientific reasoning styles and scientific approaches other than hypothesis testing and manipulation of variables. For instance, Windschitl et al. (2008) proposed a model-based inquiry framework underlying five epistemic features of scientific knowledge, promoting the idea that scientific knowledge is testable, revisable, explanatory, conjectural and generative (Windschitl et al., 2008). In that sense, students would be able to distinguish scientific over “non- or less scientific” evidence and approaches by comparing them against these criteria.

Deriving from the same idea, Kind and Osborne (2017) suggested that the best way to address scientific reasoning is to study the reasoning styles adapted in scientific communities historically, rather than to investigate the developmental abilities of the learners. They argued that cognitive science is limited in examining the specific ways of thinking that are already culturally embedded. Hence, by adapting Crombie’s (1994) examination of scientific thinking, Kind and Osborne (2017) suggested six scientific reasoning styles: mathematical deduction; experimental evaluation; categorization/ classification; probabilistic reasoning; and historical-based evolutionary reasoning. Similarly, Erduran and Dagher (2014) suggested that scientific methods as taught in science education should include investigations with hypothesis and non-hypothesis testing, manipulative non-manipulative parameter measurement.

Despite the wide range of scientific approaches, Giere et al. (2006) argued that scientific reasoning follows a general pattern that can be described as a context-specific form of problem solving. However, according to Fischer et al. (2014) the difference between scientific reasoning and science-based problem solving is that in scientific reasoning the outcome is not only a solution to a given problem, but the construction of an argument based on scientific theory. Thus, arguments constructed in science classrooms are often regarded as the product of reasoning processes based on scientific theories. For this reason, argumentation has been extensively studied in the context of science education, as an educational goal, as well as the expression of scientific reasoning. In the next section, the role of argumentation in science education will be explicitly discussed.

1.2 Argumentation in Science Education

As presented earlier in the General Introduction, the general aim of this thesis is to provide an overview of the literature regarding socio-scientific argumentation in science education. Thus, before presenting the reasons and contexts in which socio-scientific argumentation became a part of science education, a general introduction of the role of argumentation in science education will be discussed. More explicitly, the reasons supporting the introduction of the concept of argumentation in science education will be presented, while the main approaches that have been followed in teaching argumentation in science classrooms will be summarized.

In the last decades, a substantial body of literature suggests that argumentation should be taught and learned in science classrooms (e.g. Erduran & Jimenez-Aleixandre, 2008). The introduction of argumentation in science education has been generally justified in light of three main arguments (von Aufschnaiter, Erduran, Osborne & Simon, 2008). The first argument is that argumentation naturally takes place in scientific communities in order to improve scientific knowledge. Scientific argumentation has been as a critical aspect of discourse processes in science (Toulmin, 1958), since scientists use and contrast arguments regarding a phenomenon in an attempt to find the most convincing claims in light of new evidence (Lawson, 2003). As a second argument, a number of researchers have claimed that argumentation skills are the prerequisite for science learning (von Aufschnaiter et al., 2008). Argumentation creates a discourse environment in which arguers explain their own thoughts, but are also exposed to ideas of others. In this context, conceptual understanding will emerge as a product of knowledge co-construction (Newton, Driver, & Osborne, 1999). By training argumentation as well as self-reflection skills, argumentative activities, subsequently, can facilitate modelling and the development of explanations with regard to student's individual thinking (Newton, Driver, & Osborne, 1999).

Finally, argumentation has been presented as a part of public debates. According to Simon et al. (Simon, Osborne, & Erduran, 2003): “Our decision making is often based on information available through press and media accounts, which may report contested claims arising from different sources of evidence. Evaluating such reports is not straightforward, as it requires the ability to assess the validity and reliability of evidence used in scientific arguments” (Simon, Osborne, & Erduran, 2003, p. 200).

In terms of teaching approaches and educational goals, argumentation in science education can be generally described with regard to two approaches: the “arguing-to-learn” (e.g., Andriessen, Baker, & Suthers, 2013; Jiménez-Aleixandre & Pereiro-Muñoz, 2002; von Aufschnaiter, Erduran, Osborne, & Simon, 2008; Zohar & Nemet, 2002) and the “learning-to-argue” approach (e.g., Erduran, Osborne, & Simon, 2004; Kelly, Drucker, & Chen, 1998; Kuhn, 2005; Reznitskaya et al., 2001). The “arguing to learn approach” regards argumentative activities as a means to facilitate students’ conceptual understanding. As a result, the primary goal of these activities is the acquisition of content knowledge in a given area. In contrast, the “learning to argue” approach suggests that dialogic and argumentative tasks should aim to teach and improve argumentative skills per se. Towards this approach, teachers and researchers focus on the teaching and learning of argument forms and schemes in order to enable them to construct their own arguments and evaluate arguments of others.

In science education, argumentation is often examined in linear models of scientific reasoning, such as “the scientific method” (Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013). In this context, arguments are often constructed towards the end of an investigation, when students would present their results in order to receive feedback. However, this implies that, in this context, argumentation mainly acts retrospectively in order to justify claims with evidence (Edruran & Dagher, 2014). At the same time, students are expected to argue individually comparing and contrasting their ideas in order to find “the right solution”. Contrarily, in line with constructivist theories for argumentation, other approaches emphasize the social aspect of reasoning and argumentation (e.g. Mercier, 2011). Indeed, although the definitions for scientific reasoning mainly describe thinking processes on the individual level, research in scientific argumentation emphasizes the social context of argumentation tasks (Berland & McNeill, 2010; Weinberger, Stegmann, & Fischer, 2010).

The social aspect of scientific argumentation is evident in two ways. Firstly, students explain their reasoning process, by presenting the collected evidence, as a result of their investigations. This explanation requires an audience, which will engage in discourse by asking for clarifications on specific data or the applied methodology. In addition, these tasks are often organized in the form of groupwork, where students have to explain their thinking throughout the whole activity to the members of the team, in order to reach a consensus. Secondly, apart from self-explanation functions, argumentation in science education facilitates students’ engagement in dialogue by means of critique and evaluation of arguments constructed by others (Berland & Reiser, 2009; Osborne, 2010; Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013). This critique can take the form of attack against the claims, as well as the scientific methods used to acquire the presented evidence. In addition, the critique can be directed towards not only peers, but also towards scientists (“how scientists know”) (Erduran & Jimenez-Aleixandre, 2008, 2012). Furthermore, within the concept of scientific argumentation, students may question the logic and coherence between claims and evidence (Klahr & Nigam, 2004; Kuhn & Pease, 2008).

1.3 New Conceptualizations to Argumentation in Science Education

As described in the previous section, in the last decades science education researchers and policymakers have recognized the importance of argumentation as a learning outcome. Nevertheless, the goals and contexts in which argumentation in science education has been taught, learnt, and measured, have been under discussion. According to Erduran and Jimenez-Aleixandre (2008), teaching argumentation in science education is oriented towards three goals: firstly, the development of student's views regarding the Nature of Science (NOS); secondly, the inclusion of citizenship education with regard to socio-scientific issues; and thirdly, the facilitation of high order thinking skills, including argumentative skills per se. Therefore, in this section an overview of the impact of Nature of Science views on argumentation in science education will be outlined. Next, the introduction of socio-scientific issues in science education to promote citizenship skills, as well as some of the challenges with regard to the negotiation of these issues, will be discussed.

1.3.1 Nature of science views

Following the tendencies in teaching and learning scientific literacy and reasoning skills, the focus for argumentative competencies has traditionally been the acquisition of content knowledge through formal reasoning. Nevertheless, as the nature of scientific knowledge and scientific practices became more ambiguous in science education research, more discussions about the nature of knowledge and argumentative skills in science classrooms emerged. In this regard, literature on argumentation skills has been influenced by the work of Nature of Science (NOS) research (e.g. Abd-El-Khalick, 2012; Abd-El-Khalick & Lederman, 2000; Akerson & Donnelly, 2008). Some of the fundamental ideas about NOS across literature include the idea of the tentativeness of scientific knowledge, the theory-laden nature of scientific knowledge and the myth of the scientific method (Chang, Chang & Tseng, 2010).

1.3.2 Socio-scientific issues

The adaptation of NOS approaches in science teaching promoted the contextualisation of content knowledge and made way for more informal approaches of scientific knowledge.

As a result, more “ill-structured, open-ended and debatable problems” (Sadler, 2004; Teig & Scherer, 2016) were introduced in science classrooms, on which students were expected to reason and argue. These new open-ended topics, such as GMO, environmental issues or nuclear energy were termed as “socio-scientific issues” (SSI) (Sadler, 2004). The term socio-scientific has been used to illustrate the dual nature of SSI; SSI are scientific issues because they emerge from scientific or technological achievements, but they are at the same time social issues, given that they have an impact on society (Sadler, Barab, & Scott, 2007).

Socio-scientific issues were introduced in science education as part of citizenship education in an attempt to enable future citizens to argue about current scientific issues relevant to society (Driver, Leach, Millar, & Scott, 1996; Millar & Osborne, 1998). By doing so, students would be able to participate in public debates, such as the consumption of genetically modified food or the use of vaccines (Sadler et al., 2007; Zeidler, Osborne, Erduran, Simon, & Monk, 2003). Although the majority of researchers and educators regarded the introduction of SSI not only productive, but essential element of contemporary science classrooms (Driver, Newton, & Osborne, 2000; Sadler & Zeidler, 2004; Sadler et al., 2007), Erduran and Jimenez-Aleixandre (2008) argued that the legitimization of citizenship education as part of science education was not straightforward, as it depends on societies’ needs and assumes consensus regarding the goals of science education.

The main argument for the introduction of SSI in science classrooms was the claim that our societies gradually become “post-scientific” (Osborne et al., 2013) and that in such societies interdisciplinary skills and complex thinking would be prioritized over disciplinary content knowledge and expertise. With the term “post-scientific”, researchers emphasize the oversupply of information from various sources (Gilbert, 2005), that leads to two issues; firstly, accuracy (or “scientificness”) of information becomes questionable and secondly, even when information is assumed scientifically correct, it is not automatically assumed to be the only (or the best) source of information.

To illustrate these challenges, one could consider the example of human cloning as a socio-scientific topic currently presented by the media. A typical discussion about this topic would pose the question “to what degree is human cloning ethical?”. A plethora of information would be available on the topic by various sources of information, such as the newspapers or the internet. In this case, citizens are supposed to evaluate the evidence, by examining the source in order to find the most reliable information. However, this reliable, scientific information would only help them understand what are the possible procedures and implications of human cloning. Hence, the individuals would not be able to answer the question using solely scientific evidence (Sadler, 2004). Citizens should rather gather additional arguments from non-scientific sources (e.g. animal and human rights organizations) to examine potential ethical considerations regarding human cloning. Therefore, it can be easily understood that a set of different skills need to be orchestrated in order to answer complex questions, such as SSI.

1.4 Socio-scientific Argumentation

The set of skills described in the previous section, such as the ability to reason and argue about complex issues, such as SSI, were relatively recently introduced as educational goals for science education with the term “socio-scientific argumentation”. As described in the previous sections, the main aim of this thesis is to present an overview of conceptualizations and measurement approaches to assess socio-scientific argumentation. Since Study 1 and Study will present these conceptualizations in extensive detail, this section will only give a short introduction to the concept. In addition, this section will outline the conceptualizations of the reasoning processes occurring during the negotiation of SSI.

According to some researchers, the form of argumentation that is produced towards the negotiation and resolution of socio-scientific issues is a distinctive type of argumentation, termed “socio-scientific argumentation” (SSA) (Grooms, Sampson, & Golden, 2014). In the context of SSA, students have the opportunity to engage in authentic, everyday situations in which they are expected to make decisions based on multiple factors, such as social norms and values, which exceed scientific knowledge (e.g., Kuhn, 1993; Sadler, 2004). As a result, students often have to present and weight the advantages and disadvantages of a claim, seek alternatives and consider different points of view and solutions (Zohar & Nemet, 2002). Thus, SSA has been extensively studied as a form of argumentation in science education, as instruction in SSA-contexts has been associated with gains in content knowledge, conceptual understanding and general argumentative skills (e.g. Erduran, Simon, & Osborne 2004).

1.4.1 Reasoning processes during SSA

Since argumentation is often defined as the expression of reasoning processes, argumentation regarding SSI has been examined in conjunction with reasoning processes during the negotiation of SSI. These complex thinking processes have been mainly described in relation to three reasoning models; informal reasoning (e.g. Wu & Tsai, 2007), moral reasoning (e.g. Sadler & Donnelly, 2006), and socio-scientific reasoning (Romine, Sadler & Kinslow, 2017). In this section, these three models for reasoning processes during the negotiation of SSI will be described.

Informal reasoning. According to the first approach, SSI stimulate a special mode of reasoning termed “informal reasoning”, which differs from formal reasoning, as traditionally taught in science classrooms. More explicitly, this distinction between formal and informal reasoning assumed that formal reasoning integrates principles of deductive reasoning based on rules of logic and mathematics (Wu & Tsai, 2007). In addition, formal reasoning tasks are well-defined questions, in which the premises must be assumed to be true in order to reach a conclusion with certainty (Evans & Thompson, 2004). During this process, arguers should not consider external sources of information (e.g. beliefs), as they are supposed to solve the problem solely by using the contextualized information provided especially for a specific task (Evans, 2002; Evans & Thompson, 2004).

In contrast, informal reasoning takes place when individuals are asked to solve open-ended issues, with no clear-cut answers (Sadler, 2004). According to several researchers, informal reasoning, like formal reasoning, follows a logical process of constructing and evaluating arguments (Kuhn, 1993). This logical process had been described in relation to mental model theories from cognitive psychology (Evans & Thompson, 2004), one of which is the dual-process theory (Wu & Tsai, 2007). The dual-process theory represents informal reasoning including two cognitive systems (System 1 and System 2), as presented in Figure 1. This scheme illustrates the cognitive processes that take place when individuals are asked to solve an ill-structured problem. According to the scheme, System 1 is firstly activated and individuals perform automated, unconscious reasoning to solve the problem. If the problem cannot be solved by System 1, the reasoning process is transferred to System 2, where acquired knowledge is utilized and alternative hypotheses are generated (Wu & Tsai, 2007).

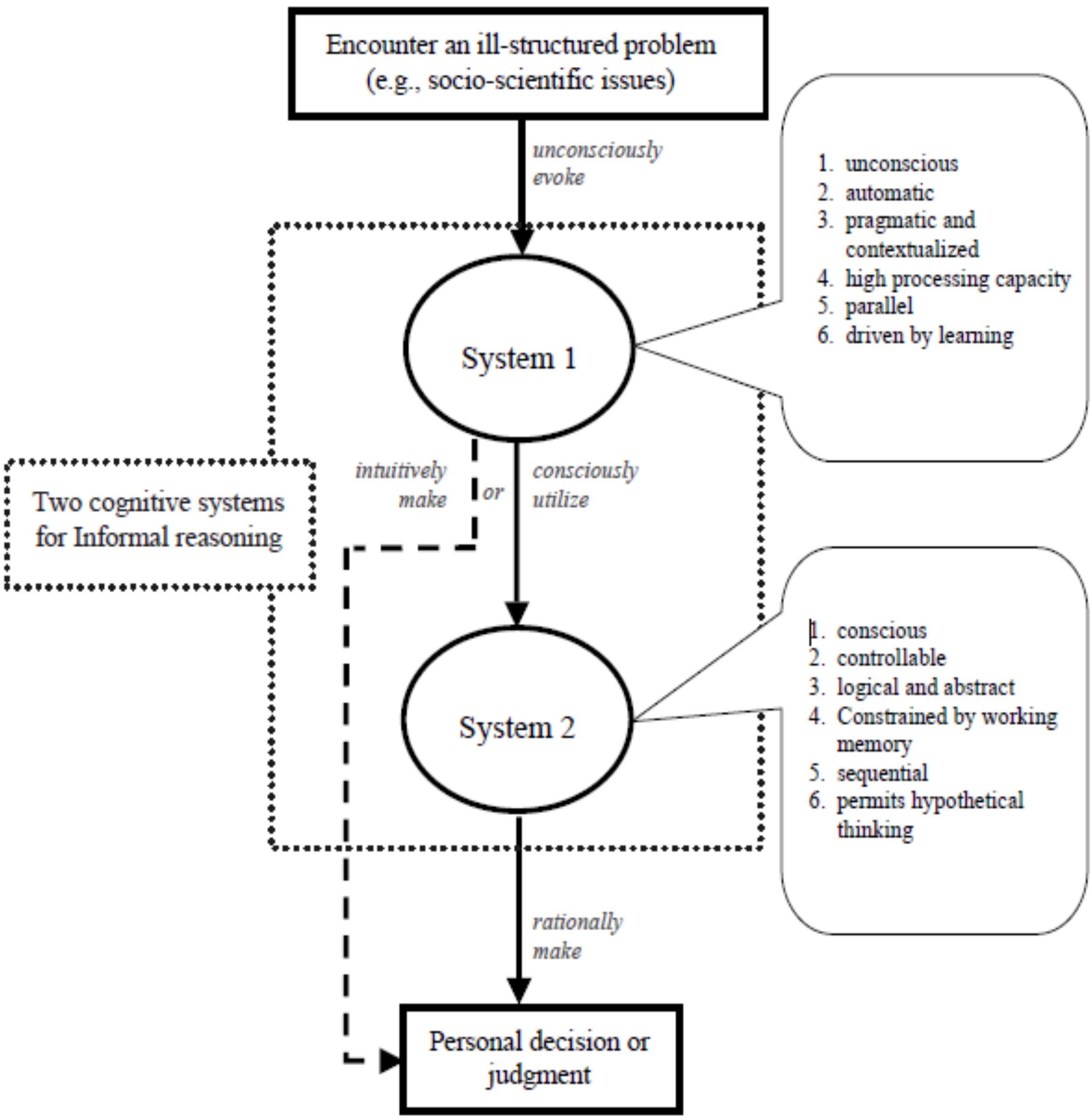

Figure 1
The dual process model of informal reasoning (Wu & Tsai, 2007)

Moral reasoning. According to another approach, the reasoning processes that take place during the negotiation of SSI are moral reasoning processes (e.g. Sadler & Donnelly, 2006). This approach suggests that discussions regarding SSI are mainly based on personal values and moral considerations, including general ideas of what is right, good, and ethical. In this direction, Sadler and Donnelly (2006) drew on moral philosophy (e.g. DeMarco, 1996), as well as religious traditions as sources of evidence. More specifically, they used the neo-Kohlbergian approach, which describes moral reasoning with regard to three dimensions: personal interests, maintaining norms and post-conventional reasoning (Mechler & Thoma, 2013). According to this theory, individuals may decide based on their personal interest, when they prioritise their personal satisfaction, as well as the protection of their rights and needs. When the decision is made based on maintaining norms, the primary goal would be to respect social and legal established rules, while during post-conventional reasoning it is assumed that social norms are negotiable commonly shared ideas (Mechler & Thoma, 2013).

Socio-scientific reasoning. According to a third approach, reasoning during the negotiation of an SSI is a distinct construct termed "socio-scientific reasoning" (SSR) (Sadler & Zeidler, 2009; Romine, Sadler & Kinslow, 2017). This assumption is based on the idea that the cognitive processes that are activated within the resolution of an SSI can be classified as naïve or low performance and sophisticated performance (Romine et al., 2017). Thus, one of the goals of instruction in science education classrooms would be the facilitation of student thinking in order to achieve sophisticated levels of SSR. The performance levels of SSR were initially proposed to be measured by four dimensions of SSR, namely: "i) recognizing the inherent complexity of SSI; (ii) examining issues from multiple perspectives; (iii) appreciating that SSI are subject to ongoing inquiry; and (iv) examining potentially biased information with skepticism" (Romine et al., 2017, p. 277).

Other conceptualizations of SSR expanded the construct by adding more dimensions; for instance, Simonneaux and Simonneaux (2009) proposed the addition of risk identification, as well as ethico-cultural considerations. Later conceptualizations of SSR integrated the construct of argumentation with a focus on specific issues, such as sustainability (Morin, Simonneaux, Simonneaux, & Tytler; 2017).

1.5 Measuring Argumentative Competencies in Science Education

Although there is a general agreement on the importance of argumentation as a desired educational outcome, argumentative competencies in education have been defined and, subsequently, measured in various ways, due to the great variety and variability of the factors that the construct encompasses. At the same time, argumentation has been studied in the context of science education for only a few decades, while the construct has a long tradition in other disciplines, such as philosophy (Rapanta, Garcia-Mila, & Gilabert, 2013). Thus, the measurement approaches, models and tools that have been used to assess argumentation in science education, and subsequently SSA, have been adapted or influenced by measurement approaches and tools that have been traditionally used to assess argumentation in other contexts.

For this reason, this section summarizes the main measurement approaches followed to assess argumentative competencies in science education by means of three themes: (1) argument as a form; (2) argument as a strategy and (3) argument as a goal (Rapanta et al., 2013). In addition to the themes, the next section will present the measurement models and quality indicators for argumentative competencies in science education.

1.5.1 Measurement approaches

Argument as a form. This theme is based on the idea that argumentative activities should shift the focus from the collection of data and warrants to back up a claim, to the establishment of a logical relationship between data-grounds-claims. To that end, the "argument as a form" approach derived from Toulmin's work (Toulmin, 1958), according to whom "a claim without grounds is not an argument and an argument without a warrant is not a legitimate argument" (Toulmin, 1958, p.98).

The scheme describes argument in terms of four basic components (Figure 2): data or fact, which is the piece of evidence to support a claim; warrants are the connection between data and claim; backings, which are explanations of why the use and type of data is appropriate to answer the given question; and finally, rebuttals, which are instances in which the claim is not valid. Toulmin's Argumentation Pattern (TAP), which constitutes one of the most influential assessment frameworks for argumentation in various contexts, is based on the idea that a statement qualifies as an argument if it contains all the structural components of the scheme. This focus on form and argument components led to later characterizations of TAP as a strictly structural assessment framework (e.g. Nielsen, 2012). However, despite the criticism that the scheme has received, a number of studies have used its original or adapted version as a tool to analyse teacher and student argumentation (e.g. Erduran et al., 2004; Mitchell, 1996). For instance, Erduran et al. (2004) examined secondary school students' verbal argumentation using an adapted version of TAP, based on which arguments were coded upon the differentiation between claims, justifications, and rebuttals.

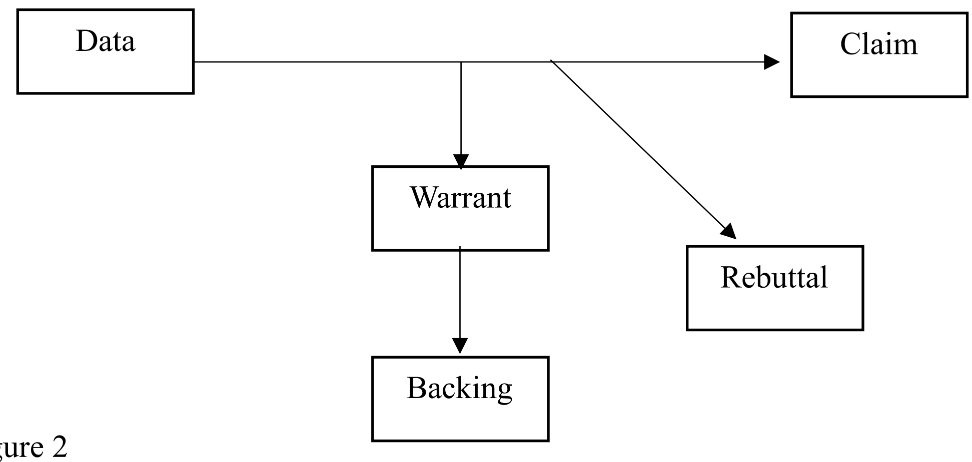

Figure 2

Toulmin's Argument Pattern (Toulmin, 1958)

Argument as strategy. Another approach for defining and measuring argumentation is "argument as a strategy" approach. According to this approach, argumentation is a social activity that involves individuals with competing arguments (Felton, 2004). In this case, arguments are analysed in terms of argument moves (Walton, 1996), which are discourse nodes or statements found in dialogical contexts. Since these dialogical moves are difficult to define, as they depend on the context, there have been many suggestions towards the differentiation of argument moves. The most dominant elaboration of this approach has been done by Kuhn (1991), who suggested that argument skills include sub-skills such as the ability (1) to construct an argument, (2) to justify an argument, (3) to construct a counterargument and (4) rebut others' counterarguments. In line with Kuhn's approach, Walton (1996) suggested a variety of argument schemes often used in support of one's claims, or to attack the opponent's argument (e.g. argument from expert opinion or argument from popularity).

Argument as goal. According to this approach, argumentative activities in education should not be examined in terms of the form, the quality of the arguments or even argument moves. Argumentation is rather seen as a "holistic" activity and should be judged as a whole. Since the quality criteria for argument as goal cannot be strictly defined, a traditional criterion has been persuasion (e.g. Walton, 1989). Although this approach has similarities to the "argument as strategy" approach, the main difference is that "argument as goal" promotes rhetorical (Perelman & Olbrechts-Tyteca, 1969), rather than dialectical student skills, which means that the social aspect of argumentation is to a certain degree neglected. This explains why this approach is as often adopted by educators or educational researchers, as the argument as a strategy approach (Rapanta et al., 2013).

1.5.2 Measurement tools for SSA

As described in the introduction of this section, SSA has been examined in various ways by using a range of assessment tools. In this section these tools will be outlined with regard to three main approaches of measuring SSA: (1) the use of already established frameworks for argumentative competencies to assess SSA, such as TAP, or its adapted versions (e.g. Erduran et al., 2004), (2) the application of frameworks that were primarily developed with a focus other than argumentation (e.g. Lakatos' scientific programmes) and

(3) the use of new argument assessment tools specially developed to capture the features of SSA (e.g. SSE-SEP framework).

Traditional argumentation schemes. The first category includes studies that have used the original or adapted versions of schemes traditionally used to assess argumentation skills in order to assess argumentation skills regarding SSI. For instance, Lin, Hong and Lawrenz (2012) used TAP to examine university students' argumentation skills regarding the construction of a power plant, the consumption of genetically modified food and the expansion of crops for manufacturing bioethanol. The coding scheme used in their study proposed the categorization of arguments into five levels. This categorization is based on the number and combination of argument components, as presented on Table 1.

Table 1

Adaptation of TAP as a scoring rubric for SSA (Lin et al, 2012)

Level	Description
1	Simple claim or proposition
2	Claims with valid data, warrant, or backing
3	Claims with either more than one valid supporting evidence (e.g., data, warrant, or backing) or one valid evidence and one weak rebuttal
4	Claim, valid evidence, and one identifiable rebuttal
5	Extended argument with claims supported by data and warrants with more than two identifiable rebuttals.

Other traditional schemes that have been applied in SSI contexts include Kuhn's (1993) and Walton's (1989) argument schemes. For instance, Iordanou and Constantinou (2015) used Kuhn's argumentation pattern to assess 11th graders' argumentation skills regarding climate change. In a case study with 11th graders Jimenez-Aleixandre (2002) used Walton's argument scheme to assess student arguments regarding environmental management. More specifically, in her coding scheme, Jiménez-Aleixandre emphasized Walton's (1989) five critical questions for arguments from expert opinion: "1) Is the expert a genuine expert in the domain? 2) Did the expert really assert the proposition? 3) Is the proposition relevant to the domain under discussion? 4) Is the proposition consistent with what other experts in the domain say? 5) Is it consistent with known evidence in the domain?" (Jimenez-Aleixandre, 2002, p.1175).

Adapted frameworks for measuring SSA. The second category refers to studies that have adapted theories and frameworks that have been originally developed to describe other constructs. An example of this category is the work of Chang and Chiu (2008), who adapted Lakatos' scientific research programmes in order to evaluate undergraduate students' informal argumentation about SSI.

Lakatos' scientific research programmes (Lakatos, 1978) were introduced in order to describe theories within a domain of science, as well as the shift from one theory to another. Chang and Chiu (2008) suggested that Lakatos' framework could be useful to assess SSA when warrants and backing are difficult to distinguish and when students' arguments are hard to change. They compared Lakatos' scientific programmes to the skills involved in informal argumentation (Table 2). More explicitly, the "hard core" foundation (HC) of the theories were compared with the ability to make claims and provide supporting reasons. The positive (PH) and negative heuristics (NH) are strategies to defend a claim are embedded in the protective belt (PB). These are strategies to address rebuttals and expand theory, thus, they were compared with the ability to qualifiers present and counter-arguments, respectively.

Table 2

Components of Lakatos' Programmes Compared with Informal Argumentation Skills (Chang & Chiu, 2008, p. 1758)

Components of Lakatos' programmes	**Skills of informal argumentation**
Hard-core (HC)	Making claims
	Providing supporting reasons
Negative heuristic (NH)	Presenting counter- arguments
Positive heuristic (PH)	Showing qualifiers
Protective belt (PB)	Evaluating arguments

SSA-specific instruments. Lastly, another approach that has been used to assess SSA is the development of new frameworks especially designed to measure the quality of argumentation in SSI contexts. In the present thesis, this category will be referred to as "SSA-specific instruments", which contains a sub-category; "SSI-specific instruments". While SSA-specific instruments have been designed to acknowledge and capture the special features of SSA, the development of SSI-specific instruments emphasized the contextual characteristics of the specific SSI under discussion.

Although there are only a few examples in this category, the most representative model is the model SEE-SEP model presented by Chang Rundgren and Rundgren (2010). The model's name is an acronym of the six subject areas which are involved in the negotiation of SSI, namely: sociology (So), environment (En), economy (Ec), science (Sc), ethics (Et) and policy (Po). These subject areas are interconnected with three aspects: knowledge, value and personal experience. These combinations result in overall 18 codes, as presented in Figure 3.

The SEE-SEP model has been applied as a tool to assess the quality of argumentation regarding SSI by other researchers, as it acknowledges the interdisciplinarity and complexity of SSI (citation) by including various fields and sources of information. For instance, Christenson et al. (2014) applied SSE-SEP to assess secondary school students' arguments about global warming, genetically modified organisms (GMO), nuclear power, and consumer consumption (Christenson, Chang Rundgren & Zeidler, 2014).

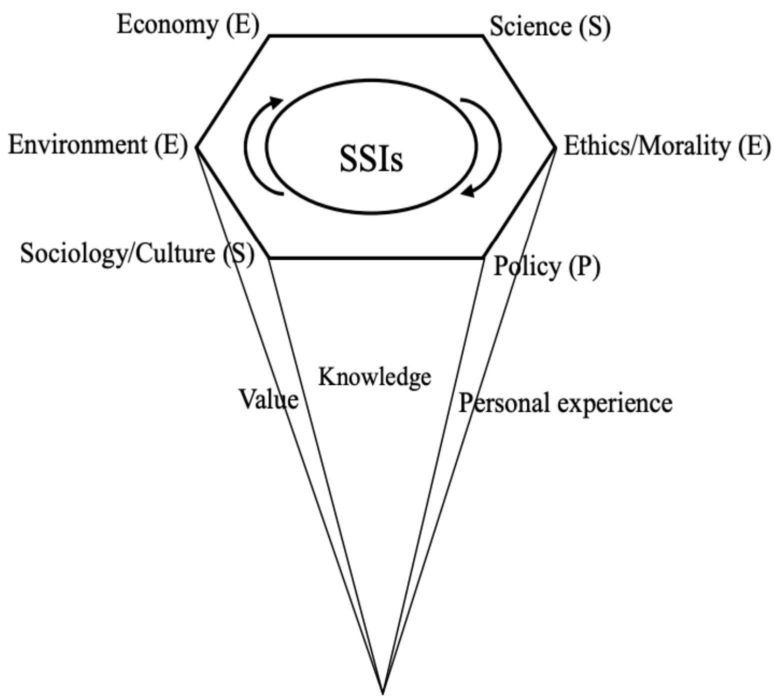

Figure 3

The SEE-SEP Framework Presented by Chang Rundgren and Rundgren (2010, P. 12)

1.6 Operationalizations in Educational Research

In the previous section, the basic definitions, as well as methodological tools for the assessment of socio-scientific argumentation, were discussed. Although theoretical frameworks and assessment tools are frequently discussed and studied separately, this thesis argues that theory and methods have a reciprocal relationship and, as a result, they should be studied in relation to each other.

To that end, a number of educational researchers have stressed out the importance of assessment methods for the interpretation of research results (e.g. Furtak et al., 2012;

Duckworth & Yeager, 2015). For instance, in a study investigating the assessment tools for personal qualities, Duckworth and Yeager (2015) pointed out that "measurement matters". They argued that, although researchers have gradually developed an interest in measuring changes and growth in various variables, such as cognitive abilities, there is a need for greater clarity about the degree to which the used tools are valid and reliable. While the reliability of the instruments relies on the quality of the measurement per se, validity reflects the interplay between conceptual ideas and methodological approaches (Duckworth & Yeager, 2015). This relationship between conceptualizations and measurement has been also highlighted by other studies. For instance, Blömeke et al. (2015) argued that the definition and the specific characteristics of competencies in higher education should direct researchers to a measurement model. The authors also recognized that this requirement creates methodological challenges that are often overlooked in educational research.

Therefore, in the next section, current issues and challenges in the operationalizations of argumentative competencies will be presented, as well as issues of operationalizations in science education research. Ultimately, the section will introduce some of the current challenges with regard to operationalizations in argumentative competencies in science education.

1.6.1 Operationalizations in science education

Meta-analyses and literature reviews reveal that constructs studied in science education are often ill-defined, rendering the process of operationalization challenging. For instance, in a meta-analysis on inquiry-based science teaching, Furtak et al. (2012) emphasized that the various conceptualizations of inquiry have an impact on the validity of any meta-analysis that endeavors to summarize its relationship with other constructs. They argued that the differentiated definitions, as well as the measurement tools applied by researchers in primary studies, would significantly affect the interpretation of their results.

Similarly, in a theoretical paper addressing thinking skills and creativity in science education, Osborne (2013) argued that the constructs are ill-defined rendering them difficult to assess. He claimed that when the constructs are not well-defined, there are no clear expectations for student performance and, as a result, the development of assessment tools becomes challenging for researchers and teachers. Similar claims regarding the ambiguity of terms and definitions and the subsequent challenges in developing assessment tools have been made for several constructs in science education, such as practical science (e.g. Gott & Duggan, 2002), nature of science (e.g Erduran, 2007) and scientific literacy (e.g. Laugksch, 2000).

1.6.2 Operationalizations in argumentative competencies

Analogously to operationalizations of constructs in science education, operationalizations of argumentative competencies in education have been presented as challenging. In a systematic literature review regarding argumentative competencies, Rapanta et al. (2013) concluded that there is a variety and heterogeneity in the conceptualizations of argumentative competencies because of the complexity of the construct. As a result, they argued that each theoretical framework focused on different aspects of argumentation resulting in a variety of definitions and quality indicators. These theoretical assumptions led to diversified operationalizations and consequently to a variety of measurement tools.

This challenge was explained by Rapanta et al. (2013) by two factors. Firstly, the theoretical approaches towards a definition of argument derive from philosophical studies focusing on how arguments are produced and evolve in a variety of contexts. Thus, the adaptation of philosophical frameworks to educational settings may be problematic, since the former focuses on theoretical discussions and the latter on specific skills and educational goals. Secondly, according to Rapanta et al. (2013) the skills that are assumed to be present during argumentation include cognitive, as well as meta-cognitive and epistemic aspects

(Kuhn, 2000). As a result, the assessment of student argumentation skills in education encompasses the same challenges with the study of complex competencies, as mentioned earlier in this section (e.g. Blömeke et al., 2015).

1.6.3 Operationalizations of argumentative competencies in science education

According to Erduran (2007) the main concern when studying argumentation in science education is methodology. One example of a methodological concern is the distinction between data and warrants, as presented by TAP (Toulmin, 1958). However, the concerns can include more general methodological approaches; for instance, the way in which researchers choose their coding tools and how they argue about their choices (Erduran, 2007). In this section, I will draw on the claim presented by Erduran (2007) that the assessment of argumentative competencies in science education is influenced by the pedagogical goals of the activity. This approach suggests that student arguments should primarily be assessed in terms of pedagogical as well as educational goals. The influence of pedagogical goals is further apparent in three levels/stages of research: (a) the description of quality indicators (b) the selection of the type of task and (c) the choice of type of assessment (Figure 4).

At the first level, pedagogical goals guide the choice of quality indicators used to assess argumentation skills. These choices can be often explained in relation to the socio-cultural contexts and the very nature of scientific knowledge that they attempt to assess (Erduran, 2007). For instance, in the case of socio-scientific argumentation, supposing that the main purpose of the given task is to argue with the use of multiple sources of information, one would expect complexity of the argument to be the main quality criterion for assessing student arguments.

In the next step, the chosen quality indicators are reflected by the tasks in which argumentation is taught and learned. Following the above-mentioned example, when complexity is chosen as a quality indicator, the argumentation task should provide the optimal conditions for students to develop complex arguments. Such conditions may include the presentation of multiple sources of evidence (e.g. refutational text, citation) or dialogical environments where students are exposed to opinions expressed by their peers.

Lastly, the quality indicators (and subsequently the pedagogical goals) should also be the foundation on which methodological tools, such as coding schemes, are designed. More explicitly, if we consider complexity as a quality indicator, the applied instrument should be suitable to capture multiple evidence used by students, and multiple reasons/ grounds used to justify a claim. Thus, with regard to this specific example, one would expect the applied coding scheme to be discourse or content analysis with weighted points on the complexity of the arguments. In doing so, Erduran (2007) suggests the establishment of specific rules for argument coding especially with regard to the different elements of argumentation.

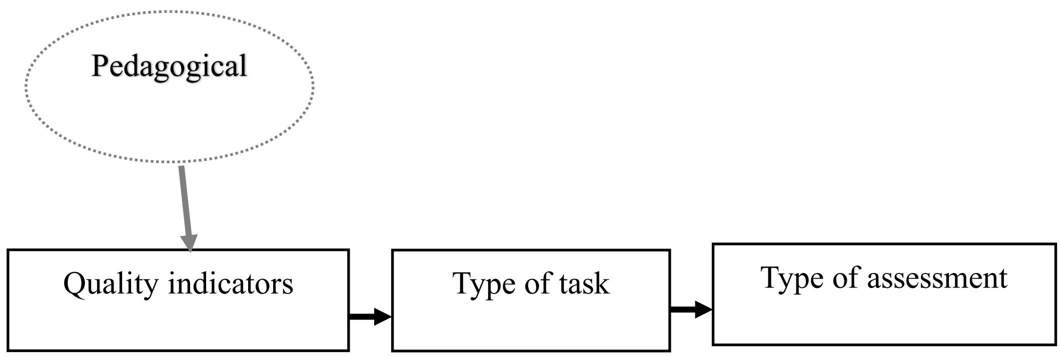

Figure 4

The Process of Operationalization of Argumentative Competencies in Science Education (adapted by Erduran, 2007).

This methodological approach is rather linear and suggests that the pedagogical goals and quality indicators are pre-determined. As a result, it would serve as a model for studies investigating the acquisition or improvement of argumentative skills in science education. Nevertheless, there exist studies with a different focus; a number of studies are only interested in exploring reasoning and argumentation patterns and modes that arise naturally in contexts such as SSI. These studies shouldn't be expected to follow a linear approach, as they aim to rather explore and describe than to assess student argumentation.

1.7 Research Goals and Questions

As stated in the previous sections, socio-scientific argumentation skills are important skills for future citizens to "minimally function in society" (Laugksch, 2000). The need for such skills is crucial especially in modern societies, where citizens often receive an excessive amount of information before they make decisions. For this reason, the concept of socio-scientific argumentation has become a part of science education and there is an increasing number of researchers addressing the topic in science education research.

However, previous studies in the topic suggest that there is clear definition of the term, and similarly to the general argumentative competencies in educational research, the concept has been measured in various ways. Despite the increase in publications on the topic, there has been no systematic review on the conceptualization and measurement of socio-scientific argumentation in science education.

Thus, the first main goal of this thesis is to provide an overview of the research in the area of SSA. More specifically, the first aim is to systematically analyze the studies that have been published in science education research, since the introduction of the term "socio-scientific", in terms of the presented terms, definitions, conceptualisations and the applied measurement approaches. This goal will be addressed in Study 1 and will be presented in relation to the following research questions:

(1) How did SSA research develop in scope and publications in the last decade?

(2) How is SSA theoretically framed in recent publications? What terminology, definitions and concepts are used to describe SSA in science education research?

(3) What methodological approaches are applied to empirically assess SSA?

- What approaches of operationalization are pursued?
- What instruments are used?
- What argument criteria are applied?

By answering these questions, this thesis aims to provide readers, researchers, as well as educators, with a comprehensive framework to facilitate the interpretation of research findings and methodological approaches in the area of SSA. At the same time the study aspires to open a discussion about the specific features of SSA and possible ways of measuring the quality of argumentative competencies in SSI-contexts. While doing so, the study aims to illustrate potential commonplaces between researchers with regard to the theoretical and methodological approaches presented in the selected studies.

The second main goal of this thesis is to examine the distinctive characteristics of SSA as a construct, by conducting an in-depth analysis of the existing literature on the topic. In parallel, another aim of this thesis is to present an overview of the operationalization processes that were followed in the selected studies. These goals are addressed by Study 2 and are in line with the following questions:

(4) Is SSA conceptualized as a special and/or issue-specific argumentation form in educational research?

(5) Is SSA measured by SSA-specific instruments and/or SSI-specific instruments?

(6) Are the methods used to assess SSA consistent with the conceptualizations of SSA?

The answers to these questions will provide researchers and readers with greater clarity on the construct of SSA. At the same time the results of Study 2 will contribute to current discussions regarding the nature of SSA, not only at a theoretical, but also at a methodological level. With regard to the analysis of the applied operationalization processes, this study will provide recommendations on how these processes can be transparent by following specific methodological steps.

2. STUDY 1- TOWARDS CONCEPTUALISATION AND MEASUREMENT OF SOCIO-SCIENTIFIC ARGUMENTATION: A LITERATURE REVIEW

2.1 Introduction

Rapid advances in science and technology can give rise to controversial issues which may become topics for public debates. Climate change, vaccines, genetically modified foods and designer babies are some examples of controversial issues that are frequently presented by the media. These controversial scientific topics with an impact on society have been called socio-scientific issues (SSI) (Sadler, 2004). The negotiation of these complex issues can be a challenging process, as it involves not only the consideration of scientific evidence, but also relevant moral, political and social aspects of the topic (Fang, Hsu, & Lin, 2019).

To address these issues, the concept of socio-scientific argumentation (SSA) was introduced in science education classrooms (Sadler, Chambers, & Zeidler, 2002). As a result, teaching in SSI-contexts and socio-scientific argumentation have become a part of curricula in countries such as Sweden (Christenson, Chang Rundgren, & Hoglund, 2012), the United States of America (Next Generation Science Standards [NGSS], 2013) and Germany (Kultusministerkonferenz [KMK], 2004) in order to promote citizenship education and 21st century skills in the context of science education. The recommendations to teach in SSI-contexts are also in line with the trending Vision II for scientific literacy, according to which the focus of science education should be shifted towards 'real-life' problems (Sadler & Zeidler, 2009). This shift is also highlighted by the most recent PISA assessment (PISA 2015), which focused on scientific literacy skills aimed at students' 'preparedness for life' (Organisation for Economic Co-operation and Development [OECD], 2017, p. 20).

In educational research, the area of socio-scientific argumentation has been relatively newly introduced (Sadler et al., 2002), and most of the publications tackling the issue emerged within the last decade as a result of a growing research interest in the topic of science education (Tekin, Aslan, & Yilmaz, 2016). This growing body of literature oftentimes provided empirical evidence in support of SSA teaching, revealing gains, inter alia, in students' argumentation skills and content knowledge (Venville & Dawson, 2010). However, it is still unclear whether there is an emerging consensus among researchers on how SSA is conceptualised and measured (Nielsen, 2012). In spite of recent attempts to conceptualise concepts related to SSA (e.g. socio-scientific decision-making; Fang et al., 2019), there are no recent reviews specifically targeting the conceptualisation of SSA in particular. To that end, and given the increasing trend in publications on the topic, there is a need to review and summarise the current state of the field in terms of terminology, concepts and instruments.

To tackle this problem, we conducted an integrative literature review in order to shed some light on: (a) definitions, terms and concepts, and (b) the operationalisations and measurements that have been used to describe SSA in educational research for the period 1999–2017. In providing such a review, we hope to contribute to the further advancement of the field, in which theoretical considerations are better aligned with methodological approaches of measurement. In what follows, we will present the theoretical basis of the article, the methods and conclusions, as well as potential guidelines for future research.

2.1.1 Argumentation in science education

Argumentation has been one of the most-cited topics in science education (Lee, Wu, & Tsai, 2009; Lin, Lin, & Tsai, 2014) and has been traditionally associated with logical reasoning and evidence-based claims (Faize, Husain, & Nisar, 2017). According to this perspective, science education should promote scientific argumentation in the form of a rational discourse between individuals who intend to make connections between evidence and ideas (Duschl, Scweingruber, & Shouse, 2007).

Nevertheless, within the last twenty years, research interest in science education has shifted from formal reasoning and scientific argumentation towards informal reasoning and socio-scientific argumentation, respectively (Erduran, Ozdem, & Park, 2015). Following the principles of science-technology-society (STS) (Yager, 1996), the SSI movement manifested the importance of discourse regarding open-ended scientific and socially relevant issues in science classrooms (Sadler, 2011). Subsequently, a number of studies suggested that teaching in SSI contexts facilitates students' argumentation skills (Zohar & Nemet, 2002; Venville & Dawson, 2010;), understanding of scientific knowledge (Walker & Zeidler, 2007) and the nature of science views (NOS) (Khishfe, Alshaya, BouJaoude, Mansour, & Alrudiyan, 2017).

2.1.2 Socio-scientific issues

Socio-scientific issues (SSI) are mostly defined as typically contentious, open-ended, ill-structured problems (Sadler, 2004; Sadler & Zeidler, 2005) which refer to the interactive relationship between science and society (Kolstø, 2001; Zeidler, Walker, Ackett, & Simmons, 2002). Moreover, they are characterized as 'societal dilemmas with conceptual, procedural, or technological links to science' (Sadler & Zeidler, 2004, p. 5). Thus, SSI introduce ethical considerations of science that stimulate students' moral reasoning and emotional development (Zeidler et al., 2002). Examples of socio-scientific issues are climate change, cloning, genetically modified foods and the use of vaccines (Zeidler & Keefer, 2003).

2.1.3 Socio-scientific argumentation

Terms and definitions. To describe the process of producing and analysing argumentation in SSI-contexts, researchers use a variety of terms, such as SSI argumentation (Rundgren, Eriksson, & Chang Rundgren, 2016; Sadler & Donnelly, 2006), argumentation regarding SSI (Christenson, Chang Rundgren, & Zeidler, 2014) or socio-scientific argumentation (SSA) (Jonsson, 2016; Kutluca & Aydin, 2016). According to Sadler and Donnelly (2006), SSA involves the negotiation of ill-structured problems. This definition implies that since SSI are open-ended problems, when individuals argue about SSI they should demonstrate perspective-taking skills (Zeidler, Herman, Ruzek, Linder, & Lin, 2013). At the same time, since SSI involve scientific and social aspects, during the process of SSA individuals should be able to coordinate scientific and non-scientific evidence by prioritizing between economic, political and moral considerations (Grooms, Sampson, & Golden, 2014). In addition to these challenges, some researchers stress the role of personal experiences and values in the negotiation of such ill-structured issues (Eggert, Nitsch, Boone, Nueckles, & Boegeholz, *2017).*

Concepts related to SSA. The concept of SSA is often described as analogous to other theoretically related concepts. For instance, socio-scientific argumentation is frequently linked to informal reasoning; a number of researchers claim that SSA is the expression of informal reasoning about SSI (Christenson & Chang Rundgren, 2015). According to this claim, informal reasoning is characterised by complexity and no clear-cut solutions, while arguers need to weigh pros and cons and make decisions based on available information (Sadler, 2004; Sadler & Zeidler, 2005). In addition, informal reasoning ‘assumes importance when the problems are more open-ended, debatable, complex or ill-structured, and especially when the issue requires that the individual build an argument to support a claim’ (Means & Voss, 1996, p. 140). Hence, since SSI are open-ended, debatable, complex and ill-structured problems by definition (Sadler, 2004), researchers regard SSA as a case of informal reasoning (Christenson & Chang Rundgren, 2015).

According to other approaches, reasoning on SSI moves beyond the conceptual scope of informal reasoning. Therefore, the SSI-related reasoning process has been termed as socio-scientific reasoning (SSR) (Sadler & Donnelly, 2006). In a recent paper, Romine, Sadler and

Kinslow (2017) distinguished between naïve and target performances in SSR based on whether they: (i) recognize the complexity of SSI; (ii) demonstrate perspective-taking skills; (iii) recognize that SSI are subject to ongoing inquiry; and (iv) treat information from various sources with scepticism (Romine et al., 2017).

Furthermore, since SSA is linked with the use and evaluation of scientific evidence, it is often considered a part of scientific literacy. More explicitly, Sadler (2011) emphasised the connection between SSA and Vision II for scientific literacy (Roberts, 2007), according to which the emphasis of scientific skills in science education should be on more contextualised issues. Based on this claim, SSI can be regarded as problems from students' everyday lives (Christenson et al., 2014); thus, SSA can be considered as an opportunity to use scientific literacy skills in the context of everyday life. Similarly, SSA has been conceptually connected to citizenship education, as citizens are expected to be able to participate in public debates (such as SSI) and in decision-making (Barrue & Albe, 2013).

Measurement of SSA. Researchers use a variety of quality criteria and argument models to examine SSA. According to Christenson and Chang Rundgren (2015), one of the most common quality indicators for SSA are the structure and the components of the argument (Chang & Chiu, 2008; Chang Rundgren & Rundgren, 2010; Erduran, Simon, & Osborne, 2004). More explicitly, components such as justifications after claims or the production of counter-arguments (counter-positions) are considered of high importance (Tal & Kedmi, 2006; Wu & Tsai, 2007; Zohar & Nemet, 2002). Other studies examine the epistemic rather than the structural components of arguments. In this case, the content of the argument is examined with regard to the appropriateness or correctness of the data and the content knowledge that the participants use for backing their claims (Sadler & Fowler, 2006; Tal & Kedmi, 2006; Zohar & Nemet, 2002). Other quality indicators that have been pointed out are the inclusion of multiple perspectives in the form of counter-arguments (Chang &

Chiu, 2008; Chang Rundgren & Rundgren, 2010; Wu &Tsai, 2007) or the inclusion of personal values in argument formulation (Chang Rundgren & Rundgren, 2010; Sadler & Donnelly, 2006).

A variety of models have been used to measure SSA. A number of studies have applied the original or adapted versions of Toulmin's Argumentation Pattern (TAP) (Toulmin, 1958) to assess argument structure (Sadler & Fowler, 2006). Moreover, models addressing the dialogical features of SSA such as Kuhn's argument moves (Kuhn, 1991) (Iordanou & Constantinou, 2015), Lakatos' scientific programs (Lakatos, 1978) (Chang & Chiu, 2008) or Walton's argumentation scheme (Walton, 1996) (Nam & Chen, 2017) have been used. Nevertheless, several researchers have developed new models and frameworks to address specific features of socio-scientific argumentation such as the multiple sources of evidence and the reasoning modes demonstrated during SSA. Examples of such models are the SEE-SEP model (Chang Rundgren & Rundgren, 2010), which was developed to evaluate the sources of participants' arguments (knowledge, values and personal experiences) and a framework developed by Sadler and Zeidler (2005) describing the three informal reasoning modes (rationalistic, emotive and intuitive) which take place during SSA.

2.2 Research Questions

Within the last 15 years, the importance of socio-scientific argumentation has been emphasized by national curricula and international assessments. However, there is no clear conceptualisation of the construct, nor are there clear recommendations for how SSA should be assessed (Erduran et al., 2015; Nielsen, 2012). To address these issues, we conducted an integrative literature review in order to present an overview of the publications on the topic to date.

Specifically, the aim of this study is to present a taxonomy of terms, definitions, assessment criteria and models which have been used to describe and assess SSA. By doing so, we seek to provide a cohesive framework regarding SSA conceptualisations and measurement, which can serve as a common basis for future research on the topic of SSA.

Our research questions were formulated with respect to the following three main foci:

(7) How did SSA research develop in scope and publications in the last decade?

(8) How is SSA theoretically framed in recent publications? What terminology, definitions and concepts are used to describe SSA in science education research?

(9) What methodological approaches are applied to empirically assess SSA?

- What approaches of operationalization are pursued?
- What instruments are used?
- What argument criteria are applied?

2.3 Materials and Methods

2.3.1 Literature search methods and selection process

To answer the research questions an integrative literature review was conducted. Two databases were used for the literature search: Web of Science (all databases) and EBSCOhost Library (ERIC and PsycArticles). The search string contained three main keywords: argumentation, social (and its derivatives) and scientific (and its derivatives). The keyword argumentation was used without its derivatives, due to a number of studies which used 'argument' as disagreement or conflict. In addition, since we were interested in studies which measured SSA, we used combinations of the following keywords: ability, skill, competence, quality, outcome, performance, study, empirical, intervention, experiment, observation, test, and questionnaire. Furthermore, to include studies in educational settings the keywords teacher, student and classroom were used. No restrictions were applied, such as language and year of publication. The two-phase search resulted in 638 hits after duplicates were removed.

Subsequently, titles and abstracts were screened for the terms ‘argumentation’ and/or ‘reasoning’ and the term ‘socio(-)scientific’. In addition, we excluded non-peer-reviewed articles (such as dissertations), theoretical articles (such as literature reviews) and articles written in languages other than English, as terms and definitions could vary across languages. We also included studies in which even though the term ‘socio(-)scientific’ was not mentioned, authors evaluated argumentation or reasoning regarding an SSI. This decision was made based on whether the topic under discussion was – according to the definition of SSI –“ill-structured, open-ended that emerge from science and have a potential impact on society” (Sadler & Zeidler, 2005). After applying these criteria, 138 studies in total were found eligible. Finally, a full-text search was conducted to examine whether the included studies mentioned the terms ‘reasoning’ or ‘argumentation’ in their research questions and whether a type of argument assessment was mentioned in their methods section. A total number of 67 studies were included. The study selection process is presented in Figure 5.

Since the goal of this study is to provide an overview for the current ways of conceptualising and measuring SSA, our aim was to conduct an exhaustive literature review. Despite the fact that we used the two of the most commonly used databases in educational research, we are aware that there might still be missing publications relevant to SSA. Nevertheless, we believe that this search is a representative collection of publications in the area of SSA between 1999 and 2017.

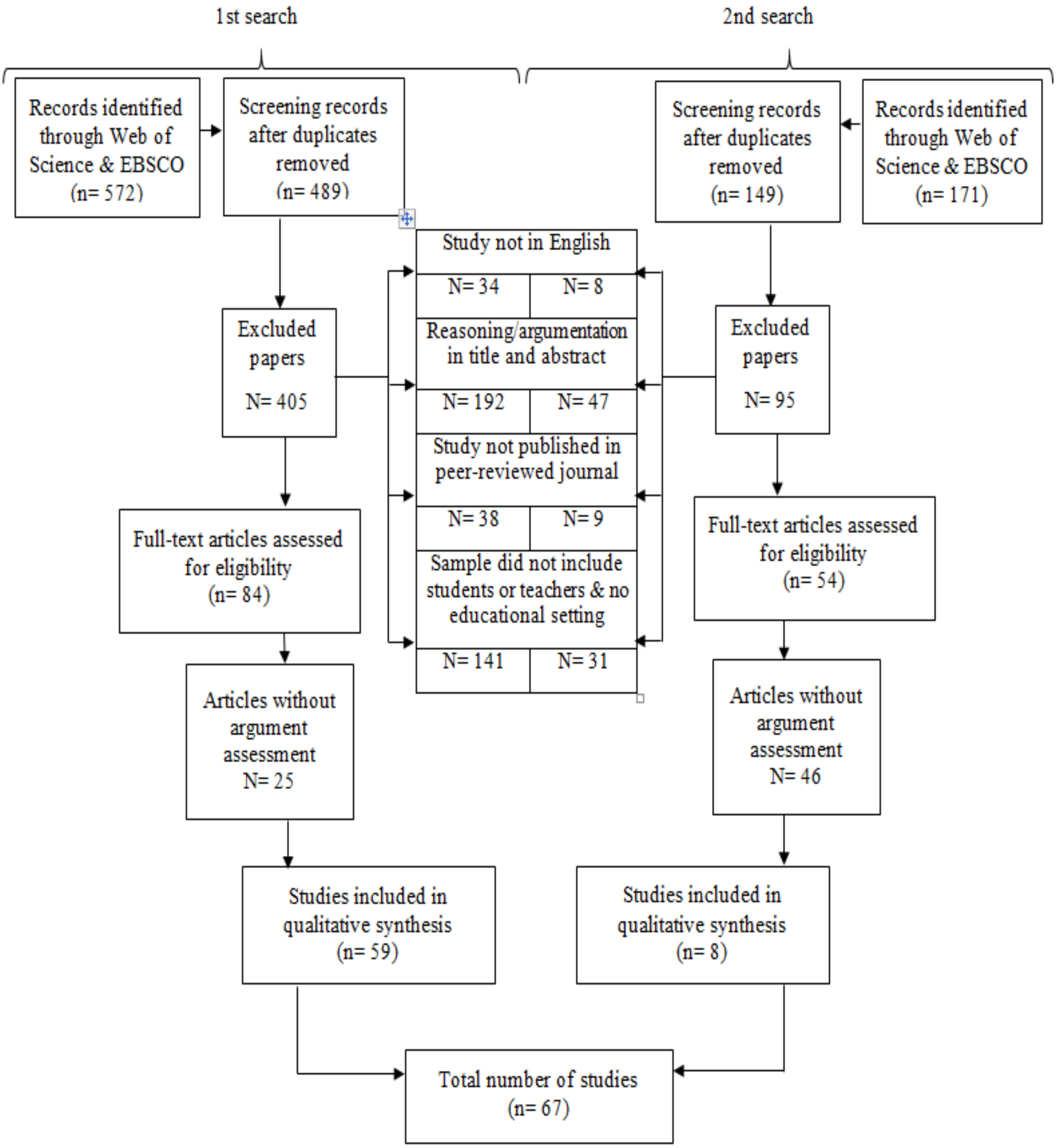

Figure 5

Study Selection Flowchart

2.3.2 Study analysis

Development of the coding scheme. After the sampling phase, a coding scheme was developed to investigate two aspects of SSA, conceptualisation and measurement.

Conceptualisation of SSA. The aspect of 'conceptualisation' referred to our first research question regarding the terms, definitions and concepts related to SSA. For this aspect, the introduction and theoretical sections of the articles (theoretical background, rationale, problem statement, etc.) were coded. The codes emerged in a bottom-up, iterative process following the principles of grounded theory (Strauss & Corbin, 1998) and content analysis (Krippendorff, 2004).

In the first coding phase, a set of ten articles was randomly selected and assessed in an open-coding procedure. Keywords relevant to the terms, definitions and characteristics of SSA, as well as SSA-related concepts were gathered. The gathered keywords produced 21 codes, which were grouped into six themes. Based on these codes and the corresponding themes, a second coding phase was conducted for a second randomly selected set of ten studies during which axial and open-coding was conducted (Strauss & Corbin, 1998). This phase resulted in 23 new codes, 15 of which were clustered into two additional themes. Finally, in the third coding phase (n=30), 15 new codes were added to the corresponding themes. The final version of the coding scheme for conceptualisation of SSA consisted of eight themes and 59 codes. For reasons of clarity and comprehensibility, a number of codes were clustered into post hoc categories. All the selected publications (N=67) were then coded based on this final version of the coding scheme.

Measurement of SSA. The aspect of 'measurement' referred to our second research question regarding operationalisations, instruments and argument assessment criteria of SSA. The methodological sections of the articles were coded for this aspect of SSA. The codes emerged using inductive and deductive approaches, as a coding scheme already existed and was adapted for this study (Rapanta et al., 2013).

The adaptation was conducted in three phases, as presented in Figure 6. In the first two phases, the selected articles were coded based on the initial coding scheme. During these phases, the adaptation occurred by constantly comparing the existing themes, categories and codes with the ones presented in the selected articles. At the same time, in phases two and three, we conducted open-coding in order to identify new codes and themes. After the process of adaptation, the final version of the coding scheme for this aspect included two themes, 13 categories and 55 codes. The total number of selected publications (N=67) was coded according to the final coding scheme for measurement of SSA.

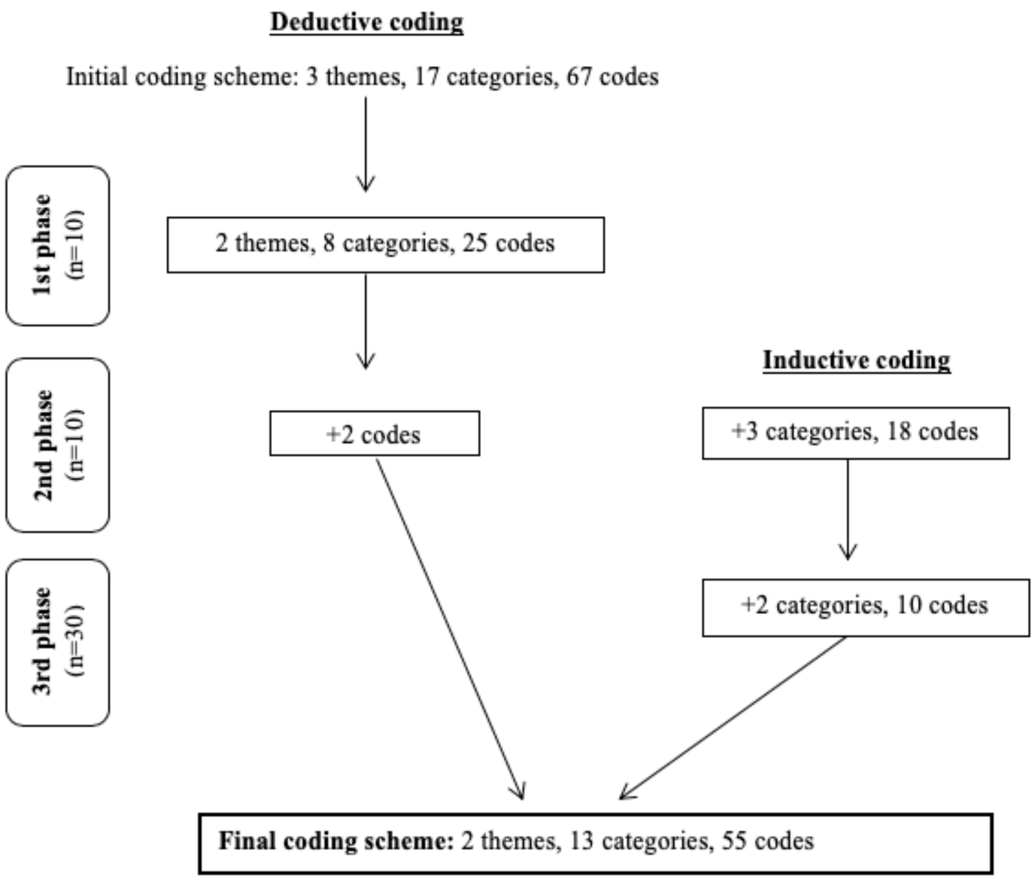

Figure 6

Adaptation Process of Coding Scheme by Rapanta et al. (2013)

After the development of the coding scheme, two independent raters coded 20% (n = 14) of the overall selected articles. Subsequently, inter-rater reliability was calculated and Cohen's kappa ranged between K = 0.70 and K = 0.89; the few remaining discrepancies were resolved through discussion. A more detailed description of the coding scheme is presented in the following section.

Description of the coding scheme. As outlined above, we developed the coding scheme to describe two main aspects of SSA, the way in which it has been conceptualised and measured in the selected articles. In our coding scheme, these two aspects are further described with respect to ten themes, 22 categories and 114 overall codes (see Table 6 & Appendix).

Within the aspect Conceptualisation, the themes 'Terms' and 'Definitions' aimed to code both the terms that have been used to describe SSA as well as the definitions for the constructs of socio-scientific issues, reasoning and argumentation. We grouped features of SSA that distinguish the construct from other, conceptually connected constructs under the theme 'Specific characteristics'; general features were categorized as 'General characteristics' of SSA. As 'Conceptually related concepts', we clustered general concepts often used to conceptually frame SSA; and as 'Empirically linked variables', we labelled codes referring to empirically tested relationships between SSA and other variables. The theme 'Quality indicators' described elements of high-quality argumentation, and for the theme 'Assessment models', we grouped codes referring to argument assessment models. For 'Measurement', we coded: 'Study characteristics', such as the sample size and type of participants; and 'Argument assessment', which included argument assessment criteria and models, as well as quality indicators. All the themes and categories contained in our coding scheme are displayed in Appendix.

Coding rules. Out of 114 codes, 99 were coded as *in vivo* codes, meaning that the code was identical or nearly identical to the coded segment. The remaining codes were keyword-based and open codes. Open codes were coded with the selection of relevant text segments following specific coding rules. For example, with regard to 'Definitions', in order to be coded, the text segment had to follow the sentence structure: '[Concept] is/ is defined/framed/described'. All keyword-based codes were coded by examining the existence of specific keywords in the segment, with the exception of the category 'argument assessment criteria'. The coding for this category was conducted by means of code descriptions, as presented in Table 3.

Table 3

Description of Codes for Argument Assessment Criteria

Code	**Description**
Structure	Evaluation of the existence of argument components (e.g. claims, justifications and rebuttals)
Use of pre-defined argument criteria	Evaluation of existence of argument criteria taught in an intervention (e.g. Toulmin's argumentation pattern)
Complexity	Evaluation of quality by counting the number of justifications/reasons
Multidimensionality	Evaluation of quality by counting the number of different sources of evidence (e.g. scientific, moral, social, political)
Use of scientific evidence	Evaluation of existence of scientific evidence
Accuracy	Evaluation of accuracy of evidence
Coherence	Evaluation of the logic connection between argument components
Argument moves	Evaluation of dialogic features of arguments (strengthen one position, weaken another)

2.4 Results

The findings of this study are presented in three sections: (a) development of SSA research, (b) conceptualisation of SSA, and (c) measurement of SSA.

2.4.1 Development of SSA research

The publication years for our sample studies ranged between 1999 and 2017 and the majority of the articles (74%) were published after 2011 (Figure 7). Moreover, the number of publications shows two peaks; the first in 2014, and the second in 2017, respectively, as displayed in Figure 7.

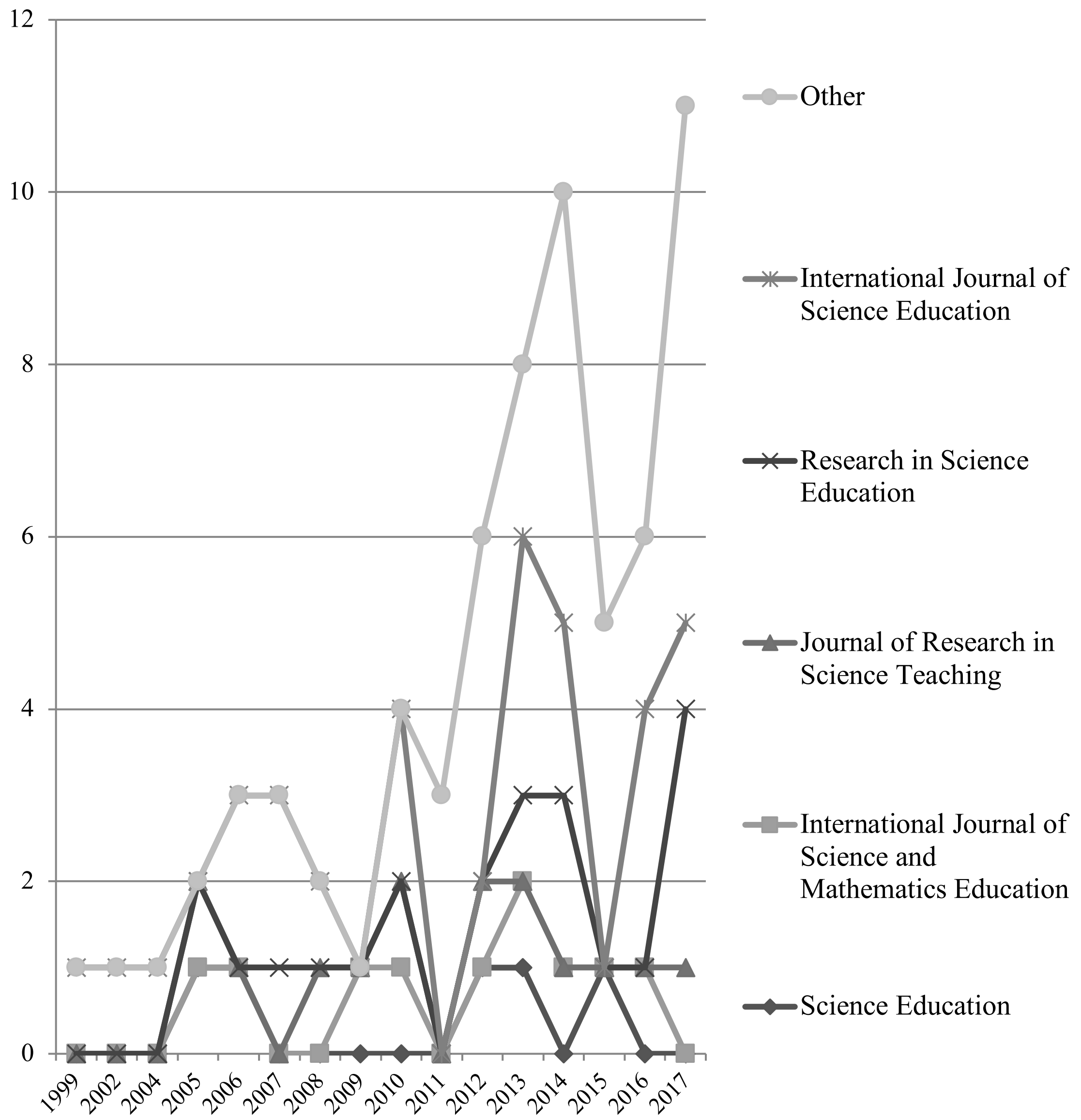

Figure 7

Publication Trends by Journal

In addition, the selected studies were published in 21 journals, out of which the International Journal of Science Education (IJSE), Research in Science Education (RSE), Journal of Research in Science Teaching (JRST), International Journal of Science and Mathematics Education (IJME) and Science Education (SE) published 61% of the sample studies overall. The number of publications per journal is presented in Table 4. Figure 7 displays the number of publications per year by journal, revealing an increase in publications by journals other than the aforementioned, especially after 2011.

Table 4

Paper Distribution by Journal (N=67)

Journal	**Publications (n)**
International Journal of Science Education (IJSE)	19
Research in Science Education (RSE)	7
Journal of Research in Science Teaching (JRST)	6
International Journal of Science and Mathematics Education (IJME)	5
Science Education (SE)	4
Eurasia Journal of Mathematics Science and Technology Education (EJMSTE)	3
Journal of Biological Education (JBE)	3
Cultural Studies of Science Education (CSSE)	2
Journal of Education in Science, Environment and Health (JESEH)	2
Journal of Science Education and Technology (JSET)	2
Teaching and Teacher Education (TTE)	2
Curriculum Journal (CJ)	1
Educational Research and Reviews (ERR)	1
Instructional Science (IS)	1
International Journal of Advanced and Applied Sciences (IJAA)	1
International Journal of Environmental and Science Education (IJESE)	1
Journal of Baltic Science Education (JBSE)	1
Journal of Science Teacher Education (JSTE)	1
Science & Education (S&E)	1

Table 4 (continued)

Paper Distribution by Journal (N=67)

Journal	**Publications (n)**
Science Education International (SEI)	1
Turkish Online Journal of Educational Technology (TOJET)	1
Chemistry Education Research and Practice (CERP)	1
Computers & Education	1

2.4.2 Conceptualisation of SSA

For reviewing the conceptualisation of SSA a mixed-methods approach was followed. *In vivo* and keyword-based codes were quantified and frequencies were calculated, whereas the open codes were analysed qualitatively (Creswell & Miller, 2000). In the following section, the results of our analysis are categorized with regard to the eight related themes and the codes with the highest frequencies are presented.

Definitions and terms. Regarding the definitions of SSA, out of 67 studies, 18 provided an explicit definition for the construct of SSA; while 10 included a definition for SSR and 18 included a definition for SSI in the format: 'SSA/SSR/SSI is/ is defined/framed/described'. Our qualitative analysis of definitions resulted in seven themes which are illustrated in Table 5.

With regard to terms, our qualitative analysis revealed that the terms socio-scientific argumentation, argumentation on/about SSI, SSI argumentation, informal reasoning and informal argumentation were most frequently used to describe SSA. The terms were also used interchangeably: "SSI argumentation, also termed informal reasoning" (Christenson & Chang Rundgren, 2015, p. 2). Socio-scientific reasoning was also termed as informal reasoning (Wu & Tsai, 2007; Wu, 2013). In addition, authors used the term socio-scientific issues or SSI, with a few exceptions, such as "Socially Acute Questions" (Simonneaux et al., 2013).

Table 5

Emerged Themes by Definitions of SSA, SSR & SSI

Concept	**Theme**[2]
Socio-scientific argumentation	SSA is argumentation regarding SSI
	SSA is the negotiation and decision-making to solve an SSI
	SSA is the expression of informal reasoning regarding SSI
Socio-scientific reasoning	SSR is the thinking process that takes place during negotiation of SSI
	SSR is informal reasoning regarding SSI
Socio-scientific issues	SSI are ill-structured, open-ended issues that emerge from science and have a potential impact on society
	SSI are contentious dilemmas with no clear-cut solutions

Specific and general characteristics. Specific characteristics of SSA were reported in several of the papers of our sample. The most frequently mentioned specific characteristic was controversy, which was mentioned by 70% of the articles. In addition, uncertainty was reported by 42% and authenticity by 30% of the articles. Regarding the specific aspects which arguers take into consideration during SSA, authors reported moral (38%), social (34%), political (22%), economic (21%), scientific (13%) and personal (13%) issues. SSA was characterized as a form of debate in 27% of the articles, and 21% of the studies highlighted the importance of perspective-taking skills while arguing on a socio-scientific topic. Media coverage was also presented as one of the distinguishing features of SSA (12%). The general characteristics of SSA included complexity (24%), knowledge integration (18%) and use of evidence (18%).

Conceptually and empirically related variables. As conceptually related to SSA, the concepts of decision-making (65%), scientific literacy (58%) and civic competencies (55%) were most frequently coded. Moreover, informal reasoning (45%) was used as a term not only to define SSA (as in section *Definitions and terms*), but also to theoretically frame SSA. Critical thinking (24%) and the Science-Technology-Society movement (21%) were also among the most frequently mentioned concepts. Concepts such as problem-solving, Vision II for scientific literacy and higher-order thinking skills were mentioned by fewer studies (≤ 10%).

Among the empirically related variables, argumentation skills (16%) and knowledge (14%) were mentioned as outcome variables of interventions involving SSA tasks. In addition, views on nature of science (NOS) (9%) were presented as a variable with reciprocal effects to SSA. Knowledge was mentioned (22%) as a predicting variable for SSA skills, as well as the arguer's personal values (13%).

Quality Indicators and argument assessment models. The most frequently coded quality indicators were the use of evidence (19%), the use of justifications (14%) and rebuttals (12%). However, 39% of the articles mentioned quality indicators categorized as 'Other', including a variety of indicators such as the use of counter-arguments (10%) and the number of reasons to support an argument (6%). 'In terms of argument assessment models, 19 out of 67 articles mentioned Toulmin's TAP, 6 of which criticized the model based on the assumption: "it (TAP) does not address the level of scientific knowledge or the number of aspects used to construct arguments." (Namdar, 2017, p. 137). Kuhn's model of argument moves (15%) was among the most frequently mentioned argument assessment models. A variety of other argument models were also coded as "Other" (24%), such as Lakatos' research programs (≤ 10%) and Duschl's argumentation model (≤ 10%).

2.4.3 Measurement of SSA

The measurement of SSA was coded and analysed with respect to two themes: (a) study characteristics, and (b) methods of argument assessment. In the following section, the codes with the highest frequencies are presented.

Study characteristics. The type of participant and sample size was our focus at first. Out of 67 articles, 54% used secondary school students as a sample, although the sample size varied among studies (M=150.13, SD= 200.99; excluding case studies). University students were also used as a sample group (25%), as well as teachers (16%), while only one study used primary school students as a sample. With regard to study design, almost half of the studies used descriptive designs, fewer studies included interventions and finally, a smaller number of studies were case studies, as presented in Figure 8.

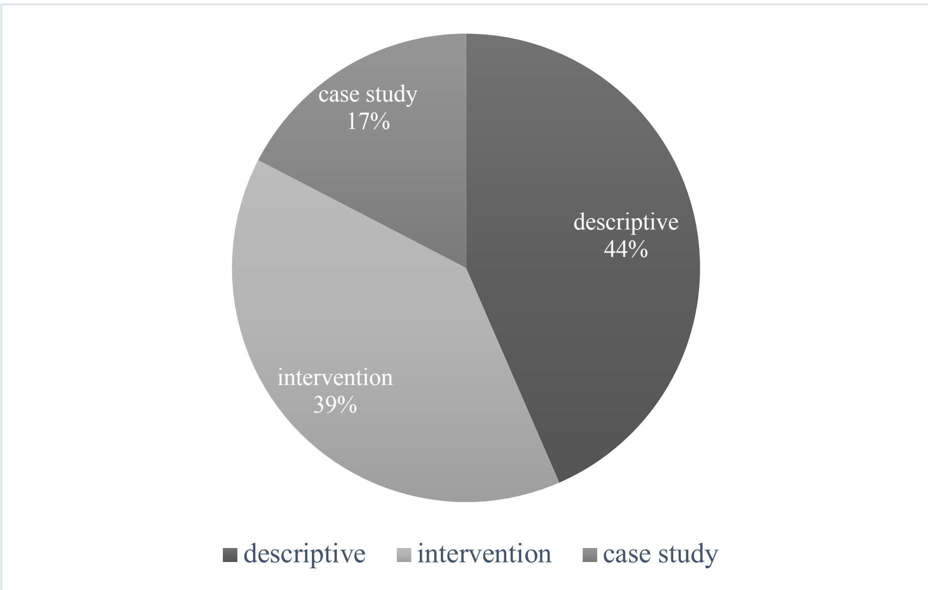

Figure 8

Percentages of Studies by Study Design

Regarding the context of measuring SSA, both discussions (31%) and written arguments (29%) were the most frequent argumentation tasks. Written arguments were mostly answers to open-ended scenarios presenting an SSI topic. Questionnaires (18%) and interviews (18%) were also used by a number of studies, as were tests (4%).

The majority of the studies (85%) used qualitative analysis to measure SSA; most (60%) quantified their qualitative data. A small number of studies used mixed methods (5%) and quantitative analysis (3%). The socio-scientific topics used in argumentation tasks included environmental issues such as climate change, as well as biotechnology issues, such as stem cell therapy. Other topics presented were: genetically modified foods, nuclear energy, eating habits and smoking, as well as animal extinction. Under the category ‘Other’, topics such as human sexuality and the colonization of Mars were included.

Methods of argument assessment. Out of 67 studies, 62 used at least one argument assessment criterion in order to assess the quality of SSA. These criteria were often presented in scoring rubrics and coding schemes. The remaining five were studies that assessed SSA through tests and questionnaires and thus, did not describe their argument assessment criteria. Furthermore, 13 out of 67 studies used reasoning assessment criteria, eight of which assessed reasoning modes. The frequencies and percentages of argument assessment criteria used in our sample studies are illustrated in Figure 9.

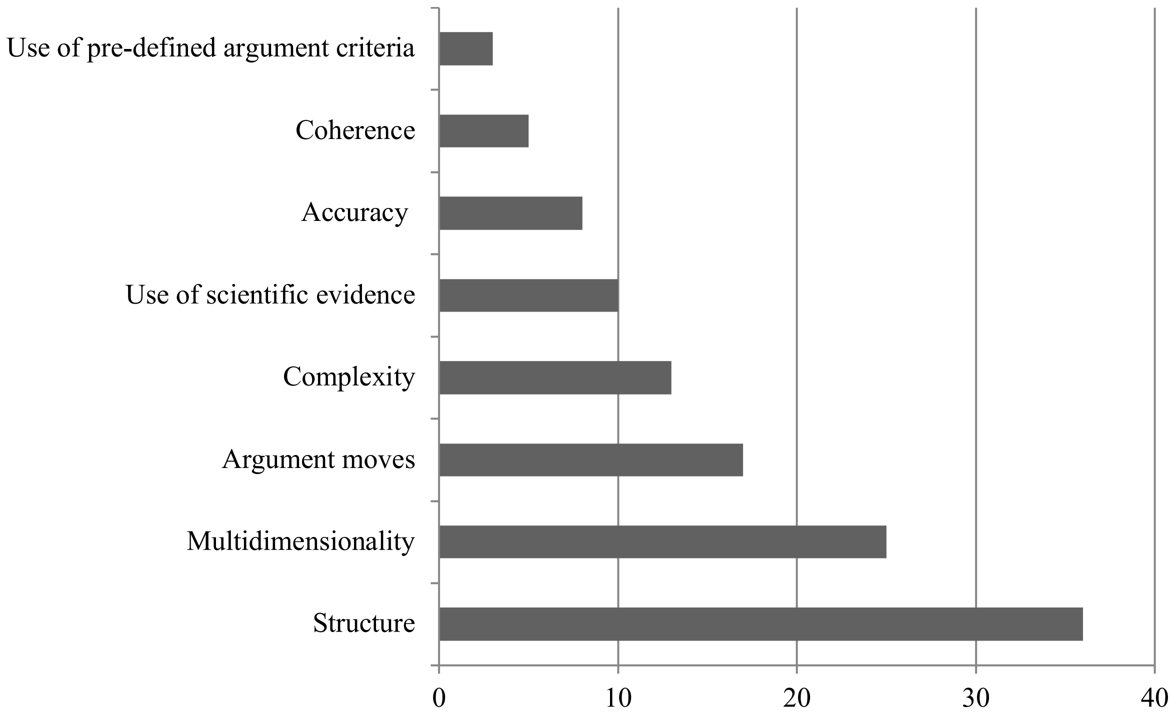

Figure 9

Frequencies and Percentages of Argument Assessment Criteria Used in the Selected Studies (N= 62)

Regarding argument assessment models, 61% of the studies used the original or the adapted version of an existing coding scheme. In addition, 29 coding schemata were cited, of which Toulmin's TAP was used with the highest frequency (18%). Other coding schemas were used in lower frequency, such as Kuhn's argument moves (6%), Walton's scheme (5%) and the SEE-SEP model (5%). In terms of quality indicators, rebuttals (15%), justifications (12%) and the use of evidence (12%) were the most frequently reported codes. Rebuttals were almost exclusively used in studies using scoring rubrics to assess argument structure.

2.5 Discussion

The purpose of this review was to summarize current research on the topic of socio-scientific argumentation in science education. Our first goal was to give a clearer idea of the concept of socio-scientific argumentation by presenting the relevant definitions, terms and characteristics. Our second goal was to present the features of a good SSI argument and possible analytical tools used to measure the quality of SSA. While addressing these goals, we were also interested in how the theoretical assumptions regarding SSA were transferred into the methodological tools within the selected articles. As in the Results section, for reasons of clarity, we discuss our main findings with regard to three aspects – the development of SSA research, conceptualisation of SSA and measurement of SSA

2.5.1 Development of SSA research

In terms of publication trends, our results suggest a gradual increase in the number of publications related to SSA, especially after 2011, with peaks in 2014 and 2017. These findings are in line with previous literature reviews in science education research. For instance, Lin and her colleagues (2014) conducted a review on research trends in science education for the period 2008–2012 in three major journals (IJSE, JRST & SE) and concluded that informal reasoning and argumentation in the context of SSI were among the most frequently investigated topics. Similarly, Erduran et al. (2015) indicated an increasing trend in studies related to argumentation in science education for the period 2006-2014, with a peak in 2014. In a more recent review on research trends on SSI, Tekin, Aslan, & Yilmaz (2016) stated that from 2004 until 2015, SSA was the most highly mentioned topic in science education journals.

At the same time, the number of journals publishing SSA-related articles has been increasing over time. This is an indication of an expanding research community; between 2014 and 2017 the number of journals other than the ones traditionally publishing SSA topics (e.g. IJSE) notably increased. This increase can be explained by the concurrent adoption of new teaching standards for science education in 2011 and 2013, which included the introduction of SSI in science education. For instance, in 2011, a new science curriculum was introduced in Sweden, which included the relationship between society and biology, chemistry and physics (Mullis et al., 2012). One year later, in 2012, the Next Generation Science Standards were released in the United States, which introduced core ideas such as the link between science and society (National Research Council, 2012). As a result, with SSI becoming a part of the new curricula, more researchers are developing an interest in assessing and fostering SSA skills in science education.

2.5.2 Conceptualisations of SSA

An adequate way to grasp the conceptualisation of an ill-defined construct, such as SSA, would be to examine the ways in which it has been defined by prior research (Allen, 2017).

Despite our attempt to summarize and present the existing definitions in an integrative way, an unexpected finding was the small number of studies which included a definition for SSA in their introduction or theoretical background. Given that all the selected studies included a measurement of SSA, it was expected that most studies would include a definition of the construct. This expectation was also based on the general guidelines for publications in educational research, according to which the definitions and conceptualisations in research articles should be included in the problem statement of the study (Institute of Education Sciences, 2013). Nevertheless, similar to our results, findings from other relevant studies suggest that researchers in the area of SSI tend to overlook definitions of the investigated constructs. For instance, a recent review on socio-scientific decision-making (Fang et al., 2019) reported that during the screening process, a large number of studies were excluded due to the fact that they did not present a definition of the construct. Nevertheless, despite the lack of definitions and the use of various terms, we summarized the provided definitions into three themes (Table 5) to provide a framework for conceptualisation of SSA.

In addition, there is an overall consensus on the conceptually related constructs and the general characteristics of SSA. Socio-scientific argumentation was generally described as argumentation regarding controversial issues, which takes place in authentic contexts (e.g. debates), and it is characterized by uncertainty. Moreover, the moral aspects of SSA are highlighted over its socio-political and scientific counterparts. This can be explained by the fact that SSA has been mainly studied in the context of science classrooms. In this case, the goal of SSI teaching is oftentimes to introduce its complex nature by presenting evidence from different sources (values, emotions, etc.), rather than by focussing on the use and quality of scientific evidence (Sadler & Donnelly, 2006).

Furthermore, SSA is generally defined by analogy to decision-making and informal reasoning, since SSA refers to 'what to do' (Nielsen, 2012) to address ill-structured, open-ended dilemmas (Sadler, 2004). The use of these concepts to define and term SSA can have two possible explanations. First, it can be explained by the fact that SSA is a newly introduced term (Sadler, 2011) and as such, it can (only) be described by other, more well-established concepts. Alternatively, it can be assumed that the majority of the studies did not use the term socio-scientific argumentation or refer to SSA's special features because they do not assume its distinctive nature.

2.5.3 Measurement of SSA

Socio-scientific argumentation has mainly been measured through qualitative analysis, by means of one or more of the eight assessment criteria (Table 5). Furthermore, more than nine argument assessment models have been applied, excluding new frameworks introduced by authors. These findings suggest that our sample studies followed various approaches for measuring SSA. This heterogeneity in measurement approaches has been also emphasized by previous reviews on argumentative competencies in education (Rapanta et al., 2013).

However, overall, SSA was frequently assessed against structural criteria and models, such as TAP. This finding agrees with another general tendency in the area of argumentation in educational research: the focus on structural rather than epistemic criteria (Christenson & Chang Rundgren, 2015; Rapanta et al., 2013). In her integrative review on argumentative competence in educational research, Rapanta et al. (2013) concluded that researchers assess argument mainly as form, and consequently, they pay closer attention to the structure of the argument rather than its conceptual or epistemic quality. This tendency has stimulated recent discussions in educational research on whether structural criteria are sufficient for assessing the dialogic features of argumentation (Nielsen, 2013; Rapanta et al., 2013). From one point of view, TAP is presented as a convenient measure, adequate to capture measurable changes, especially in large-scale assessments (Erduran et al., 2004; Nielsen, 2013). However, it is still ambiguous whether the original version of TAP can reliably grasp the controversial and complex nature of SSA (Rundgren et al., 2016).

In fact, this criticism towards structure-based criteria was also expressed by studies included in our review (Jonsson, 2016; Namdar, 2017; Venville & Dawson, 2010). Yet, what remains unclear is why, despite the criticism, a number of researchers preferred traditional structural models, such as TAP (Lin & Mintzes, 2010; Namdar, 2017), over other SSA-specific models, such as SEE-SEP (Chang Rundgren & Rundgren, 2010).

2.6 Limitations

In this section, we will discuss some possible limitations of the study to facilitate the interpretation of the results and the possibility of a replication of the study. First, our search included only studies that used ‘socio’ and ‘scientific’ among their keywords. We are therefore aware that we might have excluded studies investigating SSA which did not include the aforementioned keywords. However, this decision was justified by the scope of the study; since the aim of the study was to foster the conceptual clarity and definition of the term ‘socio-scientific argumentation’, only studies using the aforementioned keywords could serve as potential sample studies.

Another limitation of the study might be that during the selection and analysis of the studies there was no distinction between sample types (e.g. teachers and students). This was decided due to the small number of studies which used teachers as a sample. In addition, a preliminary analysis showed no patterns or differences between samples and tasks and therefore the results were presented in an aggregated way.

2.7 Conclusions and Future Directions

As socio-scientific argumentation has become a part of teaching and learning standards in science education, the research community investigating SSA is gradually growing. Our study showed general agreement among researchers on the general characteristics of SSA; SSA is an expression of informal reasoning regarding socio-scientific issues, during which arguers should take into consideration scientific and moral aspects of the issue discussed. In terms of measurement, different approaches were followed with different foci. Yet, a good SSI argument is, overall, an argument with good structure, including components such as justifications, counter-arguments, rebuttals and the use of evidence.

Despite high productivity in the research area of SSA, our review suggests that there are some conceptual and methodological gaps which must be addressed by future research. First, studies are encouraged to include definitions and clearer conceptualisations of SSA. In addition, we suggest more elaborate justification of the terminology used, as well as more detailed operationalisation of the investigated constructs. In this way, we aspire for greater consistency between the conceptualisations and the measurement of SSA.

There are two main contributions of this study in the area of SSA. The first is the systematic categorisation of existing SSA definitions into three themes as seen in Table 5: (a) SSA is argumentation regarding SSI; (b) SSA is the negotiation and decision-making to solve an SSI; and (c) SSA is the expression of informal reasoning regarding SSI.

The second is the presentation of an analytical framework regarding the eight assessment criteria of SSA: structure; use of pre-defined argument criteria; complexity; multidimensionality; use of scientific evidence; accuracy; coherence; and argument moves. These categorizations can serve as an interpretative tool for the conceptualisations and coding procedures applied in the area of SSA. More importantly, they can serve as a reference point for future researchers in attempting greater consensus and transparency.

3. STUDY 2- CRITICAL REFLECTIONS ON SOCIO-SCIENTIFIC ARGUMENTATION RESEARCH

3.1 Introduction

In science education, argumentation skills are considered as essential skills that students should acquire as a part of scientific inquiry and literacy (Erduran & Jimenez-Aleixandre, 2012; Erduran & Jimenez-Aleixandre 2007; Jimenez-Aleixandre et al. 2000; Kelly & Takao, 2002; Zohar & Nemet, 2002) By participating in argumentative activities students have the opportunity to contribute to classroom discourse and justify their claims with scientific (or non-scientific) evidence (Erduran, Ozdem & Park, 2015). For this reason, in the last decade argumentative activities have been introduced by teaching and learning standards in science education (Erduran, 2007) rendering argumentation one of the most frequently investigated and cited topics in the field of science education (Erduran, Ozdem & Park, 2015).

Although argumentation in science education has been investigated with regard to various contexts and topics (Erduran, 2007), within the last decade more emphasis has been put to informal argumentation and more specifically, on argumentation in the context of controversial scientific issues (Lin, Lin & Tsai, 2014). Controversial scientific issues, such as climate change, are oftentimes termed *socio-scientific issues* (SSI) (Lin et al., 2014; Tekin, Aslan & Yilmaz, 2016) and the form of argumentation that takes place during the negotiation of these issues is generally defined as *socio-scientific argumentation* (SSA) (Sadler, 2004).

Apart from the increase in publications regarding SSA (Tekin et al., 2016), the concept was also introduced in science curricula, as well as teaching standards (e.g. Next Generation Science Standards [NGSS], 2013). According to these standards, teachers are expected to introduce open-ended scientific issues in their classrooms, in order to highlight the relationship between science and technology, and society. At the same time, students are expected to show high quality argumentation when they argue about SSI.

However, similarly to general argumentative competencies in education (Rapanta et al., 2013), there is a variety of terms, definitions, conceptualizations, operationalizations and measurement tools currently used to assess SSA. Given the complexity of argumentative competencies and SSA in particular, previous reviews have raised concerns regarding the rationale on which researchers choose their analytical tools, as well as the role of theoretical assumptions during the process of operationalization (e.g. Erduran, 2007; Nielsen, 2012; Rapanta et al., 2013). At the same time, this heterogeneity not only raises methodological concerns about the reliability and validity of the measurement tools (e.g. Duschl, Ellenbogen, & Erduran, 1999), but it also influences the interpretation of study results in the area of SSA.

Departing from these concerns we review and summarize current empirical literature on SSA with respect to conceptuatlizations of SSA as a distinctive concept and the respective measurement tools that are used to assess SSA. This analysis lays the ground for our main focus in this paper: The analysis and critique of consistency between theory and measurement in the current literature. To that end, we draw on the paradigm of Blömeke et al. (2015), who described the conceptual and methodological challenges regarding the assessment of complex competencies in higher education (Blömeke, Gustafsson, & Shavelson, 2015). According to their view, the aim of a systematic literature review should not be to find '"the" definitions and assessment of' the concept, but to discover potential 'commonplaces' or disagreements and make suggestions to resolve them (Blömeke et al., 2015, p.4). By doing so, we aim for more clarity in the conceptualization and measurement of SSA, which will not only facilitate the interpretation of existing literature on the topic, but it will also provide theoretical and methodological insights for future research in the area of SSA.

In the following sections, we first provide an overview of current ways of conceptualizing and measuring SSA in science education research. We then discuss the importance of clarity in constructing conceptualizations, as well as operationalizations in educational research.

3.1.1 Conceptualizing socio-scientific argumentation

Despite the heterogeneity in terms and definitions, there is relative agreement among researchers regarding the general features of SSI and SSA, as well as the reasoning process involved (Ioannidou et al, manuscript in preparation). Thus, in what follows, a short review of (a) the concept of SSI, (b) the concept of informal reasoning regarding SSI, and (c) special features of SSA is presented.

Socio-scientific issues. Socio-scientific argumentation is generally defined as the argumentation that takes place during the negotiation of SSI (Sadler & Donnelly, 2006; Sadler & Zeidler, 2005). Subsequently, SSI are defined as ill-structured, open-ended problems that refer to the interactive relationship between science and society (Kolstø, 2001; Zeidler et al., 2002). Examples of SSI are climate change, cloning, genetically modified food and the use of vaccines (Zeidler, & Keefer, 2003). In science education they are regarded as the extension of the Science- Technology- Society movement (STS) (e.g. Christenson et al., 2014), as they raise issues of ethics in science and technology.

Since these topics often become topics of public debates presented by media (e.g. Craig-Haare et al., 2017), they are often introduced as decision-making tasks in science classrooms. A typical example of an SSA task includes a short introduction to the topic followed by a question; for instance: 'Do you agree that GMO should be produced and sold? Please write down your opinion as clearly as possible, and try to make your arguments the best you can.' (Christenson, Chang Rundgren, & Zeidler 2014, p.587). Students are usually asked to produce arguments for or against a decision followed by an explanation or a supporting reason.

Informal reasoning regarding SSI. Socio-scientific argumentation is commonly presented as an expression of *informal reasoning* regarding controversial scientific topics (e.g. Feierabend, Stuckey, Nienaber, & Eilks, 2012; Jonsson, 2016). Since SSI are complex and questionable issues, they could serve as ideal contexts for developing informal reasoning skills (Kuhn, 1993; Sadler, 2004). Similarly to informal reasoning, when individuals reason on SSI, they have (Eggert, Nitsch, Boone, Nueckles, & Boegeholz, 2017) to consider various possible solutions through inductive reasoning, as well as personal attitudes and opinions (e.g. Eggert et al., 2017). In addition, SSI possess one of the most essential characteristics of informal reasoning: open-endedness; an individual's decision can always be 'right', as long as it is justified by scientific, social, political, economic or ethical aspects (Sadler & Zeidler, 2009).

According to some researchers, this reasoning process is a distinct construct which is termed *socio-scientific reasoning (SSR)* (e.g. Romine et al., 2017; Sadler et al., 2007; Wu, 2013). Based on this claim, the quality of an individual's SSR skills can be assessed upon their ability to: (a) recognize the inherent complexity of SSI, (b) examine issues from multiple perspectives, (c) appreciate that SSIs are subject to ongoing inquiry, and (d) exhibit skepticism when presented potentially biased information (Sadler et al., 2007).

Features of socio-scientific argumentation. SSI are often presented as ideal settings for facilitating student argumentation (e.g. Agell et al. 2015), as they provide the opportunity for students to construct claims, justifications and counter-arguments (e.g. Isbilir, Cakiroglu, & Ertepinar, 2014; Khishfe, Alshaya, BouJaoude, Mansour, & Alrudiyan, 2017). Thus, SSI are introduced in science classrooms as debatable topics on which students have to take decisions (e.g. Christenson et al., 2012; Khishfe, 2012). During these discussions, students are exposed and are able to understand multiple perspectives (e.g. Khishfe, 2012). Therefore, by being exposed to multiple perspectives, students are more likely to integrate in their arguments evidence from various aspects of SSI under discussion including political (e.g. Christenson et al., 2012), social (e.g. Hansson, Redfors, & Rosberg, 2011), economic (e.g. Grooms et al., 2014), scientific (e.g. Wu, 2013), personal (Lin et al., 2012) and ethical aspects (e.g. Lindahl & Linder, 2013).

Although these aspects of SSA conceptualization are commonly accepted in the research areas of SSA, the applied measurement approaches differ significantly among studies. In the next section we present an overview of these approaches.

3.1.2 Measuring socio-scientific argumentation

In empirical educational research theoretical concepts are translated into tasks and procedures aiming at the access of more information about its nature through the process of operationalization (Foschi, 2014). Hence, after describing the current status of SSA conceptualization, in this section we give an overview of the ways in which SSA has been operationalized and measured in science education.

There is a wide range of methods and quality criteria used by researchers in order to identify and measure quality in SSA. However, most studies analyze SSA via qualitative data (Tekin et al., 2016) and especially through interviews (Kolsto, 2006; Sadler & Donnelly, 2006) or class discussions (e.g. Albe, 2008). Thus, researchers usually use scoring rubrics to assess the quality of oral or written argumentation with different foci. These foci are discussed with respect to quality indicators, argument assessment criteria and models

(Ioannidou et al, manuscript in preparation), which will be presented in the following sections.

Quality indicators. The assessment of SSA is often based on specific structurally or content-related elements of argumentation, which function as indicators of high quality argumentation (Christenson & Chang Rundgren, 2014). For instance, among the most commonly used structurally-oriented quality indicators for SSA are the use of justifications of claims (e.g. Lin & Mintzes, 2010), as well as the used of rebuttals (e.g. Chang Rundgren, 2013) and counter-arguments (e.g. Wu & Tsai, 2007). With regard to content-related quality indicators, the use of correct evidence (or content knowledge) is considered a feature of high quality argumentation (e.g. Zohar & Nemet, 2002).

Argument assessment criteria and models. Although quality indicators were used as 'signs' of high quality argumentation, SSA has also been assessed by overall criteria presented in the applied scoring rubrics. Thus, the difference between quality indicators and argument assessment criteria is that the former have been used as additional 'points' in the scoring rubrics, whereas the latter were used as the overall attribute of SSA quality.

The argument criteria oftentimes derive from argument assessment models, which are used in their original of adapted versions (e.g. Toulmin's TAP). Moreover, some studies use already existing? schemes and quality criteria developed for measuring other concepts (SSA-general criteria), such as scientific argumentation or argumentation (e.g. Dawson & Carson, 2017), whereas other studies developed new ones in order to assess the specific characteristics of SSA (SSA-specific criteria) (e.g. Romine et al., 2017).

An example of SSA-general argument criterion is the structure of arguments, as it is considered of great importance when assessing general argumentative competencies (Rapanta et al., 2013). This criterion is also frequently used in the assessment of SSA; to measure the quality of argumentation researchers evaluate the length, complexity and clarity of arguments (e.g. Barzilai et al., 2015; Christenson et al., 2014). These studies frequently apply the well-known Toulmin's Argumentation Pattern (TAP) (Toulmin, 1958) to assess the quality of SSA (e.g. Jimenez-Aleixandre, 2002; Chang & Chiu, 2008; Dawson & Carson, 2017). Moreover, the accuracy of evidence is frequently used as a quality criterion for SSA (e.g. Christenson & Chang Rundgren, 2015). According to this criterion, in order to be evaluated as high quality argumentation an individual's claim need to be grounded by 'correct and relevant knowledge' (Christenson & Chang Rundgren, 2015, p.4). Studies in this category frequently use argument assessment models that include epistemic criteria, such as Sadler's and Fowler's (2006) quality rubric (e.g. Isbilir et al., 2014).

In contrast, some studies use argument criteria focusing on the features of SSA, such as the complexity and multidimensionality (e.g. Christenson et al., 2012; Feierabend et al., 2012). The complexity of an argument often includes counting of the number of claims, reasons or rebuttals present in the argument (e.g. Barzilai et al., 2015). Models used to assess complexity include an assessment tool developed by Erduran et al. (2004) and Kuhn (1991). Other studies emphasize the role of multiple sources in the quality of SSA. These studies use SSA-specific models such as the SEE-SEP model developed by Chang Rundgern and Rundgren (2010), which analyses SSA with respect to six subject areas sociology/culture (So), environment (En), economy (Ec), science (Sc), ethics/morality (Et) and policy (Po) in combination with aspects of knowledge, value and experiences (Rundgren et al., 2012; Chang Rundgren & Rundgren 2010).

Finally, there are a number of studies assessing the quality of SSA according to the characteristics of the specific SSI presented in the task. These argument criteria examine whether the individual uses specific content knowledge and sources relevant to the issue under discussion. An example study which used issue-specific criteria to assess SSA is Jin et al. (2016), in which the authors adapted Walton's (1996) argument schemes in order to examine whether students constructed consistent arguments regarding energy consumption (Jin, Hokayem, Wang, & Wei, 2016).

3.1.3 The importance of consistency in operationalisations

As described in the previous sections, in the last decades SSA has been defined and measured in various ways. Consequently, the question that arises is whether there are commonalities in conceptualization or measurement of SSA and whether researchers using similar conceptualizations, use accordingly similar measurement tools. To answer this question, one has to pay closer attention to the process of operationalization and to degree in which there is consistency among these processes.

The role of operationalization in educational research is crucial, since it serves as the link between questions and answers (Kane, 2001). In other words, operationalizations connect theoretical ideas to evidence (Foschi, 2014); thus, operationalizations are the translation of theoretical concepts into tasks and procedures aiming at the access of more information about its levels (Foschi, 2014). However, previous work on argumentation in science education has posed concerns regarding the justification of measurement approaches used to measure SSA in relation to the theoretically described features of the construct (e.g. Erduran, 2007).

For instance, in a critical review of literature Nielsen (2012) concluded that although argumentation in science education has been conceptualized as dialectical, the commonly used assessment approaches focused mainly on the 'measurable constructs of core elements' (Nielsen, 2012, p. 391). Thus, he came to the conclusion that there is 'a mismatch between the intended object of study and the analytical approaches used to investigate that object of study' (Nielsen, 2012, p. 391). Similarly, in a review on methodological approaches in science education, Kelly et al. (1998) highlighted a lack of alignment between theoretical attributes of nature of science and the methods used to assess the construct. In the same vein, Erduran (2007) argued that the selection of methodology in argumentation in science education should follow a more ‘grounded approach’ (Sandoval & Reiser, 2004; Maloney & Simon, 2006), since the models and criteria currently used derive from the adaptation of theoretical models (e.g. Walton’s argument scheme) (Erduran, 2007).

This heterogeneity, as well as the mismatch between theory and assessment raise methodological concerns regarding the reliability and validity of the argument assessment tools (e.g. Duschl et al, 1999). According to this view, the heterogeneity on conceptualization and measurement of SSA, could have a potential impact on the interpretation of research findings in literature related to SSA, as differentiated operationalization and measurement (e.g. coding criteria, tests) could lead to diverge-or even contradicting- results (Thompson, 1994).

3.2 Aim of the Study

The present paper is driven by two main goals; firstly, with this critical review we aim to map the field of SSA research in terms of theory and methodology in order to support conceptual clarity and gain an overview of measures and ways of operationalization. Secondly, we aim to critically investigate the match between theory and methodology with the ultimate goal of uncovering potentials for stronger alignment between theoretical assumptions and methodological approaches in SSA research. In doing so, we put a special focus on the question of issue-specificity of SSA, as well as its distinctive characteristics from other argumentative forms, such as scientific argumentation. The overall goal of the study is to contribute to the fast growing and flourishing field of SSA research by reviewing the current state of research and highlight opportunities and possible research strands for upcoming research.

Research questions

(1) Is SSA conceptualized as a special and/or issue-specific argumentation form in educational research?

(2) Is SSA measured by SSA-specific instruments and/or SSI-specific instruments?

(3) Are the methods used to assess SSA consistent with the conceptualizations of SSA?

3.3 Materials and Methods

This study is a part of a larger systematic literature review project investigating the terms, definitions, as well as measurements regarding SSA for the period 1999-2017. In our previous study an emphasis was given to the current practices of conceptualizing and empirically operationalizing the concept of SSA in science education research (Ioannidou et al., manuscript in preparation). The present study is focused to critically reviewing the alignment between conceptualizations and operationalizations of SSA.

3.3.1 Literature search methods and selection process

In order to answer our research questions, a literature review was conducted. The literature search was conducted through two databases: Web of Science (all databases) and EBSCOhost Library (ERIC and PsycArticles) and three main keywords were used: *argumentation*, *social* and *scientific*. Furthermore, to ensure the educational settings of the studies the keywords ‘teacher’, ‘student’ and ‘classroom’ were additionally used. This search resulted in 638 hits, after the duplicates were removed.

As a next step, the screening process was conducted on the basis of the following inclusion criteria: (a) the terms ‘argumentation’ and/or ‘reasoning’ and the term ‘socio(-) scientific’ should be present in the title or the abstract of the paper (b) the paper should be a peer-reviewed, (c) empirical article, (b) written in English language. After applying these criteria, 138 studies in total were found eligible. Finally, a full-text search was conducted; to be included, the terms ‘reasoning’ or ‘argumentation’ should be mentioned in research questions of the study and a type of argument assessment should be mentioned in the methods section. After the full-text screening, a total number of 67 studies were included for full-text content analysis.

3.3.1 Study analysis

Development of coding scheme. The coding scheme was developed in an iterative coding process, as inductive and deductive coding processes were followed. For the purposes of this study, we will describe the eight codes that were used to address the research questions of this study.

The theoretical sections of the selected articles (e.g. Introduction) were coded to answer our research question addressing the conceptualization of SSA. The coding for this aspect was conducted inductively in a bottom-up process: coded segments were grouped and labelled into codes, which were later clustered into coding categories (Appendix x). Our research question, regarding the measurement of SSA, was addressed by the coding of the sections, in which the methodology of the study was described (e.g. Methods). Our codes for this aspect were created deductively, as our coding scheme was based on a previous literature review on argumentative competencies (Rapanta et al., 2013).

After the development of the coding scheme, two independent raters coded 20% (n= 14) of the overall selected articles. As a next step, inter-rater reliability was calculated; Cohen’s kappa ranged between K= 0.70 and K= 0.89 and the few remaining discrepancies were resolved through discussion.

Description of the coding scheme. The final coding scheme consisted of two coding aspects and, overall, eight codes. The theoretical sections of the selected studies were coded with respect to three codes: SSA as a special argumentation form, SSA as issue-specific argumentation form and mentioned quality indicators. The methods sections were coded with regard to five codes, namely: SSA-specific instruments, SSI-specific instruments, applied quality indicators, argument assessment criteria and argument assessment models. A more detailed description of the coding scheme is presented in the table below (Table 6), which displays the labels of the codes used with the respected descriptions and coded segments from the selected studies. Although codes (1), (2), (4) and (5) were coded as binary, codes (3), (6), (7) and (8) were coded by means of sub-codes, which are displayed in the table below (Table 6).

Table 6

Description of Codes and Example Segments

Code	Description	Example segment
Coding aspect 1: Conceptualization of SSA		
(1) SSA as a special argumentation form	• Does the author(s) conceptualize socio-scientific argumentation as a special form of argumentation? • Do the authors use the term 'socio-scientific' to describe SSA related constructs (e.g. SSR, SSI instruction)	'It is important to stress, however, that scientific argumentation is not the same as SSI' (Grooms et al., 2014, p. 1415).

Table 6 (continued)

Description of Codes and Example Segments

Code		• Description	Example segment
(2)	SSA as issue-specific argumentation form	• Does the author(s) argue that the quality indicators and the involving skills during SSA differentiate depending on the given socio-scientific issue?	'…argumentation may differ for different topics, such as global warming, genetically modified organisms (GMOs), nuclear power or consumption.' (Jonsson, 2016, p. 1826).
(3)	Conceptualized quality indicators	• Does the author(s) present elements of SSA that differentiate high from low quality of argumentation (e.g. counter-arguments or rebuttals)?	'…the skill to provide justification together with a claim is of great importance when assessing the quality of SSI argumentation..' (Christenson & Rundgren, 2015, p.2)
Coding aspect 1: Measurement of SSA			
(4)	SSA-specific instruments	• Does the author(s) use instruments specifically designed for analysing SSA?	'…the SEE-SEP model has been developed to integrate the cross-disciplinary dimensions and the four features of SSIs …' (Cristenson et al., 2012, p. 344)
(5)	SSI-specific instruments	• Does the author(s) use instruments or quality criteria specifically designed for a specific socio-scientific topic?	'..in a resource-based environment concerning GMF could be grouped into seven categories..' (Chang, 2007, p. 11)
(6)	Applied quality indicators	• Does the author(s) use elements of SSA, such as counter-arguments or rebuttals, to assess the quality of SSA?	'…were more based on justifications and counter-positions, the level of the argumentation is determined to be higher…' (Isbilir et al., 2014, p. 375)
(7)	Argument assessment criteria	• Does the author(s) use specific argument criteria (e.g. structure or use of evidence) to determine SSA quality?	'…the three subskills outlined above were used: (1) Using scientific knowledge in discussions about SSIs…' (Jonsson, 2016, p. 1832)
(8)	Argument assessment models	• Does the author(s) use argument models (e.g. Toulmin's TAP or the SEE/SEP model) to assess SSA?	*'The analytical framework was built upon Toulmin's (1958) argument pattern'* (Foong & Daniel, 2013, p.2342)

3.4 Results

In the next sections, our findings will be presented with respect to: (a) the review of conceptualization of SSA (b) the review of measurement of SSA and (c) the coherence between conceptualization and measurement of SSA.

3.4.1 Review of conceptualization of SSA

To answer our research question regarding the distinctiveness and issue-specificity of SSA we report frequencies of codes, as well as qualitative results regarding (a) SSA as a special argumentation form and (b) SSA as issue specific form. Moreover, we report frequencies of concepts related to SSA, namely SSR, SSI decision-making and SSI instruction.

SSA as a special construct. With regard to the conceptualization of SSA, only two out of 67 studies explicitly mentioned that SSA differs from other types of argumentation. Namely, Grooms et al. (2014) reported that '…scientific argumentation is not the same as SSI' (Grooms et al., 2014, p.1415), while Kim et al. (2014) suggested that 'there is some degree of difference between scientific argumentation and the type of argumentation when resolving socioscientific issues' (Kim et al. 2014, p.917). Nevertheless, overall 25 out of 67 studies used the term socio-scientific argumentation or terms related to socio-scientific argumentation. Out of these, six used the term socio-scientific argumentation (6/67), eight used the term socio-scientific decision-making (8/67), eight the term socio-scientific reasoning (10/67) and three the term socio-scientific instruction (4/67).

Concepts related to SSA. Socio-scientific reasoning. SSR was mostly mentioned to describe the reasoning process during SSI discussions. For instance, Karahan and Roehrig (2017) suggest that SSA encompasses a 'new type of reasoning that incorporates the aspects of practices and skills associated with the negotiation of socioscientific issues'. In addition, SSR was mentioned as '…a theoretical construct which subsumes aspects of practice associated with negotiation of SSI…' (Sadler et al., 2007, p. 374). Furthermore, Sadler and Zeidler (2005) presented distinctive patterns of informal reasoning in the context of SSI, namely: rationalistic, emotive, and intuitive informal reasoning.

SSI decision-making. Decision-making processes were also mentioned in relation to informal reasoning regarding SSI. According to Laius and Rannikmaee (2011), SSI decision-making is embedded in the concept of scientific literacy as through SSI discussions students are able to contribute to public debates and 'make informed and balanced decisions' (Laius & Rannikmaee , 2011, p.127). Moreover, according to Sakschewski et al (2014), high quality SSI decision making can be identified by the integration of contradicting arguments, the ability to use trade-offs and to prioritize arguments to reach a decision.

SSI instruction. The concept of SSI teaching (or curriculum) was mentioned as opposition of content-based instruction approaches (e.g. Eggert et al., 2017). It has been characterized as contextualized teaching that moves toward a more socio-cultural approach (Eggert et al., 2016, p. 2). Furthermore, SSI instruction was presented as an opportunity for students to engage in interdisciplinary science topics, which are complex and have an impact on their daily lives (e.g. Juntunen & Aksela, 2014).

SSA as issue-specific argumentation. Within our selected studies, six out of 67 studies conceptualized SSA as issue-specific. These studies argued that the reasoning and argumentation processes depend on the specific topic under discussion and, as such, SSA varies between different SSI (e.g. Chang & Chiu, 2008). More explicitly, according to Khishfe et al. (2017) '…the quality of informal reasoning could be influenced by the context of the SSI' (Khishfe et al., 2017, p. 6). Kim et al. (2014) claims that there are two context-dependent factors that could lead to differences in SSI discussions: 'the nature of available evidence and the immediacy of the topic' (Kim et al., 2014, p., 918).

In contrast, five out of 67 studies, claimed that the SSA is not issue-dependent, presenting a ‘holistic viewpoint’ for SSA (Christenson et al., 2012). According to these studies, argumentation skills are ‘domain independent’ (Yang, 2004) and can be transferred across topics (e.g. Grooms et al., 2014; Foong & Daniel. 2013).

3.4.2 Review of measurement of SSA

Our results regarding the measurement of SSA are presented with respect to (a) SSA specific (b) SSI-specific tools by means of quality indicators, argument assessment criteria and argument assessment models. The summary in Table 7 displays the conceptualizations and measurement approaches of SSA applied in the selected studie

Table 7

Summary Table for Conceptualization and Measurement of SSA (N= 67)

Study	**Definitions**			**Criteria applied**	**Model applied**
	SSI	**SSA**	**SSR**		
1) Agell (2015)					Other
2) Åkerblom & Lindahl (2017)		x		Arg,Mov	Other
3) Albe (2007)				ArgMov,SourceEv	
4) Albe & Gombert (2012)	x	x		ArgMov,UseEv,SourceEv	
5) Barzilai et al. (2015)	x			Coher,Multidisci,Struct,SourceEv	
6) Cetin et al. (2014)	x	x		Struct	TAP(O),Erdu(O)
7) Chang (2007)		x		Predef,SourceEv	
8) Chang & Chiu (2008)	x	x		ArgMov,Multidisci,Other,SourceEv	Lakat(O)
9) Christenson & Chang Rundgren (2015)		x		Accur,Struct,SourceEv	
10) Christenson et al. (2012)	x			SourceEv	SEE-SEP(O)
11) Christenson et al. (2014)	x			Struct,Other,SourceEv	SEE-SEP(O)

Table 7 (continued)

Summary Table for Conceptualization and Measurement of SSA (N= 67)

Study	**Definitions**			**Criteria applied**	**Model applied**
	SSI	**SSA**	**SSR**		
12) Cinici (2016)	x	x		Multidisci,Struct	
13) Craig-Hare et al. (2017)	x			Predef,ArgMov,Other	
14) Eggert et al. (2017)				ArgMov,Accur,Multidisci,Struct	
15) Feierabend et al. (2012)		x		Coher,Multidisci,Struct,UseEv,Other	
16) Foong & Daniel (2012)	x	x		Multidisci,Struct	TAP(A)
17) Grooms et al. (2014)	x	x		Coher,Struct,SourceEv	ZohNem(A)
18) Hansson et al. (2011)	x			UseEv,Other,SourceEv	TAP(A)
19) Iordanou & Constantinou (2015)				ArgMov,Accur,SourceEv	Kuhn(A)
20) Isbilir et al. (2014)	x			Coher,Struct	SadlFowl(A)
21) Jimenez-Aleixandre & Pereiro-Munoz (2002)				SourceEv	
22) Jin et al. (2016)				Struct,Other,SourceEv	
23) Jonsson (2016)		x	x	Accur,Multidisci,Struct,UseEv,SourceEv	SEE-SEP(A),Lakat(A)
24) Juntunen & Aksela (2014)	x		x	SourceEv	
25) Karahan, & Roehrig (2017)	x		x	Other	
26) Keskin et al. (2013)		x		Multidisci	
27) Khishfe (2012)			x	Multidisci,Struct	
28) Khishfe et al. (2017)	x		x	Multidisci,Struct	
29) Kim et al. (2014)	x	x		ArgMov,SourceEv	Duschl(A),Duschl(O),Walt(O)
30) Kolsto (2006)	x			Other,SourceEv	
31) Laius & Rannikmaee (2011)				Coher,SourceEv	
32) Lawless & Brown (2015)				Struct	
33) Lin (2014)				Struct,UseEv	
34) Lin & Mintzes (2010)	x			Accur,Struct,Other	
35) Lin et al. (2012)	x			Accur,Struct,UseEv	TAP(O)Erdu(O)
36) Lindahl & Linder (2013)	x			ArgMov,Other	
37) Lindhal & Lundin (2016)	x			UseEv,Other,SourceEv	

Table 7 (continued)

Summary Table for Conceptualization and Measurement of SSA (N= 67)

Study	Definitions			Criteria applied	Model applied
	SSI	SSA	SSR		
38) Molinatti et al. (2010)		x		Struct	TAP(O)Other
39) Nam & Chen (2017)				Predef,ArgMov	Walt(O)
40) Namdar (2017)		x		Multidisci,Struct,UseEv	Erdu(O)
41) Nielsen (2012)		x		ArgMov,SourceEv	
42) Ozturk & Yilmaz-Tuzun (2017)	x			Accur,Struct	SadlFowl(A)
43) Patronis et al. (1999)				ArgMov,Other,SourceEv	Other
44) Robertshaw & Campbell (2013)	x			Struct,UseEv	
45) Romine et al. (2017)	x		x		
46) Rudsberg et al. (2013)				Struct	TAP(O),TAP(A)
47) Rundgren et al. (2016)	x	x		ArgMov,SourceEv	Other,Lakat(O)
48) Saad et al. (2017)	x		x	Struct	ZohNem(A)
49) Sadler & Zeidler (2005a)	x	x			Kuhn(A)
50) Sadler & Zeidler (2005b)	x	x	x	Struct	
51) Sadler & Donnelly (2006)	x	x		Struct	
52) Sadler & Fowler (2006)	x	x		Struct	TAP(A)
53) Sadler & Zeidler (2005)	x	x	x	Struct	
54) Sadler et al. (2007)	x		x	Struct	
55) Sakschewski et al. (2014)	x	x		Struct	
56) Simonneaux & Chouchane (2011)				Struct	
57) Simonneaux et al. (2013)	x			ArgMov,Other	Other
58) Solli et al. (2014)	x	x		ArgMov,SourceEv	Other
59) Topcu et al. (2010)	x			Struct	TAP(A),Kuhn(A)
60) Venville & Dawson (2010)	x			Struct	TAP(A)
61) von Aufschnaite et al. (2008)				Multidisci,Struct	
62) Walker & Zeidler (2007)	x			ArgMov,Accur,Struct,UseEv	TAP(O)
63) Wu & Tsai (2007)			x		
64) Wu (2013)	x	x	x	Multidisci,SourceEv	Kuhn(A),Other
65) Yang (2004)					
66) Yu & Yore (2013)	x			Struct	Other
67) Zeidler et al. (2013)			x	UseEv,Other	Other

Note. ArgMov= argument moves, Struct=structure, SourceEv= sources of evidence, UseEv= use of evidence, Accur= accuracy, Coher=coherence, Multidisci= multidisciplinarity, (A)= adapted model, (O)= original model.

SSA-specific tools. In our sample studies, rebuttals (13/67) and the use of counter-arguments (6/67) were used to assess the quality of SSA. According to Ozturk & Tuzun (2018) high quality SSA includes 'supportive arguments, initial counterarguments, and supportive arguments for their positions, addressing counterarguments, and undertaking rebuttals' (Ozturk & Tuzun , 2018; Wu & Tsai, 2011).

Regarding SSA-specific argument criteria, multidimensionality (30/67) and complexity (16/67) were mostly used. The use of multidimensionality as a criterion was supported by the assumption that during SSI individuals used various sources of information on order to make a decision (e.g. Lindhal & Lindin, 2016; Sadler & Zeidler, 2005; Walker & Zeidler, 2005). Similarly, the use of complexity was based on the common claim that SSA is complex and should include multiple pro and contra arguments (e.g. Kaharan & Roehrig, 2017; Lawless & Brown, 2015; Rundgren et al., 2016). Regarding the argument models, the SEE-SEP model, developed by Chang Rundgren and Rundgren (2010) was the only SSA-specific tool used used by three out of 67 studies (Christenson et al., 2012; Christenson et al., 2014; Jonsson, 2016).

SSI-specific tools. Quality indicators and argument assessment criteria related to the specific characteristics of SSI were the use of scientific evidence (11/67) and the accuracy of evidence (11/67). For instance, in Aleixandre (2002) the use of relevant knowledge is considered essential 'in order to understand and decide about the issue…' (Aleixandre, 2002, p. 1175). According to Christenson & Chang Rundgren (2015), this knowledge should be grounded and relevant to the supporting claim. Issue-specific tools were mostly used to categorize arguments regarding specific SSI. For instance, in Wu (2013), the author used Kuhn's framework (1993) to categorize arguments regarding GMFs, while for the same SSI Chang (2007) used Lakatos' visual model to assess the quality of argumentation. Another study that focused on the issue-specificity of SSA was Jin et al. (2016), which assessed the quality of SSA by using Walton's scheme (1996) to categorize arguments relevant to matter and energy.

3.4.3 Coherence between conceptualization and measurement

To answer our research question, regarding to the coherence between conceptualization and measurement, we created contingency tables containing codes mentioned in the theoretical sections of the articles (theoretical background, rationale and problem statement) and codes presented in the methodological sections (methods, results and conclusions); the results are presented in the following sections.

Conceptualized and applied quality indicators. The contingency Table 8. presents the frequency of quality indicators for SSA mentioned in the conceptualization (column 1) and measurement of SSA (column 2). Column 3 ('Conceptualized & Applied') indicates the frequency of publications in which the codes mentioned in the conceptualization were also used in measurement sections (e.g. in scoring rubrics). The table shows the frequency in which justification and use/quality of evidence were conceptualized was higher than the frequency in which they were applied. In addition, the frequency in which rebuttals were applied was higher than the frequency in which it was conceptualized. Furthermore, a total of 9 studies were found to have conceptualized the measurements that they applied.

Table 8

Frequency of Mentions Regarding Quality Indicators in Conceptualization and Measurement of SSA

Quality indicators	**Conceptualiza-tion provided (f)**	**Applied in Measurement (f)**	**Conceptualized & Applied (f)**	**Total**
Complexity	8	1	1	10
Counter-arguments	8	6	1	15
Justification	10	8	2	20
Knowledge	5	0	0	5
Multiple perspectives	6	2	0	8
Rebuttals	9	13	3	25
Use/Quality of evidence	14	8	2	24
Total	60	38	9	107

Argument assessment models and argument assessment criteria. The link between conceptualization and measurement was also assessed through the link between the choice of argument assessment models and argument assessment criteria. We conducted this analysis, as we expected that studies that used specific models of assessment would use similar assessment criteria for SSA. The results in Table 9. illustrate that most of the criteria were used by more than five argument models. The only exception was the high frequency of structural criteria when TAP was applied. However, the table shows no other association between argument assessment criteria and the models applied.

Table 9

Frequencies of Argument Assesment Criteria by Argument Assessment Model

	Argument assessment model								
Assessment criterion	**Toulmin** (f)	**Erduran** (f)	**Lakatos** (f)	**Walton** (f)	**SEE-SEP** (f)	**Duschl** (f)	**Zohar & Nemet** (f)	**Sadler & Fowler** (f)	**Kuhn** (f)
Structure	10	3	1	0	1	0	2	2	0
Multidimensionality	1	0	2	2	1	1	1	0	2
Argument moves	0	0	2	2	0	1	0	0	1
Complexity	1	1	1	0	1	0	0	0	1
Use of scientific evidence	3	2	2	0	1	0	0	0	0
Accuracy	2	1	1	0	0	0	0	1	1
Coherence	0	0	0	0	0	0	1	1	0
Use of pre-defined arg. criteria	0	0	0	1	0	0	0	0	0

3.5 Discussion

The aim of this critical review was to provide an overview of the research area of SSA within the last twenty years with regard to the conceptualization and measurement of the construct, with a special focus on its specific characteristics. At the same time, we intended to critically examine the alignment between theory and methodology, in order to identify potential theoretical and methodological inconsistencies in the existing literature. In this section firstly, we discuss our main findings, we then present possible limitations of the study and, lastly, we present some conclusive comments.

3.5.1 Main findings

Conceptualization of SSA. With regard to our question, whether researchers conceptualize SSA as a special construct, our results revealed that only a small number of studies explicitly described SSA features as distinct, while this distinction was made in comparison to scientific argumentation (Grooms et al. 2014; Kim et al. 2014). However, more than one third of the studies used the term socio-scientific for describing argumentation or other related constructs (reasoning, decision-making, instruction). Thus, considering the lack of clear definitions on the topic (Ioannidou et al, manuscript in preparation) it is unclear whether the use of special terminology for SSA or SSA-related concepts is due to theoretical assumptions about its distinctive nature.

This finding is in line with previous reviews on argumentative competencies in educational research showing a lack of clarity regarding the definition of the measured constructs (e.g. Rapanta et al, 2013). Yet, since the reasoning as well as decision-making processes taking place during SSA are commonly described as different from the ones occurring during scientific argumentation, one would assume that the resulted argumentation would take different forms and should be assessed with different criteria. Thus, future research should pay closer attention to the choice of specific terminology, as well as the justification of the used terms (Erduran, 2007; Nielsen, 2012).

Measurement of SSA. In terms of measurement, almost half of the studies assessed SSA with regard to multidimensionality and one forth with regard to complexity of the argument, while only one instrument was categorized as SSA-specific: the SEE-SEP model developed by Chang Rundgren and Rundgren (2010). Furthermore, since SSE-SEP model was used by three studies the majority of the studies used measurement tools developed for measuring other constructs (e.g. TAP, Walton's scheme).

Moreover, fewer studies used issue-specific quality indicators, such as the use and accuracy of scientific evidence. The selection of these schemes can be explained due to the fact that SSA is a relatively new construct; therefore, researchers used already established instruments developed to assess general argumentative competencies (Rapanta et al., 2013).

Coherence between measurement and conceptualization. A finding of our analysis was the mismatch between theoretical considerations regarding SSA and methods applied by the researchers. A remarkably small number of studies justified and operationalized the quality indicators that were applied during measurement of SSA. An indicative example is that, although rebuttals were frequently conceptualized or applied as a quality indicator, only three studies seem to have conceptualized, as well as, applied it in measurement tools. This was an unexpected finding as, based on general guidelines in educational research, we assumed that studies would operationalize the criteria, quality indicators and models applied (e.g. Institute of Education Sciences, 2013).

Furthermore, our results showed that argument assessment models were applied with diverse foci; e.g. dialogical models, such as Lakatos' scientific programmes (Chang & Chiu, 2008), were used to assess the epistemic quality of the arguments (e.g. the accuracy or existence of scientific evidence). This finding raises questions regarding the way in which researchers decided upon the applied argument assessment models and whether they acknowledge their individual characteristics. Thus, our results suggest that future studies on SSA should describe the process of operationalization of construct in more detail, while justifying the decisions made during the process.

Towards this direction, we suggest the use of a conceptualization model presented by Baxter and Babbie (2003), according to which the abstract mental representations of concepts become measurable through the creation of nominal and operational definitions. During the construction of nominal definitions, researchers should focus on specific dimensions and indicators. Thus, only when a certain degree of agreement regarding the conceptualization is achieved, the validity of instruments can be examined.

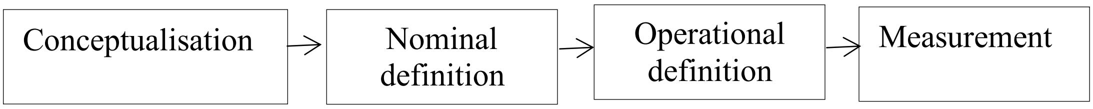

Figure 10

Diagrammatic Representation of Steps Followed During the Operationalization of a Concept (Baxter & Babbie, 2003)

3.6 Limitations

It should be acknowledged that this critical review is not without limitations. Despite the fact that we included studies from the two databases most commonly used in educational research, we are aware that there are publications that we did not include in this review. However, since there is no previous review on the topic, we consider this review as a first step to critically examine the existing literature on the topic.

Another possible limitation for this study was the bottom-up process followed in the categorization of SSA- and SSI-specific tools. Given the exploratory nature of our study, the categories were created based on the characteristics of the applied tools. Nevertheless, to avoid ambiguity, specific coding rules were applied and the calculated inter-rater reliability was satisfactory.

3.7 Conclusions and Future Directions

While SSA is becoming a topic of increased research interest, the heterogeneity in SSA conceptualizations and measurement raise questions regarding the very nature of the construct and the 'adequate' ways in which it should be measured. Current literature on SSA shows heterogeneity as well as inconsistencies in the conceptualization and measurement of the construct. Although previous reviews have shown some commonplaces in the general features of SSA (Ioannidou et al, manuscript in preparation), there is a small number of studies that explicitly address the issue of the nature of SSA as a distinct concept. This issue is amplified by the fact that, while the majority of the studies do not make any mention on the issue the distinctiveness of SSA, there are two contradicting views regarding its issue-specificity. Furthermore, while SSA has been mostly measured by SSA-general tools, there is variance in the criteria applied when specific argument models are used.

The main contribution of this study is the critical examination of the consistency between theory and measurement in the study field of SSA. By doing so, we aimed to foster the interpretation of results in current literature on SSA, while at the same time opening a discussion regarding the nature and measurement of SSA. Although we are aware that several factors lead to the selection of an assessment tool over another, such as the type of participant or the type of argument (e.g. oral or written) (e.g. Erduran, 2007), our aim was to raise attention regarding the importance of clarity in the processes of conceptualization and operationalization of SSA. In this direction, we suggested the use of the diagrammatic model presented by Baxter & Babbie (2003) and the inclusion of more nominal, as well as operational definitions of SSA. In doing so, we aspire to advance the theory, methods and analytical tools within the already growing community of SSA in science education.

4. GENERAL DISCUSSION

4.1 Summary of Study 1 and Study 2

4.1.1 Study 1

Study 1 reviewed studies on socio-scientific argumentation in science education published from 1999 to 2017. This study aimed to provide an overview of existing literature on socio-scientific argumentation, in terms of theoretical approaches in defining and conceptualizing SSA, as well as methodological approaches with regard to the applied measurement frameworks and assessment tools. To that end, Study 1 reviewed the terminology, the definitions and the concepts used to describe SSA in science education studies. Furthermore, it examined the approaches of operationalization, as well as the instruments and argument criteria that were used in order to assess SSA.

In terms of methodology, used two databases to retrieve literature and after the application of inclusion criteria, 67 studies were included in the full-text analysis. The included studies were analysed by means of a coding scheme which was developed by the authors using inductive and deductive coding approaches. The publication trends reported in the study showed an increase in publications regarding SSA after 2011 with a peak in 2014 and 2017.

Regarding the conceptualizations of SSA, the majority of the included studies did not include a definition for SSA, while when definitions were presented the terms that were most frequently used to describe SSA were *argumentation on/about SSI*, *SSI argumentation*, *informal reasoning* and *informal argumentation*. At the same time three themes emerged in the definition of SSA: a) SSA is argumentation regarding SSI b) SSA is negotiation and decision-making to solve an SSI and c) SSA is the expression of informal reasoning regarding SSI. Moreover, authors tended to describe SSA as a form of argumentation characterized by controversy, highlighting the moral and social issues that it entails.

With regard to the applied measurement approaches, most studies used secondary school students as a sample, while discussion and written arguments to address written scenarios were the tasks that were mostly chosen. The majority of the studies conducted qualitative analysis and more than half of them quantified their results. Most of the studies used criteria based on the structure and the different sources of evidence to asses SSA and while researchers applied a variety of argument assessment models, TAP was among the schemes mostly used. Regarding the applied quality indicators, rebuttals, justifications and the used of evidence were amongst the ones that were emphasized through the applied models.

Overall, Study 1 showed that although there is an increasing trend in publications regarding SSA, the construct has not yet been adequately defined and conceptualized by researchers. In addition, there is a wide range of measurement tools currently used to assess SSA in science education, while researchers seemed to emphasize structural criteria when assessing argumentative competencies in SSI-contexts. Study 1 recommended not only the inclusion of definitions and clearer conceptualizations of SSA, but also the justification of the chosen terminology for SSA. Finally, in an attempt to promote a common language among researchers in SSA research, Study 1 presented eight argument assessment criteria as emerged from the selected studies: structure, use of pre-defined criteria, complexity, multidimensionality, use of scientific evidence, accuracy, coherence and argument moves (Table 3). The use of these criteria was, ultimately, proposed as an interpretative framework for researchers and future readers of research in SSA.

4.1.2 Study 2

Study 2 aimed to investigate theoretical, as well as methodological considerations with regard to the operationalisation of SSA in studies in science education. Firstly, the study examined whether SSA is conceptualized as a special construct and, if this is the case, whether researchers identified special features in SSA that would differentiate SSA from scientific or general argumentation skills in education. Moreover, it examined whether researchers perceive the characteristics of SSA to be dependent on the SSI which is discussed. Apart from theoretical considerations, the second goal of Study 2 was to present whether the tools that have been used to assess SSA were designed: a) to capture the special features of SSA (SSA-specific tools) and b) to address arguments and characteristics of argumentation regarding specific SSI (SSI -specific tools). Finally, the third aim of Study 2 was to examine the degree to which the conceptualized characteristics of SSA as presented by the authors were reflected in the applied measurement approaches and tools.

To answer these questions, the studies selected in Study 1 (N=67) were coded with the use of a coding scheme developed by the authors. With regard to the nature of SSA, the results presented in Study 2 showed that only two studies explicitly referred to SSA as a construct different than other types of argumentation, whereas 25 studies used the term socio-scientific argumentation, socio-scientific decision-making, socio-scientific reasoning or socio-scientific instruction. With regard to whether SSA is regarded as issue-specific, there were two opposing views found in the selected studies; six studies claimed that SSA is issue-dependent, whilst five argued that SSA is issue-independent and, as a result, a transferable skill. Concerning the alignment between conceptualizations and measurements, the results showed that only 9 studies provided a theoretical justification for the quality indicators that were applied in the measurement of SSA. Moreover, Study 2 illustrated that the argument assessment models were applied to assess different assessment criteria than the ones traditionally assessed by the models. For instance, models traditionally applied to assess dialogic features of argumentation, (e.g. Kunh, 1993) were used to assess the epistemic features of arguments, such as the accuracy of information.

The results of Study 2 were discussed in light of theoretical and methodological misalignments that were found in the selected studies. Firstly, the study called for greater clarity on the use of specific terminology and the theoretical assumptions behind its use. For instance, there is a need for clarification of the conceptual differences between the terms socio-scientific argumentation, SSI argumentation and argumentation regarding SSI. In addition to the choice of terminology, there is a need for clearer distinctions, if any, between SSA and other related constructs (e.g. scientific argumentation, decision-making). Regarding the measurement of SSA, multidimensionality and complexity were the two argument assessment criteria that were used to capture special features of SSA, such as controversy and open-endedness. In the case of issue-specific measurements, the use and accuracy of scientific evidence were emphasized. However, the majority of the studies used tools developed for measuring other constructs (e.g. general argumentative competencies). This finding could be explained by the fact that SSA is a new field of study and, as a result, researchers preferred the use of models that have a long tradition in measuring argumentation skills.

In terms of the coherence between measurement ad conceptualization of SSA, Study 2 reported that the small number of studies that justified their applied quality indicators, while argument models were used in various ways. This finding raised questions about the operationalization processes followed by researchers. To address this issue, Study 2 suggested the use of a conceptualization model presented by Baxter & Babbie (2003), according to which the process of operationalization should start from the conceptualization of the construct. After the conceptualization of the construct, researchers should present nominal and, followingly, operational definitions of construct. At the end of this process and when all steps are followed the construct can be measured. With this suggestion Study 2 aimed to promote greater transparency in the process of operationalization of SSA in science education.

4.2 Theoretical and Methodological Implications

The General Introduction of the thesis discussed the main shifts with regard to the educational goals in science education. These shifts have set the ground for the introduction of SSA in science classrooms and subsequently the examination of SSA in science education research. Although argumentation in science education has been studied for decades mainly in the form of scientific argumentation (Erduran, 2007), scientific and societal changes necessitated the introduction of more complex, ill-structured issues, such as SSI, in science classrooms. Whilst the need for inclusion of SSI-teaching has been generally accepted, the terms and definitions, as well as the methodological approaches followed to measure SSA, remain ambiguous. As argued in Study 2, the intention of this thesis is not to find the "correct" way to define and measure SSA. On the contrary, Study 1 and Study 2 acknowledged the various ways in which the construct has been addressed in an attempt to open a discussion about different approaches of defining and assessing SSA.

One could claim that the differences in conceptualizations and, especially, the measurement of SSA derive from the tendency to "measure first that which seemed most feasible" (Herring, 1918, p. 558); yet this would oversimplify the task undertaken by researchers in the field of SSA. In this section, the findings from Study 1 and Study 2 will be discussed in light of some challenges in the study of SSA.

4.2.1 Conceptual and measurement approaches and educational goals

Study 1 showed that most of the studies investigating SSA examined secondary school student's argumentation. As discussed in the General Introduction, educational and pedagogical goals often direct the ways in which argumentation is operationalized and measured in science education (Erduran, 2007). Thus, we would expect that the theoretical and methodological approaches described in Study 1 and Study 2 are informed by the educational goals in science in secondary education. These educational goals are often described by curricula and may differ among different educational systems. In other words, "what makes a good argument" may differ depending on the different educational goals. Bricker and Bell (2008) argued that:

"Whereas designers and analysts typically select a single conceptualization for argumentation, it is clear that because particular ones are suited to specific purposes that a given science education enterprise might focus on multiple purposes and therefore, benefit from consideration of multiple conceptualization" (Bricker & Bell, 2008, p. 491).

According to Erduran (2007) the definition of argumentative competencies will also "dictate" the methodological approaches that will be followed. For instance, in curricula oriented towards Vision I for scientific literacy (General Introduction, section x), where the general goal is the acquisition of scientific knowledge, argumentative activities would be introduced as a means to argue with scientific evidence. As a result, an "argue to learn" approach would be followed. In this context, SSA activities would target the integration of scientific knowledge into argumentation, while the accuracy of information would potentially be used as an argument assessment criterion.

Apart from the educational goals presented in curricula, other factors, such as teacher pedagogical views and practices, may influence the choice of argumentative tasks and measurement approaches for SSA (Zeidler, 2014). In studies conducted in classroom settings, researchers often tend to follow naturalistic approaches by using tasks and procedures that are familiar to teachers and students. This means that the setting of the studies included in Study 1 and Study 2 could be influenced by teaching practices commonly used in science classrooms.

4.2.2 Contextual factors related to SSI

In literature related to SSA, SSI are commonly defined as "open-ended, ill-structured debatable problems that are susceptible to multiple perspectives and solutions" (Lindhal & Lundin, 2016). Study 1 showed that the topics that were mostly used to assess SSA were environmental issues, including global warming and climate change, and issues related to biotechnology and GMOs. Although all these topics can technically be labelled as SSI, they are different in many ways. Firstly, these topics are differently presented by the media and secondly, the amount and the types of available scientific evidence regarding these issues differs significantly.

If we consider the example of global warming and, similarly, climate change, there is currently a plethora of political discussions regarding these issues. At the same time, although presented as controversial issues for public discussion, topics such as climate change are not regarded as controversial within scientific communities. This has implications on the pedagogical goals for argumentative tasks targeting these issues. Since in the discussion about climate change the two opposing opinions use arguments from different sources of evidence (e.g. political vs scientific), it would be challenging to assess argumentation by means of multidimensionality. Hence, one would expect the focus of argumentative activities to be on the integration of scientific evidence, as well as the construction of rebuttals, rather than the use of evidence from different sources.

However, Study 1 showed that several studies, such as Christenson et al. (2012) followed a different approach. Christenson et al. (2012) addressed the topic of global warming by asking secondary school students to answer the question: “Do you believe that climate change is due to natural processes or to human activities?” (p. 346). To assess student SSA, the study applied the SEE-SEP model, which emphasizes the multidimensionality of arguments (see General Introduction, section x). The results of the study showed that students demonstrated lower levels of multidimensionality in their arguments regarding global warming compared to other issues, such as GMO and nuclear power. As explained above, given the nature of the topic, these results are not surprising. Furthermore, Christenson et al. (2012) concluded that students mainly used knowledge from environmental studies and science to back up their claims, rather than evidence from other disciplines.

As a result, in this study, students who used valid scientific evidence from scientific sources only would demonstrate lower scores than students that used pieces of information from different disciplines (e.g. economics, sociology, politics). However, it would be misleading to consider that these students performed “badly”, as the topic of global warming can be mainly backed up with arguments from science. Hence, this example illustrates how the choice of specific measurement tools for specific topics may affect the interpretation of the study results with regard to student performance in argumentation tasks.

Considering the abovementioned example, the question that arises is: Since SSI are diverse, should they be assessed differently? Study 2 showed that only a few of the selected studies used SSI-specific tools. An example of these studies is Jin et al. (2016) who used an SSI-specific tool to investigate argumentation skills regarding energy consumption issues. The study asked students to choose among contradicting views on the issue and to produce arguments to support their claims. The scoring rubric that was used to assess argumentation skills included four levels and the applied argument assessment criteria were structure and use of scientific knowledge. The tool was labelled as issue-specific because the scientific knowledge that was expected to be used by the students was the scientific mechanisms of energy transformation and matter transformation.

4.2.3 Definitions and assessment tools designed for similar concepts

As reported in Study 1, SSA has been conceptualized and measured by means of other conceptually relevant concepts, such as informal reasoning, informal argumentation and decision-making. Study 1 presented three themes to describe the presented definitions in the selected studies, one of which presented SSA as " the expression of informal reasoning regarding SSI" (e.g. Christenson & Chang Rundgren, 2015). Similarly, since argumentation is generally perceived as the expression of reasoning, SSA has also been presented as a type of informal argumentation (e.g. Chang & Chiu, 2008). While in science education informal argumentation is commonly regarded as the expression of informal reasoning, the concept of informal argumentation has not been further conceptually elaborated. For instance, in social sciences, informal argumentation is argumentation that takes place in groups of peers and informal settings (e.g. Nussbaum, 2002). Thus, one could claim that the term "informal argumentation" is not specific and clear enough to facilitate the conceptualization of SSA.

The conceptualization of SSA as informal reasoning has also implications on the applied measurement approaches. More explicitly, as mentioned in Study 1 some of the included studies, did not use any of the eight argument criteria (Study 1, section x) to assess argumentation skills. Whilst these studies included the terms socio- and argumentation in their research questions, they analysed reasoning, rather than argumentation skills regarding SSI. Thus, they used reasoning criteria to assess SSA, such as reasoning modes (e.g. Wu, 2013; Wu & Tsai, 2007) or reasoning quality (e.g. Ozturk & Tuzun, 2018). As described in the General Introduction, these studies have a different focus in terms of learning goals, and as such, they apply different quality indicators and assessment criteria.

Another term that has been used to conceptualize SSA is “decision-making to solve an SSI” (e.g. Molinatti et al., 2010). According to Nielsen (2013), argumentative activities regarding SSI are open-ended scenarios on which a decision should be made. Thus, SSI tasks are tasks about “what to do” (Nielsen, 2013, p. 374). In that sense, SSA is presented as the arguments that take place during the negotiation of an open-ended scenario. This conceptualization could have respective implications on the assessment of SSA; the educational goal of a decision-making task would be student exposure to alternative solutions and different sources of information. At the same time, students might engage in authentic situations, where they have to make decisions based on the information that is available at the time. As a result, one would expect that according to this approach the argument criteria would be focus on argument moves and argument complexity.

Nevertheless, in Study 1 studies that conceptualized SSA as decision-making focused mainly on argument structure. For instance, Molinatti et al., 2010 examined SSA as a decision-making task in secondary school students regarding the use of embryonic stem cells in research and therapy (Molinatti, Girault, & Hammond, 2010). Whereas the task was presented as a debate, SSA was measured by means of structural criteria, such as the use of argument elements (e.g. claims, rebuttals, counter-arguments). This raises questions on the way in which the decision-making task and the pedagogical goals behind them were conceptualized.

4.2.4 Concepts and assessment tools from other research fields

Another issue that became apparent in Studies 1 and 2 is the use of concepts and assessment tools from research fields other than science education to operationalize and assess SSA. This issue is embedded in the study of argumentation in educational settings, as argumentation has a long tradition in linguistic and philosophical studies (Rapanta et al, 2013). As discussed in the General Introduction, there are two approaches in the use of frameworks to assess SSA: the adoption of traditional argumentation schemes and the use of adapted frameworks for measuring SSA. In what follows, the use of traditional argumentation schemes will be discussed, as this was the approach mainly followed by the selected studies.

Study 1 showed that more than half of the selected studies used original or adapted versions of already existing schemes. This finding is in line with previous research in science education, as the adaptation of theoretical frameworks for measurement purposes has been an issue of discussion in research related to argumentation in science education (e.g. Erduran, 2004; Jiménez-Aleixandre et al., 2000; Zohar & Nemet, 2002). These discussions have focused on how and in which contexts theoretical frameworks can be utilized in the form of argument assessment frameworks (Erduran, 2007).

An example of that approach is the adaption of TAP to measure argumentative competencies in science classrooms. Study 1 reported that Toulmin's TAP (1958) was most frequently applied by the selected studies in order to measure the quality of SSA by means of the structure of the produced arguments. Originally developed by philosopher Stephen Toulmin, TAP was firstly presented in the *Uses of an Argument* (1958) as a tool for analysing arguments with regards to seven argument components (claim, data, warrant, backing, qualifier, and rebuttal). Since then, TAP has been presented by studies in science education as a useful teaching tool to help students learn how to construct sound arguments (Robertshaw & Campbell, 2013). Because TAP puts emphasis on the use of seven elements to make an argument, it has been used as a tool to judge arguments' structural quality (e.g. Osborne et al., 2004).

While generally regarded as a useful teaching tool, the use of TAP as an assessment tool has received extensive criticism because it merely focuses on the structure of the argument overlooking other criteria, such as multidimensionality or the use of valid scientific evidence (Robertshaw & Campbell, 2013; Sadler & Fowler, 2006). For this reason, TAP has often been characterized as not suitable for assessing argumentation in complex issues, such as SSI (e.g. Lin & Minzes, 2010; Nielsen, 2012). Furthermore, the framework has been presented as not appropriate to capture dialogic features of argumentation (e.g. Nielsen, 2012), such as argument moves. Since, according to one perspective on argumentation in education, all argumentative activities are dialectic activities (Nielsen, 2012), one would expect researchers to choose other measurement tools to measure argumentation skills.

Study 2 showed that TAP was mainly used to assess structure, but in some cases, structure was assessed in conjunction to other criteria such as complexity or the use and accuracy of scientific evidence. An example of these studies is Lin et al (2012) in which a scoring rubric was used to assess college students' written arguments. The rubric consisted of five levels and arguments were scored based on the number of argument elements, as well as the accuracy of the used evidence. Thus, these studies could potentially serve as an example of how strictly structural assessment models can include more argument criteria.

4.3 Practical Implications

This section presented a summary of the practical implications of this thesis. The implications will mainly concern researchers and readers in the area of SSA, who wish to navigate in the literature regarding the issue. Study 1 emphasized the various approaches in which SSA has been conceptualized and measured. The coding categories developed to analyse the selected studies can serve as a guide for researchers and readers to better understand the procedures followed by researchers to conceptualize and measure socio-scientific argumentation. At the same time, one of the main findings in the present thesis is the lack of agreement over terms and definitions. Thus, in order to facilitate the use of a common language in research in the area of SSA, this study proposes the use of the coding categories and themes that emerged by the analysis of the studies as descriptive terms for conceptualization and measurement procedures in the area of SSA.

In terms of conceptualization, Study 1 presented three themes that were used to describe SSA:) SSA is argumentation regarding SSI b) SSA is negotiation and decision-making to solve an SSI and c) SSA is the expression of informal reasoning regarding SSI. The presentation of these themes can be used for the interpretation of definitions that are not clearly stated in literature in the area of SSA. At the same time, the themes can be used by future researchers in the area of SSA as a starting point for the conceptualization of SSA. Moreover, the categorization of conceptualizations into SSA specific and SSI specific promotes the use of an additional layer of theoretical grounding of the nature of SSA. By the use of the terms "SSA-specific" and "SSI-specific" researchers can express their views regarding the distinctive nature of SSA. Furthermore, in an attempt to promote operational definitions, the use of terms "quality indicators" can be used to describe the pre-defined argument elements that are important for the definition and analysis of SSA.

In terms of measurement, the study suggested the use of the eight argument assessment criteria for the categorization of selected studies. The emerged argument criteria were: (a) structure (b) use of pre-defined criteria (c) complexity (d) multidimensionality (e) use of scientific evidence (f) accuracy (g) coherence and (h) argument moves (for the descriptions of the criteria, see Study 1, Table 1). Study 2 demonstrated that argument models are commonly used with various foci, thus, these criteria can be applied for a systematic categorization of the applied argument models. In other words, the categorization of the argument assessment criteria can facilitate the investigation of "what is being assessed" rather than "what model was applied". Thus, to increase clarity and transparency in the measurement of SSA, current and future research in SSA may use similar terminology to describe their assessment criteria. Similarly, the categorization of research instruments into SSA-specific and SSI-specific can serve as a reference point for readers and researchers to distinguish between instruments that have been developed to assess other constructs (e.g. decision-making) and instruments that have been designed to assess specific features of SSA (SSA-specific instruments) or specific features of SSI (SSI-specific instruments).

4.4 Limitations

As in most studies in educational research, the results and conclusions that were drawn from the presented studies should be interpreted in light of certain limitations. The first limitation is the use of search terms "socio*" "scien*" and "argumentation" to include studies investigating SSA. In the Study 1, it was argued that there might exist studies that investigate SSA but do not use these specific terms. Nevertheless, since our aim was to clarify the term "socio-scientific argumentation", it was necessary to include only studies that used this specific terminology. To put it in another way, this thesis indented to include only the studies using the term '"socio-scientific" because, apart from the conceptualization of SSA as a construct, it aimed to describe the definitions and terms used with regard to this specific term.

Another limitation could be the inclusion of studies that used different types of sample, explicitly, teachers and university, as well as, secondary and primary school students. Although a preliminary analysis showed that researchers in the selected studies mainly used secondary and university students as a sample, it was decided not to exclude studies that used other types of samples. This was decided because the analysis of Study 1 and Study 2 could look into the role that the type of sample could potentially play in the choice of methodological approaches followed by the selected studies. It was assumed that researchers would select specific measurement tools and argumentative tasks based on the characteristics of the sample (e.g. written argumentation for university students and oral argumentation for primary school students). Nevertheless, findings in Study 1 showed that there was no differentiation between the type of participants.

A third limitation could be the use of a coding scheme developed by the authors for the needs of the present thesis. The coding scheme was mainly developed in a bottom-up process from codes and themes that emerged by the text. Although not previously validated, the development of the coding scheme, as well as the coding categories were presented in detail in the Methods section of Study 1. Lastly, another potential limitation would be that the country in which the studies were conducted, was not included in the analysis of conceptualization and measurement of SSA. As discussed earlier in this section, as well as the General Introduction, contextual factors, such as the country and the curricula may play a role in the way in which SSA is perceived, conceptualized and measured by researchers. Thus, we would encourage future research in the area of SSA to investigate the influence of contextual factors, such as the country or the national curricula, on the selection of measurement approaches to assess SSA.

4.5 Suggestions for Future Research

The aim of the present thesis was to present an overview of literature with regard to the conceptualization and measurement of socio-scientific argumentation in science education. The results of the presented studies revealed variations and inconsistencies in the ways in which SSA has been theoretically conceptualised and measured. These results are in line with previous studies on argumentation in science education highlighting the need for grounded methodology and a more systematic approach in the process of adapting theoretical frameworks into measurement tools (e.g. Erduran, 2007). In this section, some suggestions for future research will be discussed in light of the findings of the presented studies, as well as, the key challenges in the study of SSA.

Firstly, a general suggestion for the researchers in the area of SSA would be to explicitly justify the concepts used to theoretically describe SSA, as well as the measurement approaches that they follow to assess its quality of SSA. In doing so, researchers may encounter various difficulties, such as the ill-structured nature of the construct or the controversy of the specific SSI under investigation. As discussed earlier in this section, one way to overcome challenges with regard to “what counts as good argument” in SSA is to consider the educational and pedagogical goals of the activity that the researcher aims to introduce. When the educational goals are clear, the definition of what is expected to be measured (or change, in the case of intervention studies) will be easier to describe.

Secondly, in an attempt to improve the transparency as well as the coherence in processes of conceptualization and measurement. To help future studies in the area of SSA these processes have been summarized and presented in the form of six steps, which are illustrated in Figure 11. This framework was inspired by the methodological frameworks presented by Baxter and Babbie (2003) (Study 2, 2.5 Discussion) and Erduran (2007) (General Introduction).

More explicitly, the suggested methodological framework suggests that the conceptualization of a construct, in this case SSA, could start with the construction of a nominal definition of the construct. In this phase all the characteristics of the construct should be described, including distinctive features form other constructs (e.g. decision-making). The next step would be the construction of an operational definition, which will present all the measurement procedures and indicators that can theoretically be used to assess the construct. In this step researchers could clearly define what makes a good argument in SSI contexts and what are possible assessment approaches. As a result, the following step would be to determine specific quality indicators, argument components such as justification or rebuttals, to serve as “signals” for a good argument.

The construction of quality indicators would be the connection between what is theoretically defined as a good argument and what can (and will) be measured. Thus, quality indicators would lead to specific argument assessment criteria such as the argument’s structure or complexity. As a next step, the type of task, suitable for measuring the argument criteria, would be selected. The final step would be the selection of an argument assessment model, which would capture all the argument features that were defined in the earlier steps.

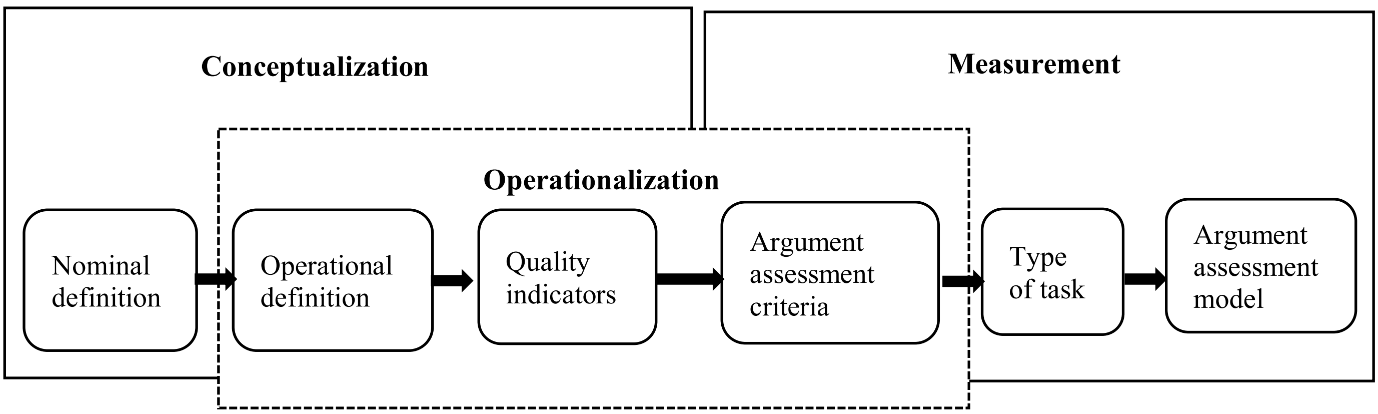

Figure 11

A Procedural Framework for the Conceptualization and Measurement of SSA

As discussed in the General Discussion, this thesis acknowledges that scientific processes are complex and cannot be easily described by linear models, similar to the “scientific method”. Nevertheless, whilst it may be difficult to follow a step-by-step procedure, this thesis suggests that procedural frameworks, like the one presented in Figure

11, will highlight important decisions that are made in studies investigating SSA. Thus, researchers are not expected to follow the procedure as a cookbook approach; they are rather encouraged to justify their decisions made within every step. At the same time, this thesis suggests that researchers should justify the omission of certain steps in the described process.

4.6 In Closing

Since its first appearance in science education research in 1999, socio-scientific argumentation has been a part of discussions in science education. As presented in this thesis, these discussions are relevant to the educational goals that science education aspires to achieve. In a world that is rapidly changing the need for scientific literacy skills for citizenship is commonly recognized. Thus, although research on socio-scientific argumentation has previously focused on the legitimization of SSA as a part of science curricula, the literature in the area of SSA is gradually evolving. In the last years a growing number of studies in the area of SSA follows intervention approaches, in an attempt to facilitate students' argumentative skills regarding controversial scientific issues. This creates an opportunity for discussion on what researchers and curricula mean by the term "socio-scientific argumentation". Subsequently, it raises questions on how the construct can be measured in science education classrooms. The two studies presented in this thesis aim to facilitate this discussion between researchers, policymakers and practitioners. Therefore, the studies contribute to the conceptualization and operationalization of SSA by promoting clarity and transparency in the theoretical and methodological processes presented in the selected studies.

References

Abd-El-Khalick, F. (2012). Nature of science in science education: Toward a coherent framework for synergistic research and development. In *Second international handbook of science education* (pp. 1041-1060). Springer, Dordrecht.

Abd-El-Khalick, F., & Lederman, N. G. (2000). The influence of history of science courses on students' views of nature of science. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *37*(10), 1057-1095.

Agell, L., Soria, V., & Carrió, M. (2015). Using role play to debate animal testing. *Journal of Biological Education*, *49*(3), 309-321.

Akerson, V. L., & Donnelly, L. A. (2008). Relationships among learner characteristics and preservice elementary teachers' views of nature of science. *Journal of Elementary Science Education*, *20*(1), 45-58.

Åkerblom, D., & Lindahl, M. (2017). Authenticity and the relevance of discourse and figured worlds in secondary students' discussions of socioscientific issues. *Teaching and Teacher Education*, *65*, 205-214.

Albe, V. (2008). When scientific knowledge, daily life experience, epistemological and social considerations intersect: Students' argumentation in group discussions on a socio-scientific issue. *Research in Science Education, 38*(1), 67-90.

Allen, M. (Ed.). (2017). *The sage encyclopedia of communication research methods* (Vols. 1–4). Thousand Oaks, CA: SAGE Publications, Inc.

American Association for the Advancement of Science. AAAS (1993) Benchmarks for science literacy: Project 2061.

American Association for the Advancement of Science. Project 2061 (AAAS). 1990. *Science for All Americans*.

Andriessen, J., Baker, M., & Suthers, D. D. (Eds.). (2013). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (Vol. 1). Springer Science & Business Media.

Barrue, C., & Albe, V. (2013). Citizenship Education and Socioscientific Issues: Implicit Concept of Citizenship in the Curriculum, Views of French Middle School Teachers. *Science & Education, 22*(5), 1089–1114.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science*, *43*(6), 737-766.

Bauer, H. H. (1994). *Scientific literacy and the myth of the scientific method*. University of Illinois Press.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765-793.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26-55.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*.

Branscomb, A. W. (1981). Knowing how to know. *Science, Technology, & Human Values*, *6*(3), 5-9.

Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education*, *92*(3), 473-498.

Baxter, L. A., & Babbie, E. R. (2003). *The basics of communication research*. Cengage Learning.

Cetin, P. S., Dogan, N., & Kutluca, A. Y. (2014). The quality of pre-service science teachers' argumentation: influence of content knowledge. *Journal of Science Teacher Education*, *25*(3), 309-331.

Chang, S. N. (2007, June). Teaching argumentation through the visual models in a resource-based learning environment. In *Asia-Pacific Forum on Science Learning and Teaching* (Vol. 8, No. 1, pp. 1-15). The Education University of Hong Kong, Department of Science and Environmental Studies.

Chang Rundgren, S.- N., & Rundgren, C.J. (2010, April). SEE-SEP: From a separate to a holistic view of socioscientific issues. In *Asia-Pacific Forum on Science Learning & Teaching* (Vol. 11, No. 1).

Chang, S. N., & Chiu, M. H. (2008). Lakatos' scientific research programmes as a framework for analysing informal argumentation about socio-scientific issues. International Journal of *Science Education, 30*(13), 1753-1773.

Chang, Y. H., Chang, C. Y., & Tseng, Y. H. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology*, *19*(4), 315-331.

Christenson, N., & Chang Rundgren, S. N. (2015). A Framework for Teachers' Assessment of Socio-Scientific Argumentation: An Example Using the GMO Issue. *Journal of Biological Education, 49*(2), 204–212.

Christenson, N., Chang Rundgren, S., & Höglund, H. O. (2012). Using the SEE-SEP model to analyze upper secondary students' use of supporting reasons in arguing socioscientific issues. *Journal of Science Education and Technology*, *21*(3), 342-352.

Christenson, N., Chang Rundgren, S. N., & Zeidler, D. L. (2014). The Relationship of Discipline Background to Upper Secondary Students' Argumentation on Socioscientific Issues. *Research in Science Education, 44*(4), 581–601.

Cinici, A. (2016). Balancing the pros and cons of GMOs: socio-scientific argumentation in pre-service teacher education. *International Journal of Science Education*, *38*(11), 1841-1866.

Craig-Hare, J., Ault, M., & Rowland, A. (2017). The effect of socioscientific topics on discourse within an online game designed to engage middle school students in scientific argumentation. *Journal of Education in Science Environment and Health*, *3*(2), 110-125.

Crawford, B., & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In *Research and the quality of science education* (pp. 309-323). Springer, Dordrecht.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, *39*(3), 124–130.

Crombie, A. C. (1994). *Styles of scientific thinking in the European tradition* (Vol. 3, pp. x-xi). London: Duckworth.\

Dawson, V., & Carson, K. (2017). Using climate change scenarios to assess high school students' argumentation skills. *Research in Science & Technological Education*, *35*(1), 1-16

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

DeBoer, G. E. (1991). Scientific literacy and the new progressivism. *A history of ideas in science education*, 173-189.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *37*(6), 582-601.

DeBoer, G. E. (Ed.). (2011). *The role of public policy in K12 science education*. IAP.

DeMarco, J. P. (1996). Moral theory: A contemporary overview. Boston: Jones and Bartlet

Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*. McGraw-Hill Education (UK).

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education, 84*(3), 287-312.

Doksæter Sivle, A., & Kolstø, S. D. (2016). Use of online weather information in everyday decision-making by laypeople and implications for communication of weather information. *Meteorological Applications*, *23*(4), 650-662.

Durant, J. (Ed.). (1992). *Museums and the public understanding of science*. NMSI Trading Ltd.

Duschl, R. A., Ellenbogen, K., & Erduran, S. (1999, April). Understanding dialogic argumentation among middle school science students. In *Invited paper at the annual meeting of the American Educational Research Association (AERA), Montreal.*

Duschl, R. A., Scweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking Science to School: learning and teaching science in Grades K-8*. Washington, DC: The National Academies Press.

Eggert, S., Nitsch, A., Boone, W. J., Nueckles, M., & Boegeholz, S. (2017). Supporting Students' Learning and Socioscientific Reasoning About Climate Change—the Effect of Computer-Based Concept Mapping Scaffolds. *Research in Science Education, 47*(1), 137–159.

Erduran, S., & Dagher, Z. R. (2014). Regaining focus in Irish junior cycle science: Potential new directions for curriculum and assessment on Nature of Science. *Irish Educational Studies*, *33*(4), 335-350.

Erduran, S., & Jiménez-Aleixandre, M. P. (2008). Argumentation in science education. *Perspectives from classroom-Based Research. Dordre-cht: Springer*.

Erduran, S., & Jiménez-Aleixandre, M. P. (2012). Argumentation in science education research: Perspectives from Europe. In *Science Education Research and Practice in Europe* (pp. 253-289). Brill Sense.

Erduran, S., Ozdem, Y., & Park, J. Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998–2014. *International Journal of STEM Education*, *2*(1), 5.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, *88*(6), 915–933.

Evans, J. S. B. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological bulletin*, *128*(6), 978.

Evans, J. S. B., & Thompson, V. A. (2004). Informal reasoning: Theory and method. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *58*(2), 69.

Faize, F. A., Husain, W., & Nisar, F. (2017). A Critical Review of Scientific Argumentation in Science Education. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 475–483.

Fang, S. C., Hsu, Y. S., & Lin, S. S. (2019). Conceptualizing Socioscientific Decision Making from a Review of Research in Science Education. *International Journal of Science and Mathematics Education, 17*(3), 427–448.

Feierabend, T., Stuckey, M., Nienaber, S., & Eilks, I. (2012). Two Approaches for Analyzing Students' Competence of" Evaluation" in Group Discussions about Climate Change. *International Journal of Environmental and Science Education, 7*(4), 581-598.

Felton, M. K. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*(1), 35-52.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... & Strijbos, J. W. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research, 2(3)*, 28-45.

Foong, C. C., & Daniel, E. G. (2013). Students' argumentation skills across two socio-scientific issues in a Confucian classroom: Is transfer possible?. *International Journal of Science Education*, *35*(14), 2331-2355.

Foschi, M. (2014). Hypotheses, operationalizations, and manipulation checks. In *Laboratory experiments in the social sciences* (pp. 247-268). Academic Press.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of educational research*, *82*(3), 300-329.

Giere, R. N., Bickle, J., & Mauldin, R. F. (2006). *Understanding scientific reasoning.* Independence, KY: Wadsworth/ Cengage Learning.

Gilbert, J. (2005). *Catching the knowledge wave?: The knowledge society and the future of education*. Wellington: Nzcer Press.

Gott, R., & Duggan, S. (2002). Problems with the assessment of performance in practical science: which way now?. *Cambridge Journal of Education*, *32*(2), 183-201.

Grooms, J., Sampson, V., & Golden, B. (2014). Comparing the effectiveness of verification and inquiry laboratories in supporting undergraduate science students in constructing arguments around socioscientific issues. *International Journal of Science Education, 36*(9), 1412-1433.

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science teaching*, *28*(9), 799-822.

Hansson, L., Redfors, A., & Rosberg, M. (2011). Students' socio-scientific reasoning in an astrobiological context during work with a digital learning environment. *Journal of Science Education and Technology*, *20*(4), 388-402.

Hartmann, S., zu Belzen, A. U., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education. *Zeitschrift für Psychologie*.

Herring, J. P. (1918). Measurement of some abilities in scientific thinking. *Journal of Educational Psychology*, *9*(10), 535.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.

Institute of Education Sciences and National Science Foundation. (2013). *Common guidelines for education research and development*. Retrieved from http://ies.ed.gov/pdf/CommonGuidelines.pdf

Iordanou, K., & Constantinou, C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education, 99*(2), 282-311.

Isbilir, E., Cakiroglu, J., & Ertepinar, H. (2014). Pre-Service Science Teachers' Written Argumentation Qualities: From the Perspectives of Socio- Scientific Issues, Epistemic Belief Levels and Online Discussion Environment. *EURASIA Journal of Mathematics, Science & Technology Education, 10*(5), 371-381.

Jiménez-Aleixandre, M. P. (2002). Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education*, *24*(11), 1171-1190.

Jiménez-Aleixandre, M. P., & Pereiro-Muñoz, C. (2005). Argument construction and change while working on a real environment problem. In *Research and the quality of science education* (pp. 419-431). Springer, Dordrecht.

Jin, H., Hokayem, H., Wang, S., & Wei, X. (2016). A US-China interview study: Biology students' argumentation and explanation about energy consumption issues. *International Journal of Science and Mathematics Education*, *14*(6), 1037-1057.

Jonsson, A. (2016). Student performance on argumentation task in the Swedish National Assessment in science. *International Journal of Science Education, 38*(11), 1825–1840.

Juntunen, M. K., & Aksela, M. K. (2014). Improving students' argumentation skills through a product life-cycle analysis project in chemistry education. *Chemistry Education Research and Practice*, *15*(4), 639-649.

Justi, R., & Gilbert, J. (2003). Teachers' views on the nature of models. *International Journal of science education*, *25*(11), 1369-1386.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of educational Measurement*, *38*(4), 319-342.

Karahan, E., & Roehrig, G. (2017). Secondary school students' understanding of science and their socioscientific reasoning. *Research in Science Education*, *47*(4), 755-782.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science education*, *86*(3), 314-342.

Kelly, G. J., Drucker, S., & Chen, K. (1998). Students' reasoning about electricity: Combining performance assessment with argumentation analysis. *International Journal of Science Education, 20*(7), 849-871.

Keskin, M. O., Samanci, N. K., & Yaman, H. (2013). Argumentation based bioethics education: Sample implementation on Genetically Modified Organisms (GMOs) and genetic screening tests. *Educational Research and Reviewes*, *8*(16), 1383-1391.

Khishfe, R. (2012). Relationship between nature of science understandings and argumentation skills: A role for counterargument and contextual factors. *Journal of Research in Science Teaching*, *49*(4), 489-514.

Khishfe, R., Alshaya, F. S., BouJaoude, S., Mansour, N., & Alrudiyan, K. I. (2017). Students' understandings of nature of science and their arguments in the context of four socio-scientific issues. *International Journal of Science Education, 39*(3), 299–334.

Kim, M., Anthony, R., & Blades, D. (2014). Decision making through dialogue: A case study of analyzing preservice teachers' argumentation on socioscientific issues. *Research in Science Education, 44*(6), 903-926.

Kind, P. E. R., & Osborne, J. (2017). Styles of scientific reasoning: A cultural rationale for science education?. *Science Education*, *101*(1), 8-31.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological science*, *15*(10), 661-667.

Kohle, F. H. (2018). The Social Media "Information Explosion" Spectacle: Perspectives for Documentary Producers. In *Digital Multimedia: Concepts, Methodologies, Tools, and Applications* (pp. 307-321). IGI Global.

Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science education, 85*(3), 291-310.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. MIT Press.

Krippendorff, K. (2004). Reliability in content analysis. *Human communication research*, *30*(3), 411-433.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science education, 77*(3), 319-337.

Kuhn, D. (2005). *Education for thinking*. Harvard University Press.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills?. *Cognition and Instruction*, *26*(4), 512-559.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science*, *9*(5), 178-181.

Kultusministerkonferenz. (2004). Bildungsstandards im Fach Physik fuer den Mittleren Schulabschluss [Standards in physics for the middle-school graduation]. Retrieved from http://www.gesetze-bayern.de/Content/Document/BayLPO_I-77

Kutluca, A. Y., & Aydin, A. (2016). An Examination of Prospective Elementary Science Teachers' Perspective towards Socio-Scientific Argumentation. *Science Education International, 27*(3), 320–343.

Laius, A., & Rannikmäe, M. (2011). Impact on student change in scientific creativity and socio-scientific reasoning skills from teacher collaboration and gains from professional in-service. *Journal of Baltic Science Education*, *10*(2), 127-137.

Lakatos, I. (1978). The Methodology of Scientific Research Programmes. In J. Worrall & G. Currie (Eds.), *Philosophical Papers: Vol. 1*. Cambridge: Cambridge University Press.

Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *The Lancet*, *378*(9790), 526-535.

Larson, H., de Figueiredo, A., Karafillakis, E., & Rawal, M. (2018). State of vaccine confidence in the EU 2018. *Luxembourg: Publications Office of the European Union https://ec. europa.eu/health/sites/health/files/vaccination/docs/2018_vaccine_confidence_en. pdf*.

Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science education*, *84*(1), 71-94.

Lawless, K. A., & Brown, S. W. (2015). Developing scientific literacy skills through interdisciplinary, technology-based global simulations: GlobalEd 2. *Curriculum Journal*, *26*(2), 268-289.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International journal of science education*, *25*(11), 1387-1408.

Lee, C. Q., & She, H. C. (2010). Facilitating students' conceptual change and scientific reasoning involving the unit of combustion. *Research in Science Education*, *40*(4), 479-504.

Lee, M. H., Wu, Y. T., & Tsai, C. C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education*, 31(15), 1999–2020.

Lin, H. S., Hong, Z. R., & Lawrenz, F. (2012). Promoting and scaffolding argumentation through reflective asynchronous discussions. *Computers & Education, 59*(2), 378-384.

Lin, T. C., Lin, T. J., & Tsai, C. C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education*, *36*(8), 1346-1372.

Lin, S. S., & Mintzes, J. J. (2010). Learning argumentation skills through instruction in socioscientific issues: The effect of ability level. *International Journal of Science and Mathematics Education*, *8*(6), 993-1017.

Lindahl, M. G., & Linder, C. (2013). Students' ontological security and agency in science education—an example from reasoning about the use of gene technology. *International Journal of Science Education*, *35*(14), 2299-2330.

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and instruction*, *14*(2), 139–178.

Mechler, H. S., & Thoma, S. J. (2013). Moral development theory: Neo-Kohlbergian theory.

Memis, E. K., & Cevik, E. E. (2016). Examination of Students' Small Groups Discussion in Argumentation Process: Scientific and Socio-Scientific Issues. *JOURNAL OF EDUCATION IN SCIENCE ENVIRONMENT AND HEALTH*, *3*(2), 126-137.

Mercier, H. (2011). On the universality of argumentative reasoning. *Journal of Cognition and Culture, 11*(1-2), 85-113.

Millar, R., & Osborne, J. (1998). Beyond 2000. *Science education for the future*.

Miller, J. D. (1993). Theory and measurement in the public understanding of science: a rejoinder to Bauer and Schoon. *Public Understanding of Science, 2*(3), 235-243.

Mitchell, S. (1996). Improving the quality of argument in higher education interim report. London: Middlesex University, School of Education

Molinatti, G., Girault, Y., & Hammond, C. (2010). High school students debate the use of embryonic stem cells: The influence of context on decision-making. *International Journal of Science Education*, *32*(16), 2235-2251.

Morin, O., Simonneaux, L., & Tytler, R. (2017). Engaging with socially acute questions: Development and validation of an interactional reasoning framework. *Journal of Research in Science Teaching, 54*(7), 825-851.

Mullis, I. V., Martin, M. O., Minnich, C. A., Stanco, G. M., Arora, A., Centurino, V. A., & Castle, C. E. (2012). *TIMSS 2011 Encyclopedia: Education Policy and Curriculum* *in Mathematics and Science. Volume 1: AK*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Nam, Y., & Chen, Y. C. (2017). Promoting argumentative practice in socio-scientific issues through a science inquiry activity. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(7), 3431–3461.

Namdar, B. (2017). Preservice Science Teachers' Collaborative Knowledge Building through Argumentation on Healthy Eating in a Computer Supported Collaborative Learning Environment. *Turkish Online Journal of Educational Technology-TOJET*, *16*(3), 132-146.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting concepts, and Core Ideas*. Washington, DC: The National Academies Press.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of science education*, *21*(5), 553-576.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Nielsen, J. A. (2012). Science in Discussions: An Analysis of the Use of Science Content in Socioscientific Discussions. *Science Education, 96*(3), 428–456.

Nielsen, J. A. (2013). Dialectical features of students' argumentation: A critical review of argumentation studies in science education. *Research in Science Education*, *43*(1), 371–393.

Nussbaum, E. M. (2002). Scaffolding argumentation in the social studies classroom. *The Social Studies*, *93*(2), 79-83.

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive science*, *21*(2), 109-146.

Organisation for Economic Co-operation and Development. (2017). OECD (2017), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. Paris: OECD Publishing,

Osborne, J. (2007). Science Education for the Twenty First Century. *Eurasia Journal of Mathematics, Science & Technology Education*, *3*(3).

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463-466.

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking skills and creativity*, *10*, 265-279.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching, 50*(3), 315-347.

Ozturk, N., & Yilmaz-Tuzun, O. (2017). Preservice science teachers' epistemological beliefs and informal reasoning regarding socioscientific issues. *Research in Science Education*, *47*(6), 1275-1304.

Patronis, T., Potari, D., & Spiliotopoulou, V. (1999). Students' argumentation in decision-making on a socio-scientific issue: implications for teaching. *International Journal of Science Education*, *21*(7), 745-754.

Pella, M. O., O'hearn, G. T., & Gale, C. W. (1966). Referents to scientific literacy. *Journal of Research in Science Teaching*, *4*(3), 199-208.

Perelman, C., & Olbrechts-Tyteca, L. (1969). The new rhetoric: A treatise on argumentation. 1958. *Trans. John Wilkinson and Purcell Weaver. Notre Dame: U of Notre Dame P*.

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, *83*(4), 483-520.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. Y. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2-3), 155-175.

Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.). *Handbook of research on science education* (pp. 729–780).

Robertshaw, B., & Campbell, T. (2013). Constructing arguments: Investigating pre-service science teachers' argumentation skills in a socio-scientific context. *Science Education International*, *24*(2), 195-211.

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

Rudsberg, K., Öhman, J., & Östman, L. (2013). Analyzing students' learning in classroom discussions about socioscientific issues. *Science Education*, *97*(4), 594-620.

Rundgren, C.-J., Eriksson, M., & Rundgren, S.-N. C. (2016). Investigating the Intertwinement of Knowledge, Value, and Experience of Upper Secondary Students'

Argumentation Concerning Socioscientific Issues. *Science & Education, 25*(9–10), 1049–1071.

Saad, M. I. M., Baharom, S., & Mokhsein, S. E. (2017). Scientific reasoning skills based on socio-scientific issues in the biology.

Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(8), 909-921.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*(5), 513–536.

Sadler, T. D., & Donnelly, L. A. (2006). Socioscientific Argumentation: The effects of content knowledge and morality. *International Journal of Science Education, 28*(12), 1463–1488.

Sadler, T. D., & Fowler, S. R. (2006). A threshold model of content knowledge transfer for socioscientific argumentation. Science Education, 90(6), 986–1004.

Sadler, T. D., & Zeidler, D. L. (2004). The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. *Science education*, *88*(1), 4-27.

Sadler, T. D., & Zeidler, D. L. (2005a). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching, 42*(1), 112-138.

Sadler, T. D., & Zeidler, D. L. (2005b). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, *89*(1), 71-93.

Sadler, T. D., Barab, S. A., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry?. *Research in Science Education*, *37*(4), 371-391.

Sadler, T. D., Chambers, F. W., & Zeidler, D. L. (2002, April). *Investigating the Crossroads of Socioscientific Issues, the Nature of Science, and Critical Thinking*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA.

Sadler, T.D. (2011). Situating Socio-scientific Issues in Classrooms as a Means of Achieving Goals of Science Education. In Sadler T. (Ed.), *Socio-scientific Issues in the Classroom: Contemporary Trends and Issues in Science Education* (Vol. 39, pp 1–9). Dordrecht: Springer.

Sakschewski, M., Eggert, S., Schneider, S., & Bögeholz, S. (2014). Students' Socioscientific Reasoning and Decision-making on Energy-related Issues—Development of a measurement instrument. *International Journal of Science Education*, *36*(14), 2291-2313.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*(3), 345-372.

Shamos, M. H. (1995). *The myth of scientific literacy*. Rutgers University Press.

Shen, B. S. (1975). Science literacy and the public understanding of science. In *Communication of scientific information* (pp. 44-52). Karger Publishers.

Simon, S., Osborne, J., & Erduran, S. (2003). Systemic teacher development to enhance the use of argumentation in school science activities. *Leadership and professional development in science education: New possibilities for enhancing teacher learning*, 198-217.

Simonneaux, L., & Chouchane, H. (2011). The reasoned arguments of a group of future biotechnology technicians on a controversial socio-scientific issue: human gene therapy. *Journal of Biological Education*, *45*(3), 150-157.

Simonneaux, L., & Simonneaux, J. (2009). Students' socio-scientific reasoning on controversies from the viewpoint of education for sustainable development. *Cultural studies of science Education*, *4*(3), 657-687.

Sjöström, J., & Eilks, I. (2017). Reconsidering different visions of scientific literacy and science education based on the concept of Bildung. *Cognition, Metacognition, and Culture in STEM Education: Learning, Teaching and Assessment*, *24*, 65.

Smith, K. (2016). *Working Scientifically: A guide for primary science teachers*. Routledge.

Solli, A., Bach, F., & Åkerman, B. (2014). Learning to argue as a biotechnologist: disprivileging opposition to genetically modified food. *Cultural studies of science education*, *9*(1), 1-23.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques*. Thousand Oaks, CA: Sage publications.

Tal, T., & Kedmi, Y. (2006). Teaching socioscientific issues: Classroom culture and students' performances. *Cultural Studies of Science Education*, *1*(4), 615–644.

Teig, N., & Scherer, R. (2016). Bringing Formal and Informal Reasoning Together—A New Era of Assessment?. *Frontiers in psychology*, *7*, 1097.

Tekin, N., Aslan, O., & Yilmaz, S. (2016). Research Trends on Socioscientific Issues: A Content Analysis of Publications in Selected Science Education Journals. *Journal of Education and Training Studies, 4*(9), 16–24.

Thomson, B. S. (2004, September). Qualitative research: Grounded theory—Sample size and validity. In *Faculty of Business and Economics 10th Annual Doctoral Conference October* (pp. 25-27).

Topcu, M. S., Sadler, T. D., & Yilmaz-Tuzun, O. (2010). Preservice science teachers' informal reasoning about socioscientific issues: The influence of issue context. *International Journal of Science Education*, *32*(18), 2475-2495.

Toulmin, S. (1958). *The uses of argument*. Cambridge, MA: Cambridge University Press.

Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, *47*(8), 952-977.

Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *45*(1), 101-131.

Walker, K. A., & Zeidler, D. L. (2007). Promoting discourse about socioscientific issues through scaffolded inquiry. *International Journal of Science Education, 29*(11), 1387–1410.

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, *3*(2), 169-184.

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Erlbaum.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human behavior*, *26*(4), 506-515.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science education*, *92*(5), 941-967.

Wold Health Organization. (2019, August). Measles in the WHO European region. Retrieved from: http://www.euro.who.int/__data/assets/pdf_file/0010/410779/Measles-Sitrep-2-Aug-2019_.pdf?ua=1

Wu, Y. T. (2013). University students' knowledge structures and informal reasoning on the use of genetically modified foods: Multidimensional analyses. *Research in Science Education*, *43*(5), 1873-1890.

Wu, Y. T., & Tsai, C. C. (2007). High school students' informal reasoning on a socio-scientific issue: Qualitative and quantitative analyses. *International Journal of Science Education, 29*(9), 1163-1187.

Yager, R. E. (Ed.). (1996). *Science/technology/society as reform in science education*. Albany: SUNY Press.

Yang, F. Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: The role of scientific thinking in environmental science decision-making. *International Journal of Science Education*, *26*(11), 1345-1364.

Yu, S. M., & Yore, L. D. (2013). Quality, evolution, and positional change of university students' argumentation patterns about organic agriculture during an argument–

critique–argument experience. *International Journal of Science and Mathematics Education*, *11*(5), 1233-1254.

Zeidler, D. L., Herman, B. C., Ruzek, M., Linder, A., & Lin, S. S. (2013). Cross-cultural epistemological orientations to socioscientific issues. *Journal of Research in Science Teaching*, *50*(3), 251-283.

Zeidler, D. L., & Keefer, M. (2003). The role of moral reasoning and the status of socioscientific issues in science education. In D. L. Zeidler (Ed.), *The role of moral reasoning on socioscientific issues and discourse in science education* (pp. 7–38). Dordrecht: Springer.

Zeidler, D. L., Osborne, J., Erduran, S., Simon, S., & Monk, M. (2003). The role of argument during discourse about socioscientific issues. In *The role of moral reasoning on socioscientific issues and discourse in science education* (pp. 97-116). Springer, Dordrecht.

Zeidler, D. L., Walker, K. A., Ackett, W. A., & Simmons, M. L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science education*, *86*(3), 343–367.

Zimmerman, C. (2005). The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning. *Final draft of a report to the National Research Council Committee on Science Learning Kindergarten through Eighth Grade. Washington, DC: National Research Council*.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 39*(1), 35-62.

Appendix

Codes and Sub-codes Used in Study 2

Code	Sub-code
(3) Conceptualized quality indicators	3.1 & 6.1 Complexity
&	3.2 & 6.2 Counter-argument
(6) Applied quality indicators	3.3 & 6.3 Justification
	3.4 & 6.4 Knowledge
	3.5 & 6.5 Rebuttal
	3.6 & 6.6 Use/quality of evidence
	3.7 & 6.7 Other
(7) Argument assessment criteria	7.1 Accuracy of evidence
	7.2 Argument moves
	7.3 Coherence
	7.4 Complexity
	7.5 Multidimensionality
	7.6 Structure
	7.7 Use of predefined argument
	7.8 Use of scientific evidence
	7.9 Other
(8) Argument assessment models	8.1 Duschl et al. (2007)
	8.2 Erduran et al. (2004)
	8.3 Kuhn (1991)
	8.4 Lakatos (Chang & Chiu, 2008)
	8.5 Sadler & Fowler (2006)
	8.6 SEE-SEP (Chang Rundgren & Rundgren 2010)
	8.8 Toulmin (1958)
	8.9 Walton (2006)
	8.10 Zohar & Nemet (2002)
	8.11 Other