Foregrounding the Background: Investigating Science Teachers' Practices in Assessing Students' Understandings of Nature of Science

Wonyong Park

St Antony's College

University of Oxford

Supervised by

Professor Sibel Erduran and Dr Judith Hillier

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

TO ALL MY TEACHERS

FOR MAKING ME WHO I AM TODAY

Table of contents

List of figures

List of tables

Abstract

The inclusion of nature of science (NOS) in science curricula has long been advocated as a key means to accomplish scientific literacy. For example, in the U.S., the *Science for All Americans* (AAAS, 1989) and *Benchmarks for Science Literacy* (AAAS, 1993) included an explicit emphasis on the teaching of NOS. More recently, *A Framework for K-12 Science Education* (NRC, 2012) also highlighted that 'there is a strong consensus about characteristics of the scientific enterprise that should be understood by an educated citizen' (p. 78). While these and other science curriculum documents around the world have emphasised the inclusion of NOS since the 1990s, there is currently a lack of empirical evidence on teachers' responses to curriculum reforms focused on NOS and particularly on their views of and practices in the classroom assessment of NOS.

In an attempt to fill this gap, this study investigates Korean science teachers' practices of classroom assessment in the wake of the recent national curriculum reform that foregrounds NOS as an explicit learning expectation. In the 2015 National Curriculum of Korea, a new subject named 'scientific inquiry and experimentation' (SIE) was introduced as a separate compulsory subject and is being taught to Grade 10 students. SIE features NOS, which has epistemic characteristics distinguished from those of scientific content knowledge, as one of its main learning goals. To understand how teachers approach the assessment of this NOS-centred curriculum, I focus on one school where three teachers taught and assessed NOS through a number of different assessment tasks. Interviews, classroom observations, assessment materials and various lesson artefacts collected over two years are used to illustrate how the three teachers viewed and practised the classroom assessment of NOS.

Several important aspects of the teachers' views and practices relating to the classroom assessment of NOS are reported. First, I discuss the mismatch that existed between teachers' NOS teaching intentions and NOS assessment intentions, and between their views and enactments of the summative assessment of NOS. The three teachers enacted the same set of NOS summative assessment activities in different ways, and their enactments were associated with their perceived purposes and foci of NOS assessment. Second, the teachers' collaborative decision making about NOS assessment revealed several challenges of NOS summative assessment, such as distinguishing ‘understandings’ and ‘views’ of NOS, ensuring fairness among different enactments of assessment tasks, and controlling the literacy and content knowledge demand when assessing NOS. Third, an in-depth analysis of one teacher’s NOS instruction suggested that her feedback and other formative assessment practices mediated the NOS learning goal of the lesson with science content knowledge, history of science and other aspects of NOS, and built connections between these elements. Throughout the lesson, the teacher created NOS learning opportunities by eliciting, responding to and acting upon evidence of student learning.

Given the centrality of high-quality classroom assessment in student learning, the findings will be useful to inform future curriculum and assessment reforms as well as to better support teachers for teaching and assessing NOS. Based on the existing literature and the findings of this study, I assert that supporting teachers for meaningful and effective classroom assessment of NOS is key to empowering students to develop an elaborate understanding of what science is and how it works.

Word count: 94,655

Acknowledgements

This thesis could not have come together without the intellectual and emotional support of many people. I thank Sibel Erduran and Judith Hillier, who have been amazing supervisors, mentors, colleagues and collaborators during the past three years, in a way that I hope to emulate one day. I am also grateful to my examiners, Alison Cullinane and Jim Ryder, along with earlier examiners Ann Childs and Ian Thompson, whose feedback, care and encouragement have strengthened the content of the thesis.

I am proud to have been part of the science education team at Oxford as well as the OARS project that has coincided with my doctoral study. Working with colleagues, including Liam, Olga, Nigel and Jessica, gave me invaluable learning opportunities that helped to shape and develop the thesis. I am also grateful to an unending list of friends and colleagues within and outside the Department of Education at Oxford, whose encouragement and wishes helped me finish this thesis amid a global pandemic. I am immeasurably indebted to my three participants, Dr Kim, Ms Nam and Ms You (pseudonyms), who graciously allowed me to learn from their classrooms and share their wisdom with the science education community through this thesis.

I am deeply grateful to Ilju Academy and Culture Foundation and the Department of Education at Oxford for their financial support during my doctoral study. My greatest love and thanks go to Mum, Dad and Brother and my extended family with whom I have shared this journey albeit remotely. On their unwavering support, patience and trust I could build this thesis.

Chapter 1 Introduction

In recent decades, many science curricula and science education reform documents around the globe commonly have stated nature of science (NOS) as a curriculum goal and content (AAAS, 1989; MEST, 2009; Millar & Osborne, 1998; MOE, 2015; NRC, 1996, 2012; The Royal Society Science Policy Centre, 2014). The idea of NOS education is to teach students not only the conceptual and procedural knowledge about science but also the epistemic and social aspects of it that are informed by research in the history, philosophy, sociology and anthropology of science. Although NOS, broadly speaking, is about 'how the scientific enterprise works' (McComas, 2020, p. 5), it would be only a slight exaggeration to say that every NOS researcher defines NOS in different ways. According to this 'consensus' view, NOS refers to science as a way of knowing, epistemology of science, or the values and beliefs underlying scientific knowledge or the development of scientific knowledge (Lederman, Abd-El-Khalick, Bell & Schwartz, 2002). On the other hand, there are NOS researchers who subscribe to a broader conception of NOS that encompasses the characteristics of not only scientific knowledge but also scientific inquiry and practice (Duschl & Grandy, 2013; Erduran & Dagher, 2014a; Hodson & Wong, 2014). This recent view also acknowledges the differences among scientific disciplines and argues for a more philosophically sophisticated approach that presents a 'more authentic view of contemporary scientific practice' (Hodson & Wong, 2017, p. 3). The term NOS in this thesis is used in this latter sense, following a recent body of criticisms about the consensus view that it undermines the interdependence of scientific knowledge and the processes and activities through which it is generated, and that there are many ambiguities in how it characterises science (Duschl & Grandy, 2013; Hodson & Wong, 2017; Irzik & Nola, 2011; Matthews, 2012).

Aside from the debate over the meaning and coverage of the term, it is evident that understanding NOS has shifted 'from a desirable goal to being a central one for achieving scientific literacy' (Dagher & BouJaoude, 2005, p. 378) over the past two decades. A scientifically literate person, as set out in the *Science for All Americans* (AAAS, 1989), 'is aware that science, mathematics and technology are interdependent human enterprises with strengths and limitations; understands key concepts and principles of science; is familiar with the natural world and recognises both its diversity and unity; and uses scientific knowledge and scientific ways of thinking for individual and social purposes' (p. 4). Most recently, the *PISA 2024 Strategic Direction and Vision for Science* also reiterated the importance of understanding 'how scientific knowledge develops over time' (p. 3), particularly with regard to tackling the worrying societal challenges such as climate change denial and vaccine hesitancy (OECD, 2020). NOS has been argued to be an integral part of scientific literacy in that it brings benefits to individuals by facilitating rational and informed decision-making, and also to society by increasing moral, democratic and responsible citizenship (Driver, Leach, Millar, & Scott, 1997; Hodson, 2014). Emerging as a core aim of primary and secondary science education (Hodson, 2014; Lederman, 2007), NOS has become increasingly addressed in the science curriculum reforms worldwide and research literature.

Although the value of NOS in science curricula is now widely accepted among science education researchers, evidence suggests some limitations in incorporating NOS in science teaching. Most notably, despite the continued research efforts over the 60 years, teachers and students still have inadequate and naïve understandings about NOS (Khishfe & Abd-El-Khalick, 2002; Lederman et al., 2019; Schwartz, Lederman, & Crawford, 2004). To a certain extent, this is connected to the fact that NOS has been included in the secondary science curriculum in a rather indirect and implicit manner (McComas & Olson, 2002; Olson, 2018; Summers, Alameh, Brunner, Maddux, Wallon & Abd-El-Khalick, 2019). Even today, science curricula continue to be organised around the structure of disciplinary knowledge in physics, chemistry, biology and (sometimes) earth science (Duschl, 2008); Most times, NOS is no more than a higher-level goal or manifesto that only exists in writing. In a recent review of recent international science education standards documents from nine countries, Olson (2018) found that NOS ideas are seldom specified as learning expectations but are mostly in ancillary materials, and lamented that '… the status of the nature of science in the international science education standards documents studied here appears disjointed and rather uninformed by scholarship in science education' (p. 656). As Olson rightly pointed out, it is hard to find secondary science curricula that foreground NOS explicitly as a learning expectation. Summers and colleagues' (2019) analysis of 96 science standards in the U.S. released between 1980 and 2016 again confirmed that there has been surprisingly little improvement with regard to the coverage of NOS and the way NOS aspects are presented. In particular, it is often criticised that the U.S. *Next*

Generation Science Standards (NGSS Lead States, 2013) includes no strongly stated learning expectations related to NOS in the body of the document (Akerson, Carter, Pongsanon & Nargund-Joshi, 2019).

There seem to be at least two possible reasons that impede NOS from being considered as core curricular elements. The first obvious difficulty is that understanding NOS often requires some prior content knowledge and inquiry skills about science. Consider a frequently mentioned NOS aspect that 'scientific knowledge is fallible and subject to change'. This NOS aspect could be exemplified by the theory change from a geocentric to a heliocentric worldview, or the development of atomic models, among other historical stories. Both cases, however, require some knowledge of astronomy (e.g., the solar system), or of physics and chemistry (e.g., atomic theory) to understand why the change evidences the intended NOS aspect. Similarly, to understand that scientists use diverse methods to investigate nature, students first need to grasp what each of the methods (e.g., inductive reasoning, experimentation, observation) entails through studying historical or contemporary cases and/or engaging in these methods in practical lessons. This means that it is crucial to teach NOS in close connection with scientific content knowledge and inquiry, and such contextualisation is one of the key issues that determine the success of introducing any NOS-informed science curricula (Allchin, Andersen & Nielsen, 2014; Bell, Mulvey & Maeng, 2016).

Another major challenge to implementing a NOS-focused curriculum is tied to the issue of assessment. This is unsurprising given that assessment plays a critical role in the way a curriculum is presented by teachers and also what students learn from it (Black, Harrison, Hodgen, Marshall & Serret, 2010; Duschl & Gitomer, 1997; Popham, 2009). The power of the assessment system in determining educational practice and attainment is particularly strong in certain countries, for example, England (Preece & Skinner, 1999), China (Ma, 2012), and finally South Korea (hereafter referred to as Korea) (Song, 2013)—the focus of this thesis. Korea has historically relied much on tests to determine students' success and shape social hierarchy by allowing them into elite schools and prestigious careers based on test results, and as a result, has developed particular perspectives on assessment-related issues such as fairness, discrimination, accountability and the role of assessment in the education system (Kwon, Lee & Shin, 2017). Given the importance of assessment in the education system in Korea, it seems critical to investigate the assessment aspect in order to understand how NOS is conceived and taught in Korean science classrooms. I will revisit this issue of the exam-centred educational system and culture in Korea later in Chapter 3.

Assessment, particularly teachers' classroom assessment, is an underexplored area in the current literature in NOS education. Although there is a plethora of research instruments developed for measuring students' understanding of NOS (Aikenhead & Ryan, 1992; Kaya, Erduran, Aksoz, & Akgun, 2019; Lederman et al., 2002) that are widely used, these researcher-developed instruments have limited applications when it comes to in-class formative and summative assessment, due to the fundamental difference in their aims. Research instruments and classroom assessments each have their distinct aims and measures of quality, and are conducted by and reported to different parties (Allchin, 2013; Hanuscin, Lee & Akerson, 2011). Such a critical difference in the aims of assessment is particularly problematic for the summative assessment of NOS. Cheung (2020) noted the difficulties of setting up a marking scheme for assessing NOS understandings with reliability and discriminatory power sufficient for high-stakes examinations and called for science educators' further discussion on what and how NOS should be assessed in summative assessments. This problem of assessment, intertwined with the highly competitive culture of education, is in fact a central issue in the introduction of NOS in the Korean science curriculum, which will be further developed in the rest of this thesis.

Bringing together the importance of assessment in the way curriculum is taught (Duschl, 2008; Duschl & Gitomer, 1997) and the lack of research on teachers' assessment of NOS, there is a clear call for research on the intersection of classroom assessment and NOS. The context of this thesis, that is, the new Korean science curriculum, is distinguished from its predecessors in that it brings NOS from the background to the fore and sets out a strong expectation that students' understanding of NOS is explicitly taught and assessed by teachers and reported to the school and parents. From a broader international perspective, this study can be seen as an examination of an 'extreme case' of NOS-focused science curricula, in the sense that having a subject dedicated to NOS is an extreme way of making it explicit. By investigating a recent case of curriculum reform in South Korea, the findings from this thesis will provide insights into the opportunities and challenges of teaching and assessing NOS on the classroom level, especially viewed from teachers' perspectives. In doing so, implications will be drawn as to how to improve the curriculum, teacher education and other systematic supports to successfully implement NOS-focused curriculum reforms. To these aims, this thesis is structured in the following manner.

Chapter 2 provides a review of relevant literature and develops a theoretical orientation for the study. I start by considering how the idea of NOS first emerged and has made its current standing in science education research and policy, and then examine the recent theories and empirical evidence on the aims, teaching approaches and various conceptualisations of NOS. The examination of NOS literature clearly indicates that there exists little classroom-based research on teachers' practices in teaching NOS compared to research conducted within university-based professional development contexts, and also that in many cases, NOS is not the main learning goal of lessons. Next, I review the current status of research on classroom assessment with a focus on its two aims (i.e., formative and summative), the conditions that determine the quality of classroom assessment, and the roles of teachers as a central actor that performs classroom assessment. The review of classroom assessment research points to the respective importance of formative assessment and summative assessment, each in facilitating student learning and accounting for the learning. At the same time, these two purposes of assessment are interrelated and mutually dependent, according to recent views. After synthesising these two bodies of literature to identify the issues of assessing NOS in particular, potential challenges, and what remains unanswered, I discuss the theoretical commitments that guide this thesis.

In the rest of the chapter, I outline the research questions that guided this thesis, along with a brief overview of how they relate to data collection and analysis methods. This thesis poses three key questions about assessing NOS in classrooms, based on my fieldwork in the school over two academic years. First, *how do teachers view and practise the summative assessment of NOS in the classroom?* (Chapter 5); Second, *how does a team of three science teachers engage in NOS summative assessment in a high school? And what does their decision-making process reveal about the potential issues in the summative assessment of NOS?* (Chapter 6); Third, *what are the instructional strategies that allow for teachers' formative assessment NOS? And what are some potential issues and challenges involved in the formative assessment of NOS?* (Chapter

7). Although both questions are concerned with classroom assessment, they illuminate different (although interconnected) aspects of it. The summative assessment question enables understanding how system-level considerations such as accountability pressure, curriculum policy, national and school culture are interpreted and implemented by teachers to shape their assessment practices about NOS. On the other hand, the formative assessment question focuses on the link between classroom instruction and assessment, particularly the contributions of this ongoing, continuous formative assessment by teachers to students' enhanced learning of NOS. In the meantime, how the interrelation between formative and summative assessment unfolded in their assessment practice is also considered throughout the thesis. This way, the two research questions allow for a holistic look into classroom assessment of NOS, ranging from the teacher and student talk in the classroom to the broader assessment policy and culture in which NOS is assessed.

Chapter 3 describes the multiple layers of context in which the study is situated. Education is a social and cultural phenomenon so cannot be separated from the unique time and place it is happening. I first provide an overview of the education culture in Korea, particularly the societal significance of tests, assessments and university entrance that has several historical origins. In addition, I also outline how these factors shape science education and teachers' classroom practice in Korea. I then describe the process in which NOS was introduced in the most recent curriculum reform in 2015, and how this change relates to the broader contexts of curriculum and assessment. Besides these upper-level contexts, I describe the context of the particular school in which my fieldwork was conducted, since it had several important socioeconomic features that need to be considered for the interpretation of findings. These contexts will often be revisited in the following chapters to consider the extent to which the current study can be generalised for NOS and classroom assessment research.

Chapter 4 provides the methodology and methods for the study. To explore how teachers assess NOS, this thesis uses an in-depth qualitative research method. I justify my choice of a qualitative single case study approach by comparing the features of the case study methodology to the aims and desired contributions of my study. Between 2019 and 2020, various types of data including teacher interviews, lesson observations, lesson and assessment materials, and other classroom artefacts were collected to describe and understand the teachers' teaching and assessment practice. Data collection methods and the process of methodological decision-making are detailed, including how the design was adapted in response to the unexpected school closures due to the Coronavirus disease 2019 (COVID-19) pandemic. Each step of the coding process and the various measures of ensuring the rigour of the analysis are explained, and the methodological and ethical issues I encountered and how I addressed them are also described. In this chapter, I also describe my own researcher positionality in relation to the context, process, interpretation of the study and my relationship with the participants, and discuss how the study can further our knowledge of NOS instruction beyond the specific context within which it was situated.

The next three chapters report the research findings. In Chapter 5, I investigate the three science teachers' views of and practices in assessing NOS with a focus on summative assessment. All school subjects need some form of summative assessment, through which the teacher evaluates and reports on student learning at certain time points. The way a teacher enacts a summative assessment can not only show their own pedagogical priorities but also reveal how their practice is shaped within the broader contexts of curriculum, policy and culture. I first present how the teachers viewed the teaching and summative assessment of NOS in the context of the new curriculum. These views are then compared with their enactments of three NOS assessment activities, which provides insights about the possible sources of disagreement and also the differences across the three teachers. Not only does the analysis provide a close illustration of teacher practices, but it also suggests the need for supporting science teachers to develop a coherent views and practices in NOS instruction and assessment.

While Chapter 5 focused on individual teachers, Chapter 6 attends to the collective aspect of the summative assessment to explore more specific issues in NOS assessment. As a team, the teachers used diverse activities to promote students' learning of NOS, contextualised within history of science and student-led investigations, but when the time came for marking and grading, they confronted several challenges. The teachers had enacted the planned activities in different manners in their lessons, with different levels of guidance and feedback to students, which raised the issue of comparability and fairness in terms of grading. In addition, the teachers found it challenging to discriminate among student answers due to the contentious and open-ended nature of

NOS knowledge. Besides, there were concerns relating to the content knowledge and reading literacy required for carrying out NOS tasks, which could potentially lead to a validity issue. Such complexities, combined with the relatively low accountability pressure for the subject, played an important role in shaping their grading practice of NOS assessment tasks. By looking into teachers' collective decision-making process, this chapter offers some possible challenges and issues that can arise in NOS summative assessment, and how the teachers dealt with them at individual and collective levels.

Recent studies on classroom assessment suggest that assessment does not only happen in the form of tests or only at the end of instruction. Rather, assessment can be a formative process as the teacher seeks and uses information about students' learning and provides appropriate feedback to students. In Chapter 7, I turn to the issue of formative assessment and explore how NOS learning can be facilitated particularly through day-to-day, informal conversations in the classroom. I focus on the instructional practices of one teacher as she taught a lesson on the paradigm shift and the development of science, where students presented examples of the paradigm shift in science and exchanged feedback with peers and the teacher. Analysis of the classroom conversation leads to identifying a number of useful practices for formative assessment employed by the teacher, but at the same time, it also reveals several moments when particular challenges arose for the teacher with giving prompt and appropriate feedback. From the analysis, I assert that in a NOS classroom, it is crucial for teachers to keep the classroom conversation focused on the target NOS and simultaneously create opportunities to broaden the NOS learning by incorporating content knowledge, historical episodes and other aspects of NOS that are relevant to the target NOS.

Chapter 8 ends with a discussion of the various themes that emerged in the thesis. First, I situate the findings within the research literature on NOS and classroom assessment. From a NOS perspective, the affordances and challenges of assessing NOS in classrooms identified in this thesis contribute to the current literature which is mostly focused on the use of researcher-developed NOS instruments. From an assessment perspective, my findings illustrate the experiences and struggles of science teachers in the face of a new curriculum where a non-traditional form of knowledge (i.e., NOS) is taught and assessed. After decades of research efforts, teaching and assessing NOS remains to be unfamiliar and challenging to science teachers. From a global perspective, the foregrounding of NOS in the new Korean curriculum is an ambitious experiment. By synthesising the findings in light of the existing literature on NOS and classroom assessment, it will become clear that this thesis will inform future research in this field and also contribute to preparing teachers for meaningful and effective teaching of NOS in secondary science.

With this thesis, I hope to initiate further discussion about the classroom assessment of NOS through an exploratory case study of three teachers. By illustrating how the traditional as well as new issues in NOS education unfolded in the context of classroom assessment, it provides a fresh perspective from which to consider NOS and propose new questions relating to the why, what and how of teaching NOS in schools. For the successful implementation of NOS in schools, the question of assessment is inevitable, given the educational reality in many countries where teaching and learning are significantly influenced by assessment (Allchin, 2013; Hanuscin, Khajeloo & Herman, 2020). Hence, an examination of teacher practices in NOS classroom assessment can provide valuable information necessary for transforming NOS education in the culture of accountability (Allchin, 2013).

Just as NOS researchers have struggled to strike a balance between preserving philosophical teachings and translating them into effective pedagogy, writing a thesis on NOS certainly required a strategic presentation of educational arguments and their often-complicated ingredients. NOS is an interdisciplinary area with its disciplinary roots in natural science, history of science, philosophy of science, and sociology of science. Accordingly, NOS ideas such as how science works in relation to society cannot be properly addressed without reference to historical cases as well as philosophical and sociological reflections (Allchin, Andersen & Nielsen, 2014; McComas, Clough & Nouri, 2020). For this reason, relevant backgrounds needed to be referred to in the thesis while interpreting teachers' practices, to provide contexts to the NOS in question. Sometimes a short description suffices (e.g., explaining what the Mpemba effect is), but at other times, these issues can be complicated and debated among historians and philosophers (e.g., how to define a paradigm). Given the nature of this thesis as an educational research project, exhaustive discussions of these historical and philosophical details are beyond the scope. These details, however, are often essential or at least helpful to explain the sources of difficulties for teachers and also to draw pedagogical implications. Therefore, I address some of these background ideas in the text where necessary or more frequently as footnotes for the convenience of readers.

Chapter 2 Literature review

This study builds on existing research on NOS and classroom assessment for developing the research questions, designing the study and interpreting the findings. Both NOS and classroom assessment are significant research areas that are discussed extensively in influential science education research handbooks (e.g., Abell & Lederman, 2007; Fraser, Tobin & McRobbie, 2012). Since the former mainly concerns the educational content to be taught in the science classroom and the latter is about a specific subset of teacher activities in the classroom, any discussion on the classroom assessment of NOS inevitably requires a synthesis of the two bodies of literature. The purpose of this chapter is to provide the existing research bases in these areas, consider the gaps in research and develop my orientation towards some key issues relating to NOS and classroom assessment.

I start with reviewing the research on NOS, particularly with a focus on the arguments for including NOS in secondary science education, competing frameworks for characterising NOS that have been proposed thus far, methods suggested for effective teaching of NOS, and teacher knowledge in NOS (Section 2.1). Given the close connection between classroom assessment and instruction, a review of research on NOS instruction will provide initial insights useful for considering the classroom assessment of NOS. Section 2.2 focuses on the main concepts and issues in classroom assessment research. Of particular interest are formative and summative assessments used by teachers. After considering some criteria for high-quality classroom assessment, I turn to the issue of what constitutes teachers' knowledge and skills required for designing, implementing and using classroom assessment. Then I discuss the specific characteristics of NOS as it relates to classroom assessment, which helps to consider the potential similarities and differences of assessing science and assessing NOS in the classroom—one of the key interests of this thesis. In Section 2.3, the focus shifts to the classroom assessment of NOS. I start by problematising the lack of attention to classroom assessment in the current NOS literature, contrary to the sustained effort to develop and use NOS assessment instruments for research purposes. I then synthesis the existing research to identify some research gaps and several key issues that need to be addressed about the classroom assessment of NOS (Section 2.4) and discuss the theoretical orientation that frames and guides the empirical study (Section 2.5).

2.1 NOS in the science curriculum

2.1.1 Justifying and conceptualising NOS in science education

NOS has emerged as a major agenda for research and education policy particularly over the past three decades (Hodson, 2014). Teaching NOS in school science has been advocated for its diverse potential benefits in achieving scientific literacy, such as enhancing students' understanding of scientific objects and processes, informed decision-making, responsible citizenship (Driver et al., 1997; Hodson, 2014; Lederman, 2007). Along with the increasing awareness of the significance of NOS among researchers, recent curriculum reform documents in many countries have introduced NOS as a major component of scientific literacy (AAAS, 1989; Millar & Osborne, 1998; MOE, 2015; NGSS Lead States, 2013). Erduran and Dagher (2014a) explained that the development of NOS-related curricular frameworks has drawn on two broader themes, namely a philosophical orientation toward science and the constructivist tradition. They explained that these two themes have shaped the policy discourse on what should be taught about NOS and how (Erduran & Dagher, 2014a).

The inclusion of NOS as an aim of science education has been advocated for at least four reasons. The first argument for NOS is that understanding NOS facilitates students' learning of content knowledge (Driver et al., 1996; Solomon, 1991). Many studies have supported the benefits of embedding the aspects of NOS into teaching science content. Galili (2012), for example, advocated using historical episodes in the field of optics for more comprehensive and effective learning of concepts. In the context of chemistry, Erduran and Mugaloglu (2014) suggested that deeper philosophical reflections on the nature of chemical knowledge would help students with difficult concepts such as the Lewis model of acid and base and the interpretation of the Bohr model of the atom. NOS is also indispensable when teaching subjects such as geology, which is an intrinsically historical and interpretive science (as compared to experimental science; Dolphin & Dodick, 2014). Dolphin and Dodick (2014) argued that making sense of the distinct nature of earth sciences is central to understanding geoscience concepts that deal with massive scales in time and space. This view is also supported empirically by studies that showed the positive correlation between conceptual understanding and metacognitive, epistemological understanding in science (Shapiro, 1994; Songer & Linn, 1991).

The second rationale for the inclusion of NOS arises from the practical utility of learning NOS. From a utilitarian standpoint, this view argues that NOS should be taught because it is useful for the personal and social lives of people, particularly by assisting their decision making (Driver et al., 1996; Giere, 1991). Driver et al. (1996) suggested, for example, knowledge of NOS is essential in deciding whether to trust experts (e.g., doctors) and understanding the reliability and possible limitations of their knowledge. Allchin (2011) echoes this point by emphasising how reliability is established in science as a core interest of NOS. Further, NOS promotes students' reasoning skills about socioscientific issues (SSIs) such as climate change (Herman, Owens, Oertli, Zangori & Newton, 2019) and genetically modified foods (Khishfe, 2012b). Besides these benefits to students, it has been suggested that NOS can be beneficial to science itself. Historian of science Stanley (2016) argued that NOS could even benefit practising scientists, by 'exposing scientists to new ways of thinking and forces them to re-examine what is already known' (p. 44). Philosophers of science such as Laplane et al. (2019) have also put forward a similar argument on the necessity of NOS for the advancement of science by enabling scientists to clarify scientific concepts, critically assess scientific assumptions or methods, formulate new concepts and theories, and foster dialogue between different sciences, as well as between science and society.

Third, NOS researchers argue that NOS can make science learning more interesting and nurture students' long-term interest in science. Infusing NOS into science lessons, particularly through the use of history of science, has proven to positively influence students' interest and motivation in science (Heering, 2000; Solbes & Traver, 2003). When history of science is used, science can be told as a story that involves real people and real events. Such 'humanisation' of science (Matthews, 2014) has been suggested to make science learning more attractive to students, especially to girls (Heering, 2000). This idea is also succeeded by more recent attempts to use storytelling, dramas and narratives in science education (Hadzigeorgiou, Klassen & Klassen, 2012; Kim, Yi & Cho, 2014; McGregor, Baskerville, Anderson & Duggan, 2019). Given that historical narratives refer to 'chain[s] of events in cause-effect relationship occurring in time and space' (Bordwell & Thompson, 1990) and can be a powerful tool for pupils to understand key figures, events and their interrelationships in the history of science (Barton & Levstik, 2004), the use of history as narratives can be a powerful tool to promote student engagement through NOS. Highlighting the long-term effect of NOS on individuals' engagement with science and technology, Ryder (2009) argued that understandings about NOS are significant for the way people engage with technology and highlighted the need for developing NOS knowledge from early ages.

Fourth, some authors find the significance of NOS in nurturing humanistic, cultural and civic values. A prominent author in this line of argument is Matthews (2014). He traces the contemporary NOS discussion back to the nineteenth-century philosopher William Whewell, who believed that teaching history of science is indispensable for understanding and promote 'intellectual culture' (p. 390). He places NOS in the tradition of liberal education that aims to 'assist the development of a person's knowledge, moral outlook and behaviour, and aesthetic sensibilities and capacities' (p. 5). Such a rationale for NOS resonates with the renewed interest in the values of humanities in cultivating citizens in a multicultural and democratic society (Nussbaum, 1997). Along the same lines, Yacoubian and Hansson (2020) recently argued that NOS can promote social justice by (a) problematising notions of science and conceptualisations of NOS, (b) challenging who science and science education are for, and (c) empowering students in relation to ongoing societal issues (p. 9). Such claims are gaining increasing attention with the rise of issues such as climate change denial and the anti-vaccination movement (Allchin, 2020), which are inherently connected to a grasp of what science is and how it works.

In addition to these traditional arguments for teaching NOS, since the 2010s, there has emerged a renewed interest in NOS in relation to the changing landscape in the production, dissemination and consumption of scientific information. Notable examples are the denial of anthropogenic climate change, and anti-vaccination conspiracy theories and false rumours that have proliferated among the public during this decade. In this 'post-truth' age where objective facts are less influential in shaping public opinions than emotional appeals (Oxford English Dictionary, n.d.), and science-related claims and information are spread fast through social media and online streaming platforms, it is essential for individuals to be able to assess the credibility and quality of scientific information that they are exposed to (Koltay, 2011). Höttecke and Allchin (2020) argued that under the increasing influence of social media, students also need to understand the nature of science communication, namely how scientific knowledge is delivered to citizens through a number of pathways—from experts directly to citizens, through gatekeepers such as journalistic media, via peer-to-peer dissemination through social networks, and so on. Their point is that assessing the reliability of scientific information requires not only how scientific knowledge is generated but also how it reaches individuals. Furthermore, arguments have been put forward to teach the nature of 'post-normal' science where 'facts [are] uncertain, values in dispute, stakes high and decisions urgent' (Funtowicz & Ravets, 1993, p. 744). Understanding the nature of the uncertainty inherent in contemporary science is essential for decision making in one's personal and political life (Allchin, 2020). For example, NOS understanding can play an important role in making policy decisions on whether to implement social distancing and enforce face coverings and other preventive measures in the wake of a pandemic when conclusive evidence is not available yet.

Despite the shared belief around the significance of NOS, however, researchers have debated, and continue to debate, how to characterise NOS and thus what should be taught about NOS. Traditionally, NOS has been approached from the 'consensus' view (or 'general aspects' view), which is based on a set of tenets that comprise the content of NOS. Lederman et al. (Lederman et al., 2002) suggested seven aspects of NOS that are relevant to K-12 science education: (a) the empirical NOS, (b) inference and theoretical entities in science, (c) nature of scientific theories, (d) distinctions and relationship between scientific theories and laws, (e) the creative and imaginative NOS, (f) the theory-laden NOS, (g) social and cultural influences on scientific knowledge (pp. 499–502). Several adapted versions of this list have been proposed, and then were used to develop instruments for measuring learners' understanding of NOS ideas (Abd-El-Khalick, Bell, & Lederman, 1998; Schwartz & Lederman, 2008). A substantial body of empirical research on students' understandings, teachers' understandings, and NOS representations in curriculum documents and science textbooks has been carried out based on this view of NOS (see Hodson, 2014; Lederman, 2007 for reviews of empirical research).

While this particular view of NOS has driven a considerable body of research to date, recent perspectives on NOS have been questioning the idea that NOS can be explicated in a set of declarative statements, and also that it should be taught in such a form. Schizas, Psillos and Stamou's (2016) article titled *Nature of science or nature of the sciences?* problematised the prevalent approach to NOS which seemingly discounts the specificities of a particular science or scientific field, and called for a new framework for NOS that is informed of heterogeneities and pluralities in scientific practices (Hodson, 2014; Park & Song, 2019). Such opposition to the consensus view has been echoed by several prominent NOS researchers (Allchin, 2017; Duschl & Grandy, 2013; Galili, 2019; Matthews, 2012). One promising alternative to the consensus view was proposed by Irzik and Nola (2011, 2014), which used Wittgenstein's concept of family resemblance as a way of conceptualising science. In their framework, science is broadly understood as:

> a cognitive and social system whose investigative activities have a number of aims that it tries to achieve with the help of its methodologies, methodological rules, the system of knowledge certification and dissemination in line with its institutional social-ethical norms, and when successful, ultimately produces knowledge and serves society. (Irzik & Nola 2014, p. 1014)

This Family Resemblance Approach (FRA) to NOS has since been expanded by Erduran and Dagher (2014a) for application for educational purposes using visual representations. While Irzik and Nola's FRA framework was focused on providing a philosophical account of science, Erduran and Dagher's work transformed it into an extended version of FRA that is based on empirical evidence on the pedagogy of NOS and can be readily applied in educational contexts. Their version of FRA includes the original 'cognitive-epistemic' NOS categories (aims and values, methods, scientific practices, scientific knowledge) proposed by Irzik and Nola, but expanded and detailed the 'social-institutional' categories (professional activities, scientific ethos, social certification and dissemination, social values, political power structure, social organisation and interactions, financial systems), drawing primarily on critical studies of science (Erduran & Dagher, 2014a). Since its proposal, FRA has been utilised for analysing how NOS is represented in the science curriculum policy documents (Erduran & Dagher, 2014b; Kaya & Erduran, 2016; Park, Wu, & Erduran, 2020) and textbooks (BouJaoude, Dagher, & Refai, 2017; McDonald, 2017; Park, Yang, & Song, 2020), and as a basis for developing and evaluating teacher education sessions (Erduran & Kaya, 2019; Kaya, Erduran, Akgün, & Aksöz, 2017; Kaya et al., 2018). Table 2.1 compares the aspects or categories of NOS that are highlighted by the consensus view and alternative approaches including FRA. The most significant point of divergence is whether the framework aims to provide a list of ideas to be learnt and assessed (e.g., the consensus view), or a categorisation of NOS aspects that could be elaborated and discussed in specific contexts (e.g., FRA) (Kampourakis, 2016; Romero-Maltrana & Duarte, 2020). Regarding how to conceptualise the content of NOS for school science, this thesis's theoretical orientation is aligned with that of FRA, as I will further elaborate in Section 2.5.

Many NOS researchers agree that 'the various aspects of NOS [should be] made "visible" within instruction through reflective discussions with students about the practice of science' (Lederman, 2006, p. 312), instead of simply engaging students in scientific inquiry activities (Abd-El-Khalick, 2005; Akerson & Donnelly, 2010; Quigley, Pongsanon, & Akerson, 2011). Various instructional strategies including visual representations (Erduran & Kaya, 2018), classroom discussions (Akerson & Donnelly, 2010; Aragón, José, & Acevedo, 2018) and reflective journals (Abd-El-Khalick, 2005) have been proposed to facilitate NOS learning in the classroom. Researchers have also stressed the importance of contextualising NOS by using historical episodes (Allchin et al., 2014; Chang, 2011; Matthews, 1994), contemporary cases (Khishfe, 2012a, 2012b; Wong, Hodson, Kwan, & Yung, 2008, 2009) and student-centred classroom inquiry activities (Clough, 2006). Allchin et al. (2014) suggested that each method of contextualisation has strengths and shortcomings so they should complement each other in the classroom implementation of NOS.

2.1.2 Instructional strategies for NOS: the explicit-reflective approach

How can NOS be taught effectively and meaningfully? In other words, how should teachers organise students' NOS learning experiences that can actually lead to an improvement in their NOS understanding? Broadly speaking, two opposing approaches to teaching NOS have been suggested: *implicit* and *explicit*. Those who advocate the implicit approach argue that students will learn NOS naturally by engaging in inquiry activities mirroring scientific practice (Lawson, 1982; Rowe, 1974). Although this seems like a reasonable inference in view of common sense, three decades of empirical research has repeatedly confirmed that simply 'doing science' does not automatically result in the learning of NOS (Abd-El-Khalick et al., 1998; Abd-El- Khalick & Akerson, 2009; Akerson, Abd-El-Khalick, & Lederman, 2000). This means that

Table 2.1 The aspects or categories of NOS in several major conceptualisations of NOS

Approach	Consensus view			
Source	Lederman et al. (2002)	McComas (2008, 2015)	Abd-El-Khalick (2012)	Lederman et al. (2014)
Motivation	• At a certain point in time and at a certain level of generality, there exists a shared consensus about NOS among philosophers, historians and sociologists. • Many disagreements about NOS are irrelevant to school science education.			
NOS aspects/categories	Empirical basis Scientific theories and laws Creativity Theory-ladenness Cultural embeddedness The myth of the scientific method Tentativeness	Tools and products of science • Empirical evidence is required • Science shares methods • Law/theory distinction Science knowledge and its limits • Science is distinct from technology and engineering • Science is tentative but durable • Science cannot address all questions Human elements of science • Creativity is vital in science • Subjectivity is a frequent element in science • Social and cultural elements impact science	Empirical Inferential Creative Theory-laden Tentative The myth of 'The Scientific Method' Scientific theories Scientific laws Social dimensions of science Social and cultural embeddedness of science	Scientific investigations all begin with a question, but do not necessarily test a hypothesis. There is no single set or sequence of steps followed in all investigations. Inquiry procedures are guided by the question asked. Not all scientists performing the same procedures may get the same results. Inquiry procedures can influence results. Research conclusions must be consistent with the data collected. Explanations are developed from a combination of collected data and what is already known. Scientific data are not the same as scientific evidence.
Related research instruments	VNOS-A (Lederman & O'Malley, 1990), VNOS-B (Abd-El- Khalick et al., 1998), VNOS-C (Abd-El- Khalick & Lederman, 2000), VNOS-D (Lederman & Khishfe, 2002), VNOS-E (Lederman & Ko, 2004), VASI (Lederman et al., 2014), Views of Science-Technology-Society (VOSTS) (Aikenhead et al., 1987), Student Understanding of Science and Scientific Inquiry (SUSSI) (Liang et al., 2006)			

Table 2.1. (continued)

Approach	Whole science	Features of science	Family resemblance approach	
Source	Allchin (2013)	Matthews (2012)	Irzik & Nola (2014)	Erduran & Dagher (2014a)
Motivation	• NOS consists in 'the multiple dimensions shaping reliability in scientific practice' (p. 518).	• The NOS aspects in the consensus list do not necessarily distinguish science from other fields. • Rather, they should be better thought of 'features of science' to be elaborated, discussed and inquired about.	• In contrast to the consensus view that addresses *individual ideas* about science, FRA consists of *a class of ideas* with regard to which different branches of science could be understood comparatively comprehensively (Erduran & Dagher, 2014a). • For any chosen pair of two scientific disciplines, 'one will be similar to the other with respect to some of these characteristics and dissimilar to one another with respect to other characteristics' (Irzik & Nola, 2014, p. 1013).	
NOS aspects/categories	Observational • Observations and measurements • Experiments • Instruments Conceptual • Patterns of reasoning • Historical dimensions • Human dimensions Sociocultural • Institutions • Biases • Economics/funding Communication	Empirical basis Scientific theories and laws Theory-ladenness Cultural embeddedness The myth of the scientific method Tentativeness Experimentation Idealisation Models Mathematisation Technology Explanation Worldviews and religion Theory choice and rationality Feminism Realism and constructivism	Cognitive-epistemic aspects • Process of inquiry • Aims and values • Methods and methodological rules • Scientific knowledge Social-institutional aspects • Professional activities • The scientific ethos • The social certification and dissemination of scientific knowledge • Social values of science	Cognitive-epistemic aspects • Aims and values • Methods • Scientific practices • Scientific knowledge Social-institutional aspects • Professional activities • Scientific ethos • Social certification and dissemination • Social values • Social organisations and interactions • Political power structures • Financial systems
Related research instruments	N/A[1]	N/A	RFN questionnaire (Kaya et al., 2019)	

[1] Example assessment questions were presented in Allchin (2011) but have not developed into a full instrument.

understanding the NOS aspects that underpin scientific inquiry should be separated from the acquisition of scientific inquiry skills themselves (Lederman, 2002), although it does not mean the two are independent. The explicit approach to NOS instruction is mindful of the distinction between inquiry skills and understanding of NOS and stresses the need to consider NOS as a cognitive learning goal and deliberately plan instruction to address NOS aspects (Lederman, 2002).

Such an approach is not without criticism. Duschl and Grandy (2013), for example, criticised that these instructional approaches are theoretically grounded upon the ideas of several philosophers of science, such as Thomas Kuhn, Karl Popper and Imre Lakatos, who generally defended science as a rational and objective activity. These philosophers drew primarily on the history of science to characterise what science is and so does the approaches to NOS instruction based on such a view. On the contrary, Duschl and Grandy argue that science needs to be understood and conceptualised through its actual practice by scientists. They argue that prioritising practice in NOS would be coherent with the ‘practice turn’ in the philosophy of science (Giere, Bickle & Mauldin, 2006; Latour & Woolgar, 1986; Pickering, 1992) as well as the foregrounding of scientific practices in recent science curricula (Duschl & Bismack, 2016; Erduran, 2015; NGSS Lead States, 2013; NRC, 2012). They highlight the discursive practices of science such as argumentation, representation and communication, emphasise the role of models, and argue that NOS learning should be ‘situated in longer instructional sequences that engage learners with scientific practices’ (p. 2125).

In recent years, theories about NOS instruction have been evolving in a way that seeks synergies between these opposing approaches. Abd-El-Khalick (2012) noted the danger of both extremes of expecting NOS learning to happen naturally through engagement in ‘doing science’ and attempting to address aspects of NOS without meaningfully engaging learners in authentic scientific inquiry. In fact, many recently suggested learning activities for NOS seem to consider both the inquiry element and the explicit NOS teaching element. For example, Lederman, Abd-El-Khalick and Lederman’s (2020) activity aims to teach the method and process of using fossils to infer facts about the past, along with the aspects of NOS such as the differences between observation and inference. In this activity, a significant amount of time is devoted to student activities to discuss the sets of bird footprints on the ground, make inferences about what happened based on the observation, which resembles the works of fossil scientists. Then the teacher can use a range of student responses to point to the desired NOS learning goals through questioning and whole-class discussion. Another example is found in Erduran, Kaya, Cullinane, Imren & Kaya's (2020) activity on protecting the endangered Bengal Tigers in North India. They first present competing claims and reasonings about helping the tigers and humans, and ask students to develop their own arguments about the issue. It is then highlighted that scientists try to keep a neutral and unbiased attitude while they seek solutions to science-related problems. These two activities exemplify the recent trend in NOS teaching where both engagements in scientific practices and explicit NOS instruction are coherently integrated in organising students' learning experience.

The explicit-reflective approach also entails the need for providing students with opportunities to *reflect* on the NOS aspects that they learnt, by constructing connections between the NOS aspect and the context in which it was illustrated (Abd-El-Khalick, 1998; Khishfe & Abd-El-Khalick, 2002). Several instructional strategies have been suggested to be useful to elicit such reflection on NOS. The most frequently used strategies are to provide pre-designed questions as students engage in inquiry activities (Williams & Rudge, 2016; Yacoubian & BouJaoude, 2010) and have whole-class discussions after the activity (Akerson & Donnelly, 2010; Qungley, Pongsanon, & Akerson, 2011). Several studies have also focused on using other instructional strategies such as writing reflection papers (Abd-El-Khalick, 2005) and text-based tasks (Park, Yang & Song, 2020) to elicit students' reflections on NOS.

2.1.3 Contextualising NOS: Historical cases, contemporary cases, student-led inquiry

Clough (2006) stressed the crucial role of contextualised NOS instruction that 'overtly draws students' attention to important NOS issues *entangled in science content and its development*' (p. 474, italics in original). In the literature, at least three major approaches to contextualising NOS can be found (e.g., Allchin et al., 2013; Clough, 2016; McComas et al., 2020): historical cases, contemporary cases and student-led inquiry activities. Allchin, Andersen and Nielsen (2014) suggested that the three approaches to NOS contextualisation can be used complementarily in classroom practice. In addition, as seen in Table 2.2, they highlighted that different modes of contextualisation have strengths and weaknesses of their own, and acknowledged that some aspects of NOS might be taught more efficiently in a certain context than in other contexts. For example, historical cases are well suited for illustrating long-scale NOS features such as theory change, whereas hands-on inquiry activities can be more effective for the features of individual practices in scientific investigations (Allchin et al., 2013). Below, I briefly discuss each of the three approaches with illustrative research examples.

2.1.3.1 Historical case studies

History of science has been used for various purposes in school science education for many decades. In 1945, the report of the Harvard Committee on General Education, led by then Harvard president James Bryant Conant, articulated a position on the education of students at the elementary and secondary school level. Specifically, the report highlighted a historical understanding of science:

> Science instruction in general education should be characterised mainly by broad integrative elements—the comparison of scientific with other modes of thought, the comparison and contrast of the individual sciences with one another, the relations of science with its own past and with general human history, and of science with problems of human society. (Conant, 1945, as cited in NSSE, 1947, p. 20)

As acknowledged in this example, history can play an important role in science education and particularly reveal many key aspects of NOS. For example, referring to how the atomic model evolved from antiquity to the recent molecular orbital model can point to the fact that scientific knowledge changes over time based on the efforts of many scientists. Investigating historical episodes in a relatively shorter timespan can also be useful. Del Mar Aragón-Méndez, Acevedo-Díaz and García-Carmona (2018) used the story of Semmelweis who discovered the relation between childbed fever and hand washing, a discovery which contributed to reducing the deaths by the disease greatly. In a preservice biology teacher training context, they used this case to illuminate the aspects of NOS such as the difference between observation and inference in science, the role of creativity in scientific activities, and the significance of communication on the acceptance of scientific findings. Similarly, Erduran and Dagher (2014a) illustrated that the story of discovering the double helix structure of DNA can point to various epistemic (e.g., non-manipulative observation in science) and social (e.g., the institutional sexism experienced by the female scientist Rosalind Franklin) aspects of NOS.

2.1.3.2 Contemporary cases

The use of contemporary cases to teach NOS has been frequently linked to SSIs, which is reasonable given their common aims such as promoting citizenship, cultivating critical thinking skills and epistemological understanding of science (Erduran & Dagher, 2014a; Zeidler, Applebaum & Sadler, 2011). These authors note that a proper knowledge of NOS is necessary to make decisions on SSIs and also can be facilitated by using SSIs as contexts (Yacoubian, 2015). Eastwood et al. (2012) used a range of SSIs including stem cell research, fluoridation of public water supplies, and fast food and health, and found that this approach helped 11th and 12th-graders use examples to describe their NOS views better than when NOS was taught in a content-driven context. Similarly, Khishfe (2012a) engaged ninth-grade students in a four-week unit about genetic engineering including a controversial SSI about genetically modified foods. In this quasi-experimental study, she taught explicitly about the empirical, tentative and subjective NOS during the units to the treatment group and found this group drew on these NOS ideas to make decisions about the controversial issue. The pedagogical significance of using contemporary cases to teach NOS is rising as we face serious challenges to the authority of science in times of global disasters such as climate change and COVID-19 (Erduran, 2020).

2.1.3.3 Student-led inquiry activities

Many enactments of NOS instruction have included some kinds of hands-on inquiry activities that model the processes and practices of scientists (Deng, Chen, Tsai & Tsai,

Table 2.2 Merits and shortcomings of modes of NOS instruction (Allchin et al., 2014, p. 473)

Mode	Merits	Shortcomings
Historical case	• helps motivate engagement through cultural and human contexts and through a narrative format • can support understanding of long-scale and large-context NOS features: esp. conceptual change, and cultural/biographical/economic contexts of research problems and interpretive biases • can support understanding of investigative NOS: problem-posing, problem-solving, persuasion, debate • can support understanding of the complexity of scientific practice, as well as historical contingency • supports analysis of process and product, since ultimate outcomes are known • when framed in inquiry mode, can develop scientific thinking skills—more efficiently than with hands-on inquiry • can foster understanding of error and revision—without risking emotions of personal failure	• may seem ‘old’ and irrelevant • difficult or time-consuming for teachers to learn background or historical perspective • if text-based only, limits the development of hands-on experimental competences • if rationally reconstructed only or presented as final-form content, does not support understanding of ‘science-in-the-making’
Contemporary case	• helps motivate engagement through authenticity and ‘here-now’ relevance • can support understanding of how science and values relate • develops scientific literacy skills in analysing SSI	• cannot be fully resolved, leaving uncertainty and incomplete NOS lessons • cannot exhibit details of the process which are not yet public or are culturally obscured
Inquiry	• helps motivate engagement through personal involvement • fosters personal integration of lessons • supports understanding of constructed interpretations, models, forms of evidence, and model revision • develops experimental competences: framing hypotheses, designing investigations, handling data, evaluating results • relates nature of scientific knowledge to inquiry skills and methods	• difficult to motivate all students, especially as a group • may be viewed as artificial exercise or school ‘game’, not as genuine science • when investigations ‘fail’, can prompt negative emotions, alienating students from NOS lessons • typically shuttered off from cultural, social or political contexts • hard to role model of ‘chance’, or contingency • requires substantive amounts of time and resources

2011). Integrating NOS with inquiry activities has been claimed to make NOS learning more engaging, authentic and meaningful to students (Yacoubian & BouJaoude, 2010). For example, Akerson, Hanson and Cullen (2007) gave K-6 teachers a guided inquiry activity related to physical sciences, coupled with explicit NOS instruction followed by reflective discussion, which was found to improve the teachers' NOS views. Price and Perez's (2018) study with science major university students used the *Cube Puzzle* activity where students used five sides of the cube to infer what is on the sixth side, and then reflected on how this activity compares with how science works. They also engaged students in another activity on the discovery of the asteroid impact that caused the Cretaceous-Tertiary Extinction to highlight the unpredictability of scientific exploration. These activities, they reported, were effective in developing students' sophisticated knowledge of the processes of science and social aspects of science (Price & Perez, 2018).

2.1.4 Diverse approaches to NOS learning

Along with the ones described so far, NOS researchers have implemented diverse methods and approaches to facilitate NOS learning. Empirical research on NOS instruction ranges from early childhood to postsecondary and adult learners, from classrooms to informal, out-of-school settings, from using texts to providing authentic research experiences to learners. Some studies are situated within specific disciplinary contexts, while other studies are centred around specific issues with multidisciplinary characters such as SSIs. In the subsequent discussion, I outline several noteworthy strands of research on instructional methods for NOS that are distinct from the three approaches to NOS contextualisation mentioned earlier.

First, there has been a long-standing interest in using textual materials to promote NOS learning. Niaz (2012, 2016) illustrated what aspects of NOS could be learnt in diverse chemistry topics such as Millikan's oil drop experiment, atomic theory, stoichiometry, and valence bond and molecular orbital models, and conducted extensive analyses of undergraduate chemistry textbooks. A recent collection of research edited by McDonald and Abd-El-Khalick (2017) represents the continuing interest in the use of textbooks to teach NOS. From an analysis of NOS-related textbook tasks, Park, Yang and Song (2020) suggested a typology of textbook tasks that could be used to facilitate NOS learning in classrooms. Meanwhile, several researchers have attended to alternative forms of texts. Brunner and Abd-El-Khalick (2017, 2020) used modified science trade books for elementary teacher education and found that the intervention positively impacted teachers' and students' NOS understandings. In García-Carmona and Acevedo Díaz's (2016) study, preservice teachers read newspaper articles on an ongoing controversy among physicists on the neutrinos experiment, followed by reflective discussion in small groups. Another recent attempt is using science classics. An example is Kiang and Colanero's (2020) undergraduate modules, where they used famous science classics ranging from Euclid's *Elements* and Newton's *Principia* to Rachel Carson's *Silent Spring*. They used these texts to illuminate various aspects of NOS, such as the rationality of science, the scientific revolution, moral responsibility, value-ladenness and cultural embeddedness of scientific knowledge (Kiang & Colanero, 2020).

Second, engaging learners in authentic research experience has also been suggested to facilitate NOS learning. Being 'authentic' means that the learner's experience should be in alignment with that which 'occurs within the scientific community by practising scientists in efforts to gain an understanding of the natural world' (Schwartz & Crawford, 2006, p. 337). Bell, Blair, Crawford and Lederman's (2003) study with high-ability secondary students who participated in an eight-week science apprenticeship programme reported that students' ideas about NOS remained largely superficial and naïve despite the experience. On the contrary, Schwartz, Lederman and Crawford's (2000) study with preservice teachers showed a positive impact of placement with university scientists. They suggested that research experience can be effective when combined with reflective activities such as journal writing and seminars, as they did in their programme (Burgin & Sadler, 2015; Schwartz et al, 2000).

Third, several authors have proposed using art as a medium to reveal and represent NOS. Galili (2013), for example, used the famous fresco *The School of Athens* (1511) as a means to address the rationalist versus empiricist views about scientific knowledge. He also argued that Picasso's Cubist paintings could be used to draw parallels with the advent of twentieth-century physics that changed our understanding of simultaneity and frames of reference (Galili, 2013). Similarly, Güney and Şeker (2017) considered the socio-cultural aspects of science that artworks represent. They argued that artworks such as the fresco *Galileo demonstrates the law of falling bodies* (1841) by Giuseppe Bezzuoli and the frontispiece *Elements of the Philosophy of Newton* (1738) written by Voltaire can shed light on aspects such as the relationship of science and the church, and the mutual influence of major figures in the history of science (Güney & Şeker, 2017). The key idea underpinning these authors' proposals is the power of artwork where individual objects and people in the paintings can tell us stories about the time that it depicts, and often symbolise important messages about nature and humans. In a somewhat related context, art has also been advocated as a tool to express and represent learners' understanding of NOS visually. Erduran and Kaya (2018) used preservice teachers' drawings about different aspects of NOS to assess the quality of their NOS understandings, showing the value of drawing as a representational tool for teaching and learning NOS.

The potential of art is also related to the fourth line of research that uses stories and narratives as pedagogical tools for teaching NOS. Inspired by the universal value of storytelling in teaching and learning (Egan, 1989), science educators have used stories from history of science to teach aspects of NOS. Hadzigeorgiou, Klassen and Klassen (2012) used the story of Nikola Tesla, who defeated Thomas Edison in the famous 'current war' between direct and alternative current, to develop students' understanding of the social, human, personal dimensions of science. The use of stories and narratives are not limited to historical examples. Kim, Yi and Cho (2014) produced a documentary entitled *Life as a Scientist: People in Love with C. elegans, a Soil Nematode* based on the lives of scientists' lives in a university lab. This documentary was found to have enabled university students to recognise the role of negotiation, cooperation and collaboration in advancing scientific knowledge (Kim, Yi & Cho, 2014). Similarly, a few authors have focused on the potential of drama and role-play (Bentley, 2000; BouJaoude, Sowwan & Abd-El-Khalick, 2005; McGregor, Baskerville, Anderson & Duggan, 2019; Cakici & Bayir, 2012) as a medium to teach NOS. For example, McGregor et al. (2019) engaged nine to ten-year-old students in New Zealand and England in drama activities, where the students worked in-role as atmospheric and material scientists, combined with reflective discussions. They found that such an experience improved students' understanding of the nature of scientists' work as well as scientists' dispositions such as perseverance and collaboration (McGregor et al., 2019).

Fifth, there is an increasing interest in using digital technologies as a vehicle to improve NOS learning by providing interactive and user-centred experiences. This trend is in tandem with the recent emphasis on cultivating teachers' and students' digital literacies that can assist efficient learning and lessen the cognitive load during learning (Ng, 2011). An early such effort is found in Walker and Zeidler's (2003) study, where they taught high school students an online-based unit on genetically modified foods that consisted of seven 1.5-hour sessions, from which they could reveal students' conceptions of NOS. Schellinger et al. (2019) developed the Habitat Tracker curriculum to foster fourth and fifth-grade students' NOS learning. In their module that took place across formal and informal settings, students used a website to conduct background research and data analysis, and an iPad application to guide observation and reflection during field trips to the wildlife centre. Digital technologies have also been used in initial teacher training contexts. Mansour et al. (2016) used online software that supports learners' collaborative knowledge construction through reasoned dialogues, to improve UK preservice teachers' NOS conception. The participants were engaged in an asynchronous online discussion about NOS and its teaching, followed by a face-to-face discussion. DeCoito (2020) used digital scientific timelines and a digital video game as a means to teach NOS to preservice teachers. She found that as the preservice teachers developed a detailed digital timeline of science, they developed better understandings of the sociocultural aspects of science. Also, she found that playing an interactive online game called *History of Biology* over a four-month period improved preservice teachers' conceptions about a number of aspects of NOS (DeCoito, 2020).

Sixth, researchers have also attended to the potential of informal, out-of-school environments such as museums, fields, zoos, aquaria, botanic gardens, magazines and films (Ahrenkiel & Worm-Leonhard, 2014; Rocha Fernandes, Rodrigues & Ferreira, 2018). Reiss and McComas (2020) argued that informal learning has an important role in bridging school lessons with lifelong learning in terms of NOS. Dhingra (2003) investigated how different genres of television programmes mediate US high school students' views of the ethics and validity of science, final form science, science as portrayed by its practitioners, and school science and television science. Quigley, Pongsanon and Akerson (2011) taught NOS to K-2 students in a Saturday science class, which addressed the scientific inventions that help travel, such as wheels, cars, trains, bridges and aeroplanes. They reported that the informal setting allowed the instructors to flexibly adapt lesson plans to meet individual learners' interests and include many interactive, hands-on activities without the pressure of formal assessment (Quigley et al., 2011). Science camps are also a widely used context for engaging learners in NOS learning. For example, Leblebicioglu and colleagues worked with sixth and seventh-grade students in a science camp focused on NOS and scientific inquiry, using nearby plants, soil, animals and water as contexts. After using an explicit approach to NOS during the camp, the students developed a better understanding of the nature of scientific inquiry (Leblebicioglu et al., 2017), and such improvement was retained after two months (Leblebicioglu et al., 2019). These studies show examples of NOS learning transcending the boundary of formal science education and suggest that the synergy of school science and informal science learning can be pursued for meaningful and lasting NOS learning.

There are also a number of other recent initiatives in NOS education that are worth noting. For example, making NOS more accessible to diverse learners in terms of disability (Enderle, Cohen & Scott, 2020), gender and ethnicity (Bianchini, Cavazos & Helms, 2000; Yacoubian & Hanson, 2020) have been considered by several researchers, in line with the general interest in such themes within the science education community. In addition, researchers have begun to consider the nature of related subjects such as technology (Waight & Abd-El-Khalick, 2012), engineering (Antink-Meyer & Brown, 2019; Pleasants & Olson, 2018), and even in the broader interdisciplinary context of STEM (Chesky & Wolfmeyer, 2015; Erduran, 2020; Park, Wu & Erduran, 2020). Behind these 'extensions' of NOS research is a realisation that scientific practice is increasingly interdependent upon and inseparable from closely related STEM disciplines, so the knowledge about the nature of these disciplines will eventually contribute to a better understanding of NOS and enhanced scientific literacy (Park, Wu & Erduran, 2020). Examples such as these show the diversity and richness of scholarship on the effective and meaningful instruction of NOS.

2.1.5 Teachers' knowledge of and challenges in teaching NOS

Given the role of teachers in student learning, preparing teachers for teaching NOS to students is a core task for the NOS-oriented curriculum reform to lead to classroom practice and eventually take effect. From existing research on teacher knowledge for teaching NOS, at least two important points can be drawn. First, teaching NOS requires knowledge in diverse domains throughout scientific content knowledge, history and philosophy of science, instructional strategies and so on (Abd-El-Khalick, 2013; Wahbeh & Abd-El-Khalick, 2014). For this reason, Erduran and Dagher (2014a) stressed that improving teachers' knowledge for successful teaching of NOS should build on the successful models in teacher education research, such as pedagogical content knowledge (PCK). Abd-El-Khalick (2013) proposed three dimensions of teacher knowledge required for teaching NOS (or NOS-PCK): science content understanding, NOS understanding, and inquiry pedagogical understanding & skills. In particular, he emphasised that teacher NOS understanding should include both domain-general and domain-specific (or content-situated) knowledge of NOS (Abd-El-Khalick, 2013).

Second, evidence suggests that teachers' understandings of NOS are necessary for teaching NOS but not sufficient (Abd-El-khalick & Lederman, 2000). Lederman's (1999) study of five high school teachers revealed that teachers' conceptions of science do not necessarily influence their instructional planning and classroom practice, but only do so when supported by teachers' level of experience, intentions, and perceptions of students. NOS researchers have identified other factors that mediate the translation of teachers' NOS understanding into teaching practice, such as teachers' beliefs about the importance of teaching and learning NOS, external pressures regarding high-stakes examinations, classroom management issues, and availability of instructional materials for teaching NOS (Abd-El-Khalick, Bell, & Lederman, 1998b; Brickhouse, 1990). Recent views have drawn on the concept of identity to explain the relationship between teachers' knowledge and practice regarding NOS instruction (Akerson, Carter, & Elcan, 2016; Erduran & Kaya, 2019). Akerson et al. (2016) suggested that developing a professional identity as teachers of NOS is a key condition for successful NOS teaching. From a study that involved preservice and in-service elementary teachers, they found that having an accurate understanding of NOS and actively translating and infusing the concept of NOS into their teaching practice (i.e., developing agency in

NOS teaching) was crucial in building an identity as a teacher of NOS (Akerson et al., 2016). In a similar line, Erduran and Kaya (2019) proposed that developing the ‘epistemic identity’ is an essential task in educating teachers for teaching the epistemic nature of science to students.

When NOS instruction is introduced as part of curriculum reform, it is also important to consider teachers’ views, responses and actions through the lens of curriculum implementation (Roehrig, Kruse & Kern, 2007). Studies on science education reforms have identified several major conditions that facilitate or impede the successful implementation of curriculum reforms by teachers. When teachers implement a science curriculum reform, they are influenced by their knowledge and beliefs about discipline (Roehrig & Garrow, 2007), skills to implement instructional methods for the new curriculum (Adams & Krockover, 1997), teachers’ perceived ability of students (Cronin-Jones, 1991) and their beliefs about teaching and learning, such as the transmission of knowledge, maintaining the rigour of the curriculum and preparing students for examinations (Roehrig et al., 2007; Tobin & McRobbie, 1996).

When it comes specifically to teaching NOS, there are reported constraints that teachers experience, arising from students’ cognitive and affective abilities (Abd-El-Khalick, Bell, & Lederman, 1998), lack of teachers’ NOS-specific skills (Schwartz & Lederman, 2002), insufficient educational resources (Wong, Kwan, Hodson, & Yung, 2009), and teachers’ attitudes and beliefs about the importance of NOS (Lederman, 1999). Particularly in relation to physics, Höttecke and Silva (2011) identified four major obstacles in implementing history and philosophy of science (HPS) in physics classrooms: (1) the subject culture of teaching physics, where physics is taught as truth and a collection of facts, and content is ‘hard facts’ rather than a negotiable matter; (2) physics teachers’ attitudes, beliefs and skills regarding HPS instruction, for example, not conceiving NOS as an explicit teaching objective, using HPS only as a context for the scientific content, and the lack of relevant instructional skills; (3) the organisation of the curriculum, where history is only an overall objective with no aimed competencies directly linked to HPS; and (4) the content of textbooks that only addresses HPS as a minor component. Although these studies offer valuable insights into teachers’ perspectives and experiences of NOS instruction, one apparent limitation of these studies is that they were situated within ‘science’ curricula where NOS and

HPS were either subsidiary curriculum goals or not present at all.

In response to these well-documented challenges, preparing science teachers for accurate and effective teaching of NOS has arisen as a key task and concern among researchers. In a recent review of research on NOS and teacher education, Clough, Herman and Olson (2020) suggested five fundamental principles of effective NOS instruction in relation to teachers: First, teachers must have an accurate understanding of NOS. Second, NOS instruction should be purposefully planned and implemented consistently throughout the course of science instruction. Third, teachers should explicitly direct students' attention to NOS ideas, while requiring students to mentally engage in wrestling with those ideas. Fourth, teachers should be able to use various contexts ranging from decontextualised, moderately contextualised, to highly contextualised and scaffold back and forth between these contexts. Fifth, specific instructional settings such as inquiry learning and SSIs can provide opportunities for addressing NOS. Sixth, students' NOS understanding must be accurately assessed in the form of homework, teacher-developed assessments, and high-stake external exams.

Another recent study by Mesci, Schwartz and Pleasants (2020) drew on the framework of pedagogical content knowledge of teachers to explore the knowledge required for teachers to effectively teach NOS. They suggested five key constructs that are essential to teaching NOS: orientation towards teaching NOS, knowledge of students' understanding of NOS, knowledge of science curriculum of NOS, knowledge of assessment in science, and knowledge of topic-specific representations and instructional strategies (Figure 2.1). Through the examination of two preservice teachers' experiences during a professional development programme including teaching practicum, they also found several additional factors that affected the practice of teaching NOS, such as subject matter knowledge, teacher self-efficacy, lesson planning and general pedagogical knowledge (Mesci et al., 2020). Models such as this highlight that for effective teaching of NOS a number of conditions need to be met, including teachers' knowledge, experience and instructional orientation. Nevertheless, there is still little discussion around teacher knowledge and skills required for assessing NOS in the classroom.

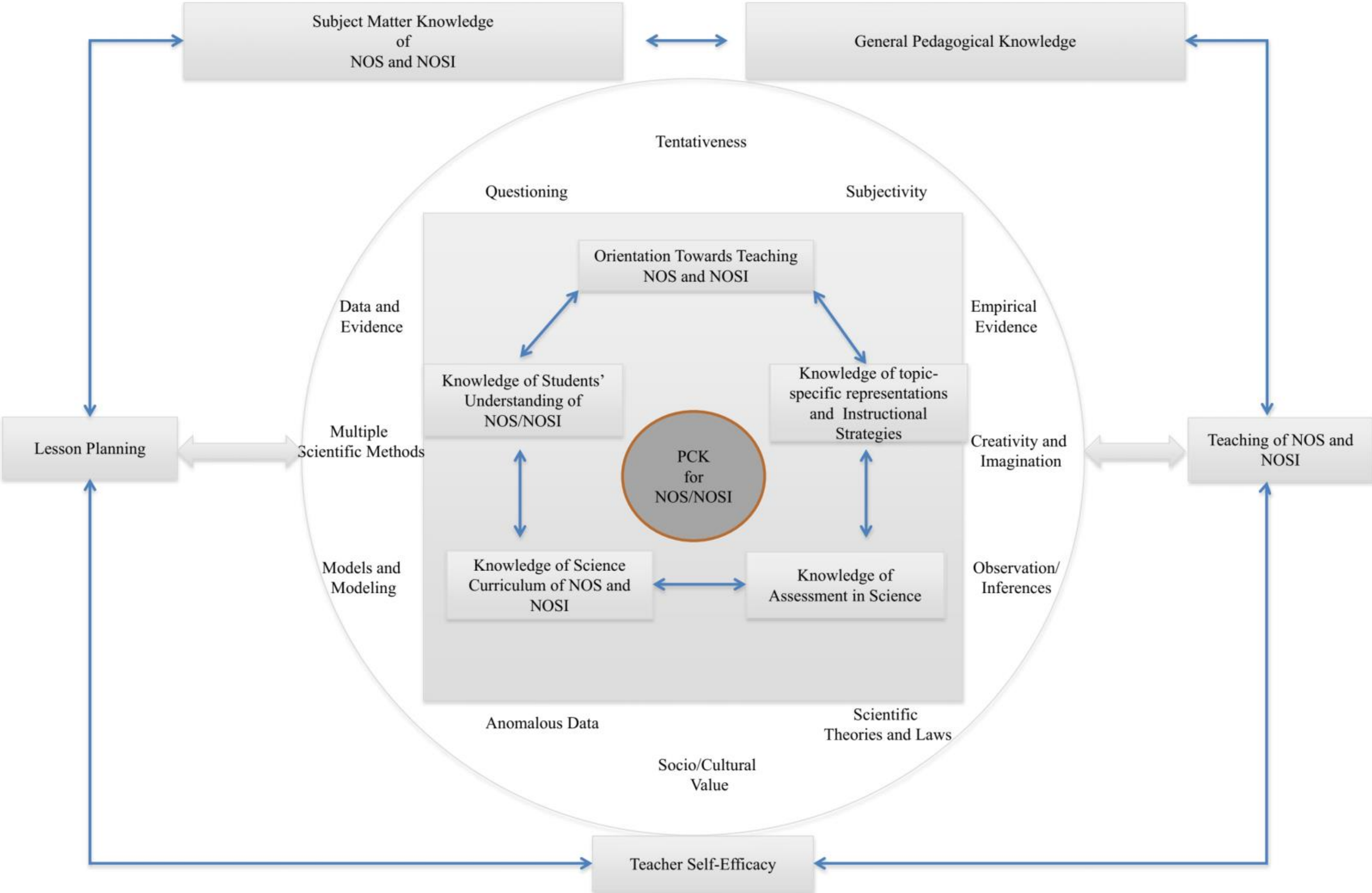


Figure 2.1 Key elements of teacher knowledge for teaching NOS (Mesci et al., 2020, p. 286)

2.1.6 Summary of literature in NOS education

How we conceptualise and theorise NOS carries significant implications for its teaching and assessment. In this thesis, I use the recent FRA model to characterise NOS, which acknowledges the interconnectedness of scientific knowledge and scientific inquiry, the domain-specificity of NOS, and therefore opens the possibility for addressing controversial issues in NOS. As will become clear in the findings, interpretation of classroom interactions, and teaching and assessment strategies regarding NOS can be more effectively made with an open-ended approach rather than based on a predetermined set of NOS aspects. For example, the classroom interactions about paradigm shifts in science that will be presented in Chapter 7 are more effectively captured when various aims, methods and social impacts of science are considered holistically, compared to using the ‘tentativeness’ lens from the consensus view.

The review of research on the conceptualisation and pedagogy of NOS suggests two major concerns in the current landscape of NOS research. The first involves the importance of examining teachers’ instructional practices in classroom settings. Instructional practice is a manifestation of a teacher’s knowledge of NOS, its teaching, pedagogical knowledge and other educational and practical considerations (Cohen, Raudenbush & Ball, 2003; Mesci et al., 2020). A focus on practice can thus shed light on how NOS-related issues unfold in individual instructional situations and can suggest ways to improve NOS teaching and teacher education. Second, studies consistently point to the need to set NOS as an explicit learning aim and suggest assessing NOS as a major way to do so (Clough, Herman & Olson, 2020; Mesci et al., 2020). However, little has been discussed on how the elements of assessment can be considered in the context of NOS to improve student learning. In the following sections, I consider theories on classroom assessment and their application to NOS to frame this thesis study.

2.2 Classroom assessment

2.2.1 Purposes of classroom assessment: formative and summative

In many education systems, assessment is an essential component that significantly drives instructional practice and student learning (McMillan, 2018). In the broadest sense, assessment in education can be divided into external assessment and internal (classroom) assessment. External assessments are designed and administered by countries or international bodies to audit or monitor learning and are based on specific standards for knowledge and skills (McMillan, 2018). These assessments often have major consequences (i.e., high stakes) for students, teachers, schools and policymakers, as in the cases of the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS). Classroom assessment, on the contrary, is primarily concerned with teacher-student interactions, observations and evidence of student learning produced in the course or at the end of instruction (McMillan, 2018). Classroom assessment, therefore, occurs naturally as part of teaching and learning, as teachers and students ask questions, report on assignments and make decisions for the next steps (NRC, 2001).

Educators have stressed the importance of ensuring the quality of classroom assessment on multiple grounds. McMillan (2018) stated that high-quality assessment can have positive consequences for both students and teachers. For example, there is considerable evidence that classroom assessment can increase students' motivation when properly planned and enacted (McMillan & Hearn, 2008; Nikou & Economides, 2016), and help to build a positive teacher-student relationship (Crimmins et al., 2016). Assessment can also promote students' sense of autonomy and responsibility for their own learning (Andrade & Cizek, 2010), especially by means of self-assessment and peer assessment. On the teachers' side, assessment can influence teaching practice by encouraging teachers to focus on what is highlighted in the assessment (McMillan, 2018), and allowing them to better identify each students' strengths and weaknesses and take actions to maximise learning.

An influential taxonomy of classroom assessment was proposed by Black (1998), who suggested that assessment can serve three purposes: supporting learning, reporting the achievements of individuals, and satisfying public demands for accountability (p. 24). He then labelled the first as *formative* and the second and third as the *summative* purpose of assessment. Although there often exists a misleading distinction that formative assessment occurs during a lesson and summative occurs at the end of it, it should be noted that the expressions formative and summative correspond to different *purposes* of assessment rather than different *forms* or *types* of it (Dolin, Black, Harlen

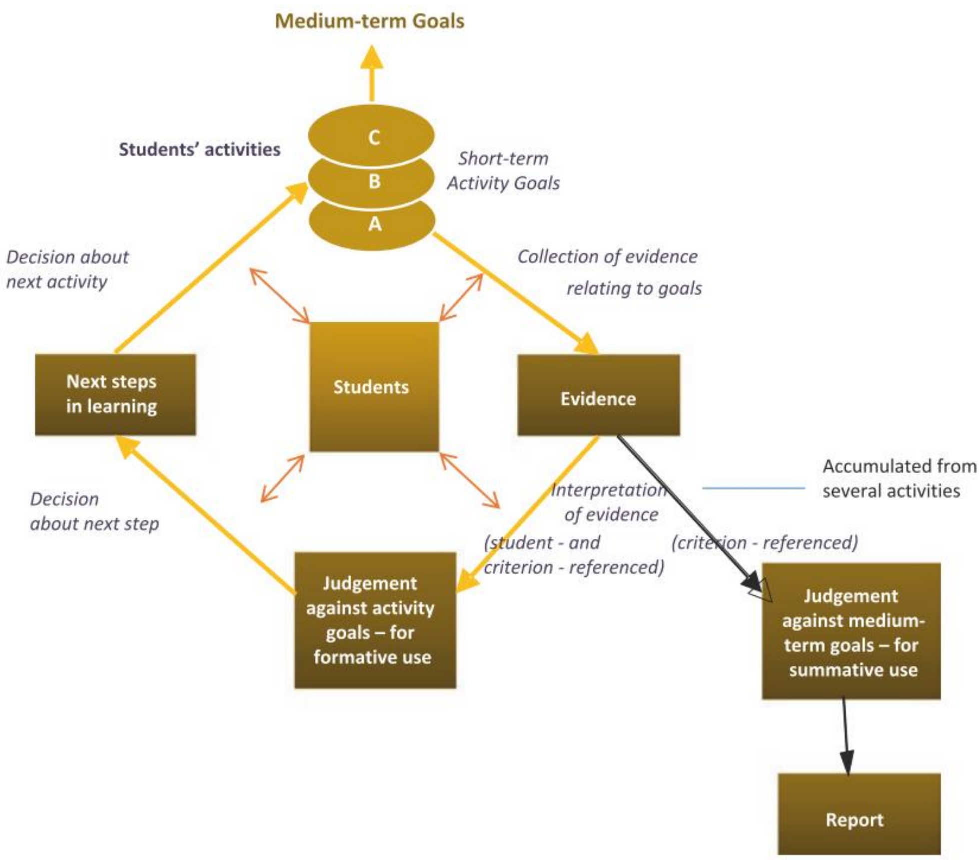


Figure 2.2 Assessment process that encompasses both summative and formative purposes (Dolin et al., 2018, p. 57)

& Tiberghien, 2018; Gardner, 2010). One assessment activity by teachers can be used both formatively and summatively; the evidence can simultaneously inform the next steps of teaching and indicate the overall achievement of a student to be reported to relevant stakeholders. Figure 2.2 illustrates that classroom assessment is an iterative and interactive process, showing that the formative-summative distinction relies on for what purpose the collected evidence is used. In the following, I examine each of the two purposes of assessment in detail.

2.2.2 Formative purposes of assessment

Formative assessment focuses on the role of assessment in supporting learning, often called 'assessment *for* learning' (or AfL). The concept was made popular in the late 1990s by Black and Wiliam (1998). The central interest of formative assessment is teacher and student feedback as a core element of assessment and learning process. In formative assessment, teacher practices such as sharing success criteria with learners, classroom questioning, comment-only marking, peer and self-assessment, formative use of summative assessment are crucial to effective learning (Black et al., 2003).[2] One widely accepted definition of formative assessment was given by Black and Wiliam (2009) as follows:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (p. 9)

Wiliam (2010) pointed out that such a definition highlights several key aspects of formative assessment: (1) Learners and peers, not just the teacher, can be agents of formative assessment; (2) The collected evidence should be used in decision-making; and (3) Formative assessment should inform instruction. From the above definition, Black and Wiliam (2009) suggested that formative assessment is concerned with 'the creation of, and capitalisation upon, 'moments of contingency' in instruction for the purpose of the regulation of learning processes' (2009, p. 6). Moreover, Shavelson et al. (2008) classified formative assessment with respect to its degree of formality. He suggested three main types of formative assessment in terms of formality. According to his classification, *formal* and *embedded* formative assessment comes from curriculum developers or teachers in advance of the instruction. *Planned-for-interaction* formative assessment is something that a teacher prepared deliberately to

---

[2] Although formative assessment is often taken to be synonymous with AfL (and summative assessment with AoL), some researchers distinguish between the two. For example, Black, Harrison, Lee, Marshall and Wiliam (2003) states: '*Assessment for learning* is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes "*formative assessment*" when the evidence is actually used to adapt the teaching work to meet learning needs'. (p. 10)

achieve the central learning goals of the lesson at the right moment. Lastly, *on-the-fly* formative assessment refers to an unexpected, unplanned occurrence of a 'teachable moment' during instruction (Shavelson et al., 2008).

This last type of formative assessment can occur informally during day-to-day activities undertaken by teachers and students (Black & William, 1998). As such, formative assessment occurs during classroom talk where the teacher and students engage in instructional dialogues, from which the teacher can recognise students' conceptions, communication skills, interests and use this information to guide instruction (Ruiz-Primo & Furtak, 2007). Duschl and Gitomer (1997) referred to such classroom talk as *assessment conversation*. According to them,

> The intent of an assessment conversation is to engage students in the consideration of a diversity of ideas or representations produced by class members and then to employ evidence and age-appropriate adaptations of scientific ways of knowing to foster a dialogue about what does and does not fit with the emerging thematic structure of the lesson. (pp. 39–40)

The timeframe for interpreting and teaching is more immediate in these informal formative assessments compared to formal formative assessment, and acting in response to evidence is more spontaneous and flexible (Ruiz-Primo & Furtak, 2004). Formative assessment can occur when, for example, the teacher or students respond with a question, elicit other students' opinions, rephrase what other students said. Such assessment conversations provide teachers with information about students' conceptions, strategies, mental models, language use and communication skills (Ruiz-Primo, 2011). From an analysis of the informal formative assessment in three middle school classrooms, Ruiz-Primo and Furtak (2007) suggested the 'ESRU model' of formative assessment, where the teacher *elicits* a question, the student *responds*, the teacher *recognises* and *uses* the information to support learning. Table 2.3 lists the specific formative assessment strategies that they found in the science lessons, categorised by the three dimensions of teacher practices.

2.2.3 Summative purposes of assessment

Besides its formative purposes, assessment in the classroom also plays the function of

Table 2.3 Strategies for ESRU cycles by dimension (Ruiz-Primo & Furtak, 2007, p. 63)

Eliciting	Recognising	Using
Epistemic frameworks		
Teacher asks students to: • Compare/contrast observations, data or procedures • Use and apply known procedures • Make predictions/provide hypotheses • Interpret information, data, patterns • Provide evidence and examples • Relate evidence and explanations • Formulate scientific explanations • Evaluate quality of evidence • Suggest hypothetical procedures or experimental plans • Compare/contrast others' ideas • Check students' comprehension	Teacher: • Clarifies/elaborates based on students' responses • Take votes to acknowledge different students' ideas • Repeats/paraphrases students' words • Revoices students' words (incorporates students' contributions into class conversation, summarises what student said, acknowledge student contribution) • Captures/displays students' responses/explanations	Teacher: • Promotes students' thinking by asking them to elaborate their answers (why, how) • Compares/contrasts students' responses to acknowledge and discuss alternative explanations/conceptions • Promotes debating and discussion among students' ideas/conceptions • Helps students to achieve consensus • Helps relate evidence to explanations • Provides descriptive or helpful feedback • Promotes making sense • Promotes exploration of students' own ideas • Refers explicitly to the nature of science • Makes connections to previous learning
Conceptual structures		
Teacher asks students to: • Provide potential or actual definitions • Apply, relate, compare, contrast concepts • Compare/contrast others' definitions or ideas • Check their comprehension		

measuring and reporting on students' achievement at a particular time, which is referred to as the summative purpose of assessment. For this reason, it is also called 'assessment of learning' (AoL) as opposed to AfL. The dominance of external large-scale summative assessments on classroom practice often makes teachers imitate those exam questions when designing their own summative assessment (Black & Wiliam, 1998). This characteristic of summative assessment in the classroom has traditionally led practitioners and researchers to consider the quality criteria such as validity, reliability and fairness more seriously than for formative assessment. In this thesis, I use 'summative assessment' to refer to school-based summative assessments planned and carried out by teachers within the context of classroom instruction. It thus excludes external, large-scale standardised assessments administered by governments and other authorities unless so specified. Likewise, although classroom teachers are often also responsible for administering external summative assessments in their classroom, these assessments will not be within the scope of this thesis.

A review of studies on teachers' summative assessment practices by Moss (2013) found several trends. First, it is known that students' perceptions of the classroom assessment environment and purposes impact their motivation to learn. Second, there is a discrepancy between what teachers perceive as an effective summative assessment and what they actually use in classrooms, which might be attributed to subject characteristics, teachers' knowledge of assessment. Third, teachers' judgements of student achievement differ from the results of standardised tests for various factors. For example, some studies reported teachers' tendency to rate a student in relation to the same year group in the school (norm-referenced assessment), rather than rating against the assessment criteria (criterion-referenced assessment) (Martínez, Stecher & Borko, 2009).

In the classroom, evidence for summative assessment can be collected in a range of ways such as paper-and-pencil tests, performance tasks, and written and oral products. Here I will focus on the assessment planned by the teacher in place of external summative tests designed and administered by the state. Aside from traditional paper-and-pencil tests, one popular method for summative assessment is a *performance-based* assessment that 'allows students to demonstrate their understandings and skills (to a teacher or an outsider) as they perform a certain activity' (NRC, 2001, p. 31). Performance-based assessments require students to be active participants and can be used as an alternative method for assessing students' higher-order skills that involve the application of knowledge in many countries (Darling-Hammond et al., 2010). An example of performance-based assessment in science lessons would be to ask students to plan, design and carry out an experiment about a certain topic and write up their findings. These kinds of activities allow assessing not only the cognitive understanding of the subject matter but also students' laboratory and communication skills (Darling-Hammond et al., 2010).

Since summative assessment serves the function of reporting student learning, it often leads to the award of qualifications such as pass/fail, grades and marks. These qualifications play important roles in student selection, admission to higher education, school reputation and so on, which makes ensuring validity and reliability of summative assessment a crucial consideration. Harlen (2005) reviewed the research evidence on teachers' summative assessment and suggested that providing detailed assessment criteria, describing developmental progression for learning goals, and allowing teachers to use information from across all student activities can be helpful to improve the validity and reliability of teacher summative assessment.

2.2.4 Conditions for high-quality classroom assessment

What determines the quality of classroom assessment? Key quality criteria such as validity and reliability have traditionally been developed for validation of standardised, large-scale assessments, so it should be first noted that concepts in large-scale assessments often need adaptation to be applied to classroom assessment (Brookhart, 2003). Core quality criteria for large-scale assessment such as validity, reliability and fairness should be adapted when applied in the context of classroom assessment because the purposes and characteristics of classroom assessment differ from those of large-scale assessment. For example, classroom assessment is connected to teachers' instruction, so it is difficult to be standardised across multiple schools and contexts (Bonner, 2013). For this reason, the validity of classroom assessment partly depends on its alignment with instruction in the classroom (Brookhart, 2013). Similarly, traditional psychometric approaches to ensuring reliability such as increasing the number of test items or using multiple scorers are often not an available option in the classroom assessment context (Parkes, 2013).

Calling for an 'indigenous' measurement theory suitable for classroom assessment, Brookhart (2003) pointed out that classroom assessment deviates from traditional large-scale assessment in several aspects: (1) Items or tasks are dependent on and nested within particular instructional contexts; (2) Items or tasks are not mutually independent but linked to one another in the learning sequence; and (3) Assumptions about large sample sizes do not apply to smaller class sizes. Acknowledging such fundamental differences, McMillan (2018) suggested eight key criteria for ensuring high-quality classroom assessment: clear and appropriate learning targets, alignment of assessment methods and learning targets, validity, reliability, fairness, positive consequences, alignment with standards, practicality and efficiency.

While these conditions are mainly applicable to classroom summative assessment, formative assessment has some additional quality criteria more suitable for its aim of facilitating learning through assessment. Particularly on formative assessment conversations in informal settings, Ruiz-Primo (2011) suggested five key conditions for effective assessment conversations to occur. First, effective assessment conversations should be guided by clear learning goals, which in turn can determine what constitutes evidence of having met these goals. Second, effective assessment conversations are dialogic and interactive, where students claim and justify their ideas and engage in discussing each other's arguments. Third, the teacher should strike a balance between supporting and challenging students to achieve the learning goals, providing appropriate instructional scaffolds. Fourth, a classroom environment that can support students' social participation should be formed. Fifth, assessment conversations should be an experience for students to immerse themselves into the culture of the scientific community as they engage in the argumentative practices in the classroom.

2.2.5 Teacher assessment literacy, assessment expertise and assessment identity

Classroom assessment is planned, designed and enacted by teachers, which makes the role of teachers essential to be considered. Researchers have been interested in the components that constitute teachers' ability to conduct an assessment, which is often referred to as 'assessment literacy'. Early ideas and standards on assessment literacy emphasised the measurement and test design aspects of assessment, and teachers' knowledge and skills with regards to these aspects (e.g., AFT, NCME & NEA, 1990). Since the 2000s, however, the concept of assessment literacy has been evolving as educational demands by society and learning environments have changed. Brookhart (2011) was critical of the measurement-centred characterisation of assessment literacy on the grounds that (1) it overlooks the formative use of assessment that occurs more frequently in lessons, provides feedback and informing lessons, and (2) it does not consider the changing accountability and standards-based reform contexts in the United

States. The latter issue was also connected with exercising equitable classroom assessment to students with diverse backgrounds and learning needs (DeLuca, 2019). Informed by the constructivist theory of learning, recent notions of assessment literacy consider such wider social and cultural contexts that dynamically interact with teachers' assessment practice, by viewing it as 'a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledge with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students' (Willis, Adie, & Klenowski, 2013, p. 242).

Specifically with regard to science teachers, Magnusson et al. (2006) proposed that science teachers' knowledge for assessment consists of two categories: knowledge of the dimensions of science learning that are important to assess, and knowledge of the methods and strategies of assessment. The former refers to what aspect(s) of the content knowledge should be assessed in a particular unit. The latter refers to which methods of assessment are more appropriate for a particular unit of study. Abell and Siegel (2011) offered a similar but more comprehensive model of science teacher literacy for assessment. Their model consists of two core elements (views of learning, assessment values and principles) and knowledge of assessment purposes, knowledge of what to assess, knowledge of assessment strategies, knowledge of assessment interpretation and action-taking (Abell & Siegel, 2011).

In harmony with the expanding concept of assessment literacy, Looney, Cumming, van der Kleij and Harris (2018) took a step further to propose 'teacher assessment identity'. Drawing on research on teacher identity from the sociocultural perspective, they suggested that teachers' identity as professionals, beliefs about assessment, perceptions of their roles as assessors, and disposition towards enacting assessment play a central role in teachers' assessment work. Their assessment literacy concept stresses the emotional dimensions of teacher identity related to assessment that pertain to how it 'feels' to assess students, as well as the agency (i.e., the ability to pursue valued assessment goals) of teachers in classroom assessment (Looney et al., 2018). Similarly, Lyon (2013) conceptualised and exemplified that science teachers' 'assessment expertise' can be organised around three elements: designing aligned and theoretically cohesive assessment, using assessment to support students' science learning, and equitably assessing language minorities. Together, research on teachers' assessment

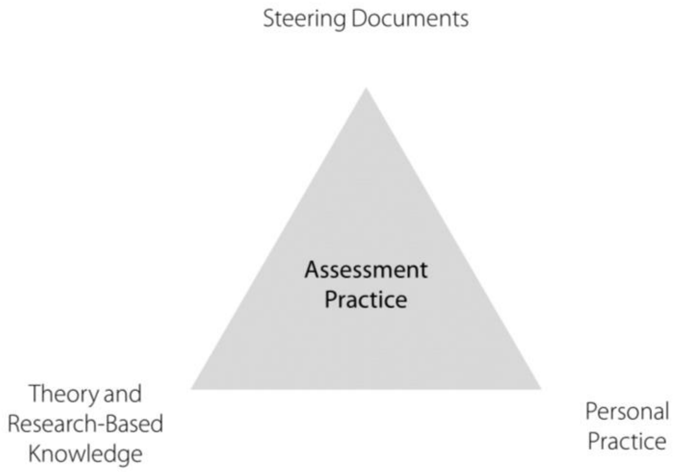


Figure 2.3 Three major factors related to assessment practice (based on Engelsen & Smith, 2010)

literacy, assessment expertise and assessment identity highlight that teachers' abilities to assess student learning is shaped as they negotiate their assessment knowledge in relation to various sociocultural and contextual factors such as prior experience, training, classroom and curriculum contexts, and policy (DeLuca & Braund, 2019).

2.2.6 Classroom assessment in the context of science teaching and learning

Science teachers engage in the practice of planning and conducting the assessment of students' learning throughout the school year. Classroom assessment of science learning involves the process of collecting information related to students' learning for making decisions (Bell, 2007; Champagne, 2014). Teachers use assessment for providing feedback to themselves and students (formative assessment) and monitor students' progress or improvement (summative assessment) (Bell, 2007). Recent policy developments in many countries have emphasised that teachers' formative and summative assessment have a significant impact on student achievement and motivations for learning (Black & Wiliam, 1998; Willis, 2010). Taking a wider perspective on classroom assessment, Black and Wiliam (2018) argued that teachers' implementation of classroom assessment is influenced by the subject discipline, theories of learning and pedagogy, and wider social, political and cultural contexts of education. Similarly, Engelsen & Smith (2010) indicated that there are three major factors that drive teachers' assessment in classrooms: steering documents (e.g., national and state curricula), theory and research-based knowledge and personal practice (Figure 2.3). Other studies have suggested a range of knowledge domains that teachers should be equipped with for assessment, such as views of learning, assessment values and principles, knowledge of assessment purposes, knowledge of what to assess, knowledge of assessment strategies, knowledge of assessment interpretation and action-taking (Abell & Siegel, 2011; Magnusson, Krajcik, & Borko, 2006).

Summative assessment in the science classroom generally serves the purpose of understanding students' progress and overall learning achievement at a certain point of instruction such as the end of a term (Bell, 2007; Black & Wiliam, 1998). This goal can be achieved in various ways. For example, teachers use written tests, observations, portfolios of student work, in-class activities and tasks to summatively assess student learning (Dolin et al., 2018). The collected information is used for reporting individual students' progress to parents, school and students themselves. In some cases, such information is 'high stakes' because it contributes to the evaluation of schools and teachers, funding and also students' entrance to prestigious schools and universities (Dolin et al., 2018). For this reason, summative assessment is often perceived as a major and infrequent event with little feedback expected (Sadler, 2014). Recent views tend to pay more attention to the role of summative assessment in enhancing classroom learning. For example, Black and colleagues (2011) stressed the interaction between formative and summative assessment by embedding summative tasks within formative in-class work. In addition, as seen in Figure 2.4, they noted that 'the design of educational activities and associated assessments is influenced by the theories of pedagogy, instruction and learning, and by the subject discipline, together with the wider context of education' (p. 551). This diagram well depicts that summative assessment in the classroom is not an isolated and independent event as traditionally conceived.

Although the studies described so far provide a useful background to teachers' knowledge and practices in science assessment, it is important to recognise their limits, particularly in terms of the extent to which these studies can provide accounts for the assessment of NOS. As noted earlier, NOS is mostly taught by science teachers, but the content and aim of teaching NOS have distinct aspects from those of science. This

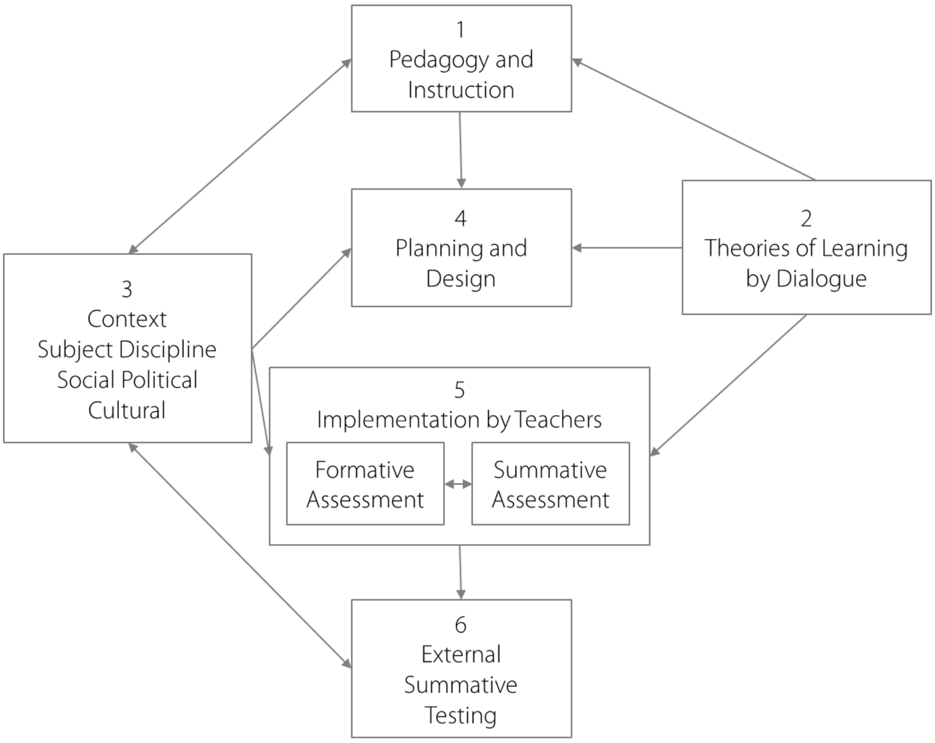


Figure 2.4 Model for assessment in relation to pedagogy (Black & Wiliam, 2018, p. 556)

brings our attention to the relationship between classroom assessment and subject matter.

2.2.7 Subject- and topic-specificity of classroom assessment

Research on teacher knowledge has highlighted the need for understanding 'teachers' cognitive understanding of subject matter content and the relationships between such understanding and the instruction teachers provide for students' (Shulman, 1986, p. 25). Shulman (1987) proposed the concept of PCK to describe the teacher knowledge for effective teaching of subjects. PCK is defined as:

> that special amalgam of content and pedagogy that is uniquely the providence of teachers, their own special form of professional understanding ... Pedagogical content knowledge ... identifies the distinctive bodies of knowledge for teaching. It represents the blending of content and pedagogy into an under- standing of how particular topics, problems, or issues are organised, represented, and adapted to diverse interests and abilities of learners, and presented for instruction. Pedagogical content knowledge is the category most likely to

> distinguish the understanding of the content specialist from that of the pedagogue. (Shulman, 1987, p. 8)

Veal and MaKinster (1999) proposed a three-level taxonomy of PCK. General PCK refers to teacher knowledge about the concepts and strategies specific to disciplines, such as science, art, mathematics, language, and history. Domain-specific PCK concerns different domains or subject matters within a particular discipline. This includes PCK specific to physics teaching, chemistry teaching, biology teaching and so on. Finally, there are topic-specific PCK. Thermodynamics and atomic structures are taught in both physics and chemistry but introduced differently, which requires teachers of each subject more topic-specific knowledge to properly addressing these topics (Veal & MaKinster, 1999). Given such domain- and topic-specificity of teacher knowledge, it can be reasonably argued that NOS needs separate attention from scientific content knowledge when it comes to teacher knowledge. The knowledge for teaching physics concepts would not be the same as that for teaching the interactions of science and society, although these two topics are not entirely isolated. This is also true for teacher knowledge of assessment, which is a core component of PCK (Abell & Siegel, 2011; Xu & Brown, 2016).

A similar argument can be made about teacher assessment practices. While some practices of classroom assessment seem to be more or less shared across subjects, there exist features that are more specific to certain subjects and topics being taught (Black et al., 2003). McMillan (2001) investigated 150 US secondary teachers of science, social science and English and identified statistically significant cross-subject differences in terms of factors used in determining grades, type of assessment and cognitive level of assessment. Science teachers were found to use teacher-made assessments more frequently than provided ones, relative to English teachers. In terms of the cognitive level of the assessment, science assessments required less higher-order thinking but more recall than English assessments (McMillan, 2001). Marso and Pigge's (1993) review of research on classroom assessment practices showed that essays are more frequently used in English, history and social studies classrooms, and science and mathematics teachers conduct tests, in general, more often than other subject teachers. Black et al. (2003) concluded from a study with secondary schools in England that cross-subject divergences in teachers' classroom assessment practices are 'not inherent in the subject, but rather are consequences of the way that the subject is interpreted in the school curriculum' (p. 68). In their study, mathematics and science teachers tended to identify the subject matter as definitive of their subject, which in turn dictates the instructional aims to be shared in class, whereas English teachers allowed a range of goals appropriate for each student's needs (Black et al., 2003).

The fact that the content and method of classroom assessment should be adjusted to suit the subject characteristics naturally takes us to the issue of assessing NOS. As discussed earlier, NOS ideas are different from scientific knowledge in several important aspects, which necessitates the employment of a different teaching approach characterised as explicit and reflective instruction (Sections 2.1.1 and 2.1.2). This, in turn, raises questions about the extent to which existing accounts of classroom assessment in science are valid in the context of assessing NOS. Can we use the same strategies and quality criteria for science assessment and NOS assessment? If not, how should they differ? In the next section, I will turn to the research on the assessment of NOS specifically, to consider this point.

2.3 Assessment of students' NOS understandings

2.3.1 NOS assessment: by whom, and for whom?

Existing studies on the assessment of NOS, whether theoretical or empirical, are almost exclusively focused on assessing learners' NOS knowledge by researcher-developed instruments (Hanuscin et al., 2020). Neither of the two recent, comprehensive reviews of NOS assessment (Abd-El-Khalick, 2014; Lederman, Bartos & Lederman, 2014) that appeared in authoritative handbooks discussed how to assess NOS in classrooms. The preponderance of small, medium and large-scale administration and evaluation of NOS questionnaires is, in a way, reminiscent of the time in the 1990s when science educators focused on identifying learners' misconceptions in various areas of science and developing methods to change them. Such studies have made valuable contributions to identifying the gaps in students' and teachers' NOS ideas and developing effective instructional methods to improve them.

What appears to be missing in the discourse is the question of who does the assessment of NOS and to whom the collected evidence brings benefits. Research instruments are pre-designed by researchers, so there is no chance of teachers' instructional considerations being incorporated in the assessment items. On the contrary, classroom assessment is planned, enacted and used by teachers and provides information and qualifications to students, the teacher, parents, schools and so on (see Section 2.2.3). As noted by McComas et al. (2020), NOS assessment approaches that require a significant amount of time and training would be implausible to use for classroom teachers. It is therefore necessary to distinguish NOS assessments into those conducted by researchers for research purposes, and those conducted by teachers for school purposes. Recognising such a lack of consideration for NOS assessment in the classroom, several researchers have highlighted that there is an urgent need for designing and providing assessment instruments for classroom use by teachers (Bell, 2008; Ryder, 2009).

2.3.2 Researcher-developed instruments for assessing NOS

Since NOS was first introduced as a learning goal in school science, there has been a great interest among science educators in assessing learners' understanding of NOS (see Abd-El-Khalick, 2014 for a comprehensive review of existing instruments). The most widely used instrument at present is *Views of Nature of Science* (VNOS; Lederman et al., 2002) and its several variants, which consist of open-ended questions about NOS. More recently, the *Views about Scientific Inquiry* (VASI) questionnaire has been developed separately from VNOS to assess students' understanding of the nature of scientific inquiry (Lederman et al., 2014; see **Table 2.1** for a longer list of NOS instruments). In addition to questionnaires, other techniques such as Draw-a-Scientist-Test (DAST), concept mapping, classroom observation and small group discussion have also been occasionally used to collect information about students' NOS views (Hodson, 2014).

While researcher-developed instruments can provide some insights to inform classroom instruction and assessment of NOS, it should be noted that classroom assessment has aims that are distinct from those of standardised measurement for research purposes. There are three major reasons that make it particularly difficult to use researcher-developed instruments for classroom assessment (Allchin et al., 2014).

First, most existing instruments are based on context-general ideas of NOS, whereas learners' responses tend to be context-dependent (Sandoval & Morrison, 2003; van Dijk, 2011). Classroom assessment occurs in a specific curricular and instructional context and for collecting information about student learning situated within that particular context, rather than context-independent knowledge of NOS (McComas 2020, Chapter 4). Another reason is that teachers have other considerations for classroom assessment such as the limitations of time and motivation, and the institutional context of accountability (Bell, 2008; Sadler, 2014). Stressing the importance of tailoring the assessment instruments to meet the needs of individual teachers and students, Bell (2008) suggested several alternative approaches to assessing NOS for formative and summative purposes such as draw-a-scientist test, informal conversation and science notebooking. While these can be useful strategies for collecting information about students, there remains a question of how they can be adapted to serve the purpose of summative assessment.

Finally, as noted by Allchin (2011), these research instruments do not aim to 'evaluate' students' understanding of NOS so much as to describe and characterise students' views about NOS from a rather neutral viewpoint. While such a descriptive understanding of students' ideas is indeed a part of teachers' professional responsibility, from an assessment perspective, this is not sufficient information for both formative and summative purposes. Consider the currently most popular instrument, VNOS-B, for example (Abd-El-Khalick, Bell & Lederman, 1998). It contains six open-ended questions about NOS, responses to which are usually marked as one of the three categories, naïve, mixed and informed (e.g., Akerson & Abd-El-Khalick, 2000; Bell et al., 1998) based on the coding scheme. From this information, researchers are interested in the distribution of students' views across the three categories as well as the qualitative aspects of them, comparing these results across different school and grade levels, countries and genders, and tracking changes between pre- and post-interventions in diverse grade levels and instructional settings. While these functions have some overlap with the aims of classroom assessment, they are quite distinct from what teachers need and are interested in in the classroom context. Teachers are keen on facilitating learning by providing feedback to student responses (formative functions of assessment) and producing evidence of students' development at a particular time (summative functions of assessment). They embed assessment in diverse forms of learning activities such as teacher-student interactions, peer interactions, group activities and portfolios, besides written tests done in silence. These differences call for considering the classroom assessment of NOS as a distinct activity of teachers that has different aims, methods and criteria for quality from the assessment through research instruments such as VNOS.

2.3.3 Research on classroom assessment of NOS

In view of the explicit-reflective approach (see Section 2.1.2) where the understanding of NOS is, by definition, 'intentionally planned for, taught and assessed' (Lederman, Schwartz, Abd-El-Khalick, & Bell, 2001, p. 137), both formative and summative assessment should be considered an essential condition for the success of NOS instruction. In contrast to the richness of research on teachers' instructional planning and teaching practice of NOS, however, there is currently minimal research, both theoretical and empirical, on teachers' instructional practices and strategies for assessing NOS in classrooms (Allchin, 2013; Hanuscin et al., 2011). The existing studies on classroom assessment of NOS are briefly summarised in Table 2.4.

One fact that we know from the literature is that in many cases, teachers tend not to assess students' NOS understanding in their classrooms, neither formatively nor summatively. In a study by Abd-El-Khalick, Bell and Lederman (1998a), none of the 14 teachers who participated in their study made any attempt to formally assess NOS learning, although most of them believed in the importance of teaching NOS in schools. In a study with elementary teachers on a 3-year professional development programme, Hanuscin et al. (2011) echoed this point by reporting that teachers mostly relied on using the instruments provided by researchers, but no attempts were made to use their own formative/summative assessment strategies to determine students' learning of NOS. Such discrepancy between the perceived importance and assessment practice of NOS has been confirmed in another recent study of English science teachers teaching Key Stage 4 (Brock & Taber, 2019).

Park, Yang and Song (2020) touched on the problem of classroom assessment of NOS by focusing on a specific form of assessment tasks, namely text-based tasks. From their analysis of textbook tasks on NOS, they found three different ways in which

Korean SIE textbooks prompt the learning of NOS. They identified three types of NOS-related tasks: (a) guiding to NOS ideas (by performing simple tasks from memory, identifying NOS in a historical case, explaining a historical case based on NOS, identify details about NOS through a historical case, or compare different NOS aspects), (b) expanding NOS ideas (by searching for historical examples that show a certain NOS aspect, thinking about reasons for a NOS aspect, or evaluating the applicability of NOS), and (c) thinking critically about NOS (by discussing limitations of a NOS aspect, or taking sides and defend one's own view about an issue with NOS understanding) (Park, Yang & Song, 2020, p. 436). Although their study focused on textbook tasks, the findings are relevant to classroom assessment more generally, because such a typology of NOS activities can be used for both formative and summative purposes.

Several studies have considered the formative assessment of epistemic practices of science. An early study by Duschl and Gitomer's (1997) focused on the potential of 'assessment conversations', an instructional strategy to make public students' ideas and reasoning, in fostering students' skills in the evaluation of knowledge claims and the construction of reasoning. They highlighted the need for teacher-researcher collaboration to promote the use of assessment conversations in the classroom since teachers 'are not used to using student information to guide and revise instructional decision making' (p. 65). Similarly, Ruiz-Primo and Furtak (2006) found that how teachers use informal formative assessment during instruction may be associated with student learning in scientific inquiry, using the ESRU cycle. These studies, although not directly concerned with NOS, suggest that the epistemic practices of science can be taught effectively through the use of instructional strategies that allow for ongoing, frequent assessment of student learning during instruction.

Another related, and problematic, phenomenon in the NOS scholarship is that there have been surprisingly few empirical investigations on how NOS is (and should be) assessed at the classroom level. The scarcity of research on the classroom assessment of NOS could have possibly stemmed from the fact that in most cases NOS is treated implicitly in science curricula and not considered as separate learning outcomes that should be assessed. Among the few exceptions that addressed the classroom assessment of NOS is Akerson et al.'s (2010) study. They found that experienced primary teachers used various assessment strategies such as a modified version of VNOS-D2 (Views of Nature of Science; Lederman & Khishfe, 2002) and student journals, and that teachers considered their students' developmental levels, special needs and academic abilities when assessing the NOS understanding (Akerson et al., 2010). Brock and Taber (2019) reported that English science teachers find NOS particularly hard to assess fairly due to the ambiguity of assessment criteria, which became a major source of challenge in their teaching of NOS in schools. However, since these studies were conducted in the context of a university-based professional development programme (Akerson et al.) or a school curriculum where NOS is almost absent (Brock and Taber), currently there is minimal understanding of teachers' assessment practices of NOS in science classrooms.

In Schwartz et al. (2002)'s study with 13 US secondary teachers who were engaged in a year-long professional development focused on NOS and scientific inquiry, the teachers did not make efforts to assess NOS, similar to other studies. According to them, the teachers tended to mention NOS explicitly and assume that it was clearly understood by students. They also suggested two possible reasons that the teachers did not assess NOS: First, the teachers were not comfortable using formal assessment methods to assess students' NOS views because they considered doing so would be 'unfair'. Second, they struggled with understanding the NOS concepts themselves, which made them sceptical about whether they could assess them in the classroom (Schwartz et al., 2002). Mesci et al. (2020) explored the teaching practices of two US preservice teachers who participated in a 13-month teacher development programme. These two preservice teachers, who were selected as the most successful cases among the 11 participants of their programme, planned for and used some formative assessment techniques during their teaching practicums. They used 'exit tickets' and essay writing to identify students' prior knowledge, misunderstandings or confusions about the target NOS aspects, and acted upon this information to improve their lessons by giving feedback and more specific examples. Drawing on the framework of PCK, they proposed teachers' knowledge of assessment of NOS to be a major element of 'PCK for NOS' (Mesci et al., 2020).

2.3.4 Developmental appropriateness and learning progressions for NOS

One main issue related to assessing NOS is the emerging arguments for the developmental approaches to NOS, taking the cue from the research on learning progressions (LPs) (Duncan, Rogat & Yarden, 2009; Smith, Wiser, Anderson &

Krajcik, 2006). NOS researchers have been keen on structuring the NOS learning in a developmentally appropriate manner (Abd-El-Khalick, 2014; Kampourakis, 2016; Lederman et al., 2014), which is also central to the assessment of NOS given that assessment strategies must be suitable for the grade level of students. For example, Erduran and Dagher (2014a) capitalised on the importance of articulating the scope and sequence of NOS across grade levels by distinguishing two dimensions of such articulation: *vertical* and *horizontal*. Vertical articulation is concerned with ensuring the coherence in the educational experiences that a child gets as they progress through the school years. In NOS terms, this corresponds to organising the aspects of NOS from relatively simple and straightforward ideas (e.g., '*Scientific knowledge is tentative and subject to change*', '*Scientists use multiple methods'*, '*Science and society are interrelated'*) to more sophisticated ones (e.g., variations in experimental methods, different ways in which science influences society and vice versa, cross-domain differences in epistemic values). In comparison, horizontal articulation refers to structuring NOS learning experiences across different topics and domains so that students can develop connections among seemingly discrete scientific domains, topics and instructional units drawing on their understanding of NOS. An example is to teach about the roles and methods of observation in science with an example from forces, and then extend this to observation in other examples and domains such as Darwin's observation of finches, Watson and Crick's observation of DNA double helix, and the recent detection of the Higgs boson and gravitational waves using cutting-edge technologies. Similarly, Abd-El-Khalick (2014) suggested a 'developmental framework' that involves a progression of NOS benchmarks spanning from primary schools to universities and teacher education programmes. Such developmental considerations are also partly reflected in NGSS (Table 2.6). NGSS presents key aspects of NOS and sets out the progression of each NOS aspect from a simple statement (K–2 level) to sophisticated and elaborate ideas related to it (high school stage), although the benchmarks are still largely focused on declarative statements as seen in the table. Given that teachers, as frontline practitioners, are usually keen on their own students' developmental needs and trajectories (Heritage, 2013), these considerations should also be important factors for successful classroom assessment of NOS.

Table 2.4 Examples of previous studies that addressed classroom assessment of NOS

Author	Region	Grade level	Participants	Context	Data sources	Findings related to classroom assessment
Abd-El-Khalick, Bell & Lederman (1998a)	United States	Secondary	14 preservice teachers	A NOS-focused teacher preparation programme	Lesson plans, lesson videos, portfolios, clinical observation notes	• No preservice teachers made an attempt to formally assess NOS. • This means that a significant discrepancy exists between their beliefs in the importance of teaching NOS and failure to assess NOS.
Hanuscin et al. (2011)	United States	Primary	3 experienced teachers	A single-school professional development programme focused on NOS and inquiry	Interviews, questionnaires, classroom observations, classroom artefacts	• While the teachers had general knowledge of assessment, they made no attempt to use formative/summative assessment strategies to assess students' NOS understandings. • Instead, teachers only relied on the research instrument (VNOS-D) as evidence of the impact of their NOS instruction.
Brock & Taber (2019)	England	Upper secondary (Key Stage 4)	6 in-service teachers	A 11-18 academy with above-average students	Interviews	• Teachers perceived that NOS content is exceptionally difficult to assess fairly. • Teachers perceived that NOS-related assessment criteria are often unclear and insufficient.
Akerson, Cullen & Hanson (2010)	United States	Preschool and primary (K-4)	10 in-service teachers	A NOS-focused 2-year professional development programme	Interviews, action research designs, lesson plans, assessment tools	• Experienced primary teachers used various assessment strategies. • Teachers considered their students' developmental levels, special needs and academic abilities when assessing NOS.

Table 2.5 (continued)

Author	Region	Grade level	Participants	Context	Data sources	Findings related to classroom assessment
Schwartz, Lederman, Khishfe, Lederman, Matthews & Liu (2002)	United States	Secondary	13 in-service teachers	A 1-year professional development programme with NOS and scientific inquiry focus	Lesson videos, lesson plans, reflections, student work	• Teachers mentioned NOS explicitly and assumed that it was clearly understood by students. • Teachers were not comfortable using formal assessment methods to assess students' NOS views and thought that doing so would be unfair. • Teachers' own struggles with NOS concepts made them question the possibility of assessing NOS.
Ryder & Leach (2008)	England	Upper secondary	7 experienced in-service teachers	Advanced level science courses using specially designed lesson materials	Lesson audios	• Teachers varied in their use of the four characteristics of effective teaching about the epistemology of science: making appropriate statements about NOS, linking NOS to science content, working with students' ideas of NOS, stating and justifying learning goals.
Herman, Clough & Olson (2013)	United States	Secondary	13 early career in-service teachers	An intensive science teacher preparation programme with NOS focus	Lesson observations, interviews, classroom artefacts	• Four of the 13 teachers (high implementers) included NOS-related questions in their summative assessments, in a variety of contexts. • Five of the 13 teachers (medium implementers) used decontextualised NOS questions in their summative assessments.
Mesci et al. (2020)	United States	Secondary	2 preservice teachers	A 13-month teacher development programme with NOS and scientific inquiry focus	Open-ended surveys, interviews, observations, lesson plans, reflections, teaching materials	• Participants did not include assessment in the first draft of the lesson plan. • After the expert feedback, they planned for and used formative assessment techniques.

Table 2.6 Example of a developmental approach to NOS in NGSS (NGSS Lead States, 2013, Appendix H, p. 98)

	K–2	3–5	Middle school	High school
Scientific investigations use a variety of methods.	• Scientific investigations begin with a question. • Scientists use different ways to study the world	• Scientific methods are determined by questions. • Scientific investigations use a variety of methods, tools, and techniques	• Scientific investigations use a variety of methods and tools to make measurements and observations. • Scientific investigations are guided by a set of values to ensure the accuracy of measurements, observations, and objectivity of findings. • Science depends on evaluating proposed explanations. • Scientific values function as criteria in distinguishing between science and non-science.	• Scientific investigations use diverse methods and do not always use the same set of procedures to obtain data. • New technologies advance scientific knowledge. • Scientific inquiry is characterised by a common set of values that include logical thinking, precision, open-mindedness, objectivity, scepticism, replicability of results, and honest and ethical reporting of findings. • The discourse practices of science are organised around disciplinary domains that share exemplars for making decisions regarding the values, instruments, methods, models, and evidence to adopt and use. • Scientific investigations use a variety of methods, tools, and techniques to revise and produce new knowledge

2.4 Summary of literature review

In this chapter, I have examined the theoretical and empirical research bases in the intersections of three main concepts of this thesis, namely NOS, teachers and classroom assessment. In sum, there seem to be a variety of potential factors that stem from the interplay of teachers and assessment, and of science teachers and NOS. These range from epistemic considerations such as the controversial nature of NOS knowledge to pedagogical and systemic considerations such as developmental appropriateness and the influence of external examinations. Also, it has been identified that there is a significant lack of interest and effort in researching teachers' experiences of engaging in the assessment of NOS. Most of the existing studies only anecdotally, if anything, addressed the assessment issues related to NOS, but their initial findings provide a useful starting point. For example, *why* did the teachers think it is unnecessary and unfair to assess NOS? How much NOS knowledge should a teacher have to be qualified to assess it? Most importantly, if teachers did assess NOS in classrooms, what would it look like? How would their practices and concerns play out in formative and summative assessment? What would be teachers' considerations specific to formative and summative assessment, respectively? This thesis will explore these issues about NOS assessment through an empirical case study.

Another deficit in the current landscape of NOS research concerns methodological simplicity. In their critique of the research trend on NOS, Kelly, Chen and Crawford (1998) pointed out that the majority of empirical investigations of NOS rely on surveys and interviews without including observational data with teachers and students. They viewed such practice as bearing the risk of 'mentalistic modes' of describing human action (p. 30), which cannot fully represent people's behaviour alone. Instead, they argued for a social semiotic research programme of NOS that focuses on language use in classrooms and how meanings about NOS are co-constructed through classroom conversations (Kelly et al., 1998). Although their criticism dates back two decades, NOS research still suffers from the lack of observational research and overreliance on surveys and interviews, with only a few exceptions (Erduran, Kaya, Cilekrenkli, Akgun & Aksoz, 2020; Hanuscin et al., 2011; Ryder & Leach, 2008). Given this gap in current NOS literature, this thesis has the potential to make a significant contribution to research through observation of actual practices in the classroom assessment of NOS.

2.5 Theoretical orientation to NOS and classroom assessment

Having reviewed existing studies on NOS, classroom assessment and the classroom assessment of NOS, in this section, I describe my theoretical commitments for this study. These commitments were developed through my previous engagement with NOS research (Park, Erduran & Wu, 2020a; Park, Erduran & Wu, 2020b; Park & Song, 2019; Park, Yang & Song, 2019; Park, Yang & Song, 2020), my other studies about the broader links between science education and history, philosophy and sociology of science (Park, 2020; Park & Song, 2018), my experience of teaching NOS in the teacher education context, and my reading of relevant studies in science education and classroom assessment that have been offered in this chapter.

2.5.1 NOS

I use NOS to mean ideas about scientific knowledge and scientific inquiry that are drawn from history, philosophy and sociological studies of science. As it pertains to this thesis, I consider NOS to encompass both the aspects of scientific knowledge and scientific inquiry (Irzik & Nola, 2014; Hodson, 2014). My view is aligned with that of Allchin (2011, 2013) who argued that the focus of NOS education should move from listing up what science is to providing a systematic and unifying perspective to 'whole science'. Also, I take NOS to be a cognitive learning outcome (Abd-El-Khalick et al., 1998; McComas, Clough & Nouri, 2020) but also a functional aim of science education (Allchin, 2011; Ryder, 2001), in the sense that NOS education should eventually aim to equip learners with the competence to interpret the reliability of scientific claims in their personal and social decision making (Allchin 2011). Moreover, I agree with Matthews' (2014) assertion that NOS is to be 'elaborated, discussed and inquired about' (p. 18) rather than to be learnt as declarative statements. For example, in the context of the 'paradigm shift' that is the topic of Chapter 7, I believe that the instructional goal should not be rote learning of the statement that 'Science develops through the shift of paradigms' or identifying this statement as correct in a multiple-choice question. Instead, my position is that the learning goal should include inferring such an aspect of NOS from examples and applying one's understanding to appreciate examples in light of the general idea of the paradigm shift. In other words, the focus of instruction and assessment should be on learning *about* the paradigm shift idea and its connections with examples, besides learning *that* it is a feature of science. Similar arguments can be made for other tenets in the consensus list and other NOS ideas that are not on the list.

In terms of what to teach about NOS (i.e., the selection of NOS aspects for the curriculum), I depart from the consensus view in that I think the NOS curriculum should be open-ended about NOS aspects rather than listing them up as a limited number of tenets. The list-based approach is often advocated on the grounds that there always needs to be some form of didactic transposition, in the sense that we cannot give every detail about the topic but should highlight the ones that align with the learning goal (Kampourakis, 2016). Although I agree with the need for didactic transposition, I consider that there are more effective ways to transpose the philosophical ideas about science than reducing the rich and diverse NOS ideas to a number of statements. Consider the comparison of laws and theories from the consensus list, which I think is indeed a valuable NOS topic to be discussed in the classroom. It is certainly an *example* of what science is, but it is difficult to justify including this NOS as one of the seven (or more) NOS to be taught in schools over many other ideas about what science is and how it works. Educators should be open to teaching other examples of NOS ideas that are not on the list. This is the main reason why I use FRA as a conceptualisation of NOS since it offers a possibility to include broader issues and ideas about NOS compared to other frameworks (Dagher & Erduran, 2017).

As for the effective instruction of NOS, I am committed to the explicit-reflective approach to NOS instruction that is supported by much empirical evidence (Abd-El-Khalick et al., 1998; Abd-El- Khalick & Akerson, 2009; Akerson, Abd-El-Khalick, & Lederman, 2000). Therefore, I consider that teachers should explicitly aim to teach NOS in their classroom and communicate the NOS learning goal to students, rather than expecting NOS to be learnt naturally from learning scientific content knowledge, reading historical episodes in science, or ‘doing science’. Students must be given opportunities to reflect on the connections between these activities and NOS, and between diverse aspects of NOS (Abd-El-Khalick, 2012). This is not to say that every science lesson should highlight NOS; The emphasis on NOS as an instructional goal can vary depending on what other goals the teacher has for the lesson. Still, I view that NOS learning can be best achieved when in tandem with other goals such as content knowledge and inquiry skills, as I will further develop in Section 8.6.

2.5.2 Classroom assessment

On the nature of classroom assessment, my point of departure is aligned with that of Brookhart and McMillan (2019), who consider classroom assessment as ‘a process that teachers and students use in collecting, evaluating and using evidence of student learning’ (p. 4). I then distinguish formative and summative assessment based on whether the purpose of such a process is chiefly to support learning, report achievement or satisfy the accountability demands (Black, 1998). This way of classification is distinguished from the one based on the timing of classroom assessment (i.e., whether it occurs during or at the end of a lesson). Nevertheless, I view the relationship of formative assessment and summative assessment to be reciprocal rather than exclusive, given that a classroom activity can be used to both support learning and report achievement (Harlen, 2011).

When it comes to formative assessment, following Ruiz-Primo and Furtak (2007), I focus on two broad types of formative assessment: formal and informal. I consider formal formative assessment to be those embedded in the curriculum and planned by the teacher before instruction to check for student learning on the topic. This could be the teacher asking a question about a specific aspect of learning at the start of the lesson, or using a quiz, brainstorming and other forms of group activities that can produce evidence of student learning (Bell & Cowie, 2001). On the contrary, in the case of informal formative assessment (often called ‘on-the-fly formative assessment’ or ‘assessment conversations’), the collection and interpretation of information occur as part of teacher-student interaction in the classroom, when the teacher elicits student ideas, the student responds and the teacher acts upon the response (Bell & Cowie, 2001; Ruiz-Primo & Furtak, 2007). However, such cycles may not always be complete. An example of an incomplete cycle is when a teacher asks mostly yes-or-no questions to students that can elicit minimal information about student learning, and then merely revoices student responses without using the information. In Chapter 7, I discuss how the teacher used the planned activity (i.e., formal) and characterise her use of information collected during whole-class interactions about the activity to promote students' NOS learning (i.e., informal), based on this conceptualisation of formative assessment in the classroom.

2.5.3 Classroom assessment of NOS

Discussion about classroom assessment of NOS should be based on the aforementioned aspects of NOS and classroom assessment. Putting together the two, I consider classroom assessment of NOS to mean the process in which teachers collect evidence of student learning about what science is and how it works, to either promote learning or report achievements. Starting from such a conception of NOS classroom assessment, this thesis aims to explore the diverse themes that can emerge at the interface of NOS and classroom assessment. My position is that if we were to promote NOS as a learning goal for achieving scientific literacy, both formative and summative assessment should be intentionally planned for and enacted in the science classroom (Hanuscin et al., 2020). This, in turn, highlights that science teachers should have the ability to practise the classroom assessment of NOS and that they need to be provided with training opportunities to develop such ability. At the same time, however, any suggestions about classroom assessment should be practical; I concur with McComas, Clough and Nouri (2020) that 'assessment approaches that require a significant amount of time to administer, along with time and expertise (training) to score, are unreasonable for classroom teachers' (p. 98). Therefore, a core task for NOS researchers is to suggest effective as well as realistic ways to promote NOS classroom assessment.

As argued in Section 2.3.1, the classroom assessment of NOS should be distinguished from assessing NOS for research purposes. I agree with the view of Ryder (2009) that most research instruments developed for educational research are insufficient and inappropriate for use in the classroom, due to the reasons discussed in this literature review (Section 2.3.2). When it comes to the learning goals for NOS instruction, my view aligns with Allchin (2013) who said that the goal of NOS education is knowledge and understandings rather than views and perspectives. It follows that what the teachers should assess in the classroom is closer to whether a student *understands* what science is and how it works rather than whether they *agree* or *disagree* with a particular view about science (Allchin, 2013). Such a take is crucial in considering, for example, the suitability of using 'views-oriented' questionnaires for summative classroom assessment, as I analyse in Chapter 6 (see also Section 8.2 where I discuss this point further). This position also suggests that not all students' ideas about NOS should be treated as equally valid and acceptable, as Ryder and Leach (2008) stressed.

2.6 Research questions

With such theoretical positioning, this thesis investigates how teachers practise the summative and formative assessment of NOS. The goal is to shed light on an underexplored territory within NOS research and make a contribution that can eventually lead to improving the practice of NOS assessment in classrooms. This thesis is centred on three questions about science teachers' classroom assessment of NOS:

RQ1. How do teachers view and practise the summative assessment of NOS in the classroom? (Chapter 5)

RQ2. How does a team of three science teachers engage in NOS summative assessment in a high school? What does their decision-making process reveal about the potential issues in the summative assessment of NOS? (Chapter 6)

RQ3. What are the instructional strategies that allow for teachers' formative assessment NOS? What are some potential issues and challenges involved in the formative assessment of NOS? (Chapter 7)

The first question focuses on the summative assessment aspect and asks how the teachers' views and the contexts of curriculum and accountability aligned with their practices. The second question specifically looks at the collective aspect of their NOS summative assessment. The third question asks how the teachers enacted formative assessment while they teach NOS-centred lessons. It is important to note that, in this study's context, 'summative assessment' (RQ1 and RQ2) and 'formative assessment' (RQ3) do not refer to two separate events but rather two different purposes for which a single set of activities were used. Although the entanglement of formative and summative assessment is not unique to this study (Dolin et al., 2018), there were

Table 2.7 Overview of study design

	RQ1	RQ2	RQ3
Study objectives	To identify each teacher's views and practices of NOS summative assessment	To identify the practices of summative assessment and the challenges of NOS summative assessment	To characterise the teacher's NOS formative assessment strategies
Focus	Summative assessment	Summative assessment	Formative assessment, instructional strategies
Subject	Three teachers who enacted summative assessment of NOS	A team of three teachers who collectively planned and marked NOS assessments	One teacher who enacted the formative assessment of NOS
Main data (*primary sources)	Assessment plan* Assessment meeting recording Interviews* Lesson recordings* Lesson and assessment materials	Assessment plan Assessment meeting recording* Interviews* Lesson recordings Lesson and assessment materials	Lesson recordings* Lesson materials Interviews
Analytical approach	A mixture of deductive (key assessment concepts, NOS concepts) and inductive coding (emergent codes from data)	A mixture of deductive (key assessment concepts, NOS concepts) and inductive coding (emergent codes from data)	Conversation analysis: (1) ESRU cycle coding, (2) inductive coding

specific moments where such interdependence played a role in this study. For example, potential concerns about summative assessment frequently influenced the way the teachers enacted lessons and used feedback, and this, in turn, impacted decisions on marking, grading and fairness. These dynamics will be addressed while I answer each research question, and will also be revisited in the Discussion chapter. Whereas RQ3 is focused on analysing one teacher's instructional practice at a specific time in the first year of research to gain an in-depth understanding, RQ1 and RQ2 draw on the entire data set collected from three teachers and across two years to address the question. Table 2.7 outlines the focus of each research question, corresponding data sources and analytical methods, which will be further articulated in Chapter 4.

Chapter 3 Research context

3.1 Introduction

Assessment, as for most other activities and processes in education, does not exist in isolation from the society and culture in which it occurs. This position stands in contrast with the behaviourist approach, where testing discrete skills and knowledge removed from the context makes sense (Lund, 2008). From sociocultural perspectives, however, assessment is considered to be a cultural activity whereby individuals negotiate meaning and identity through participation in the activity with the cultural, social and historical backgrounds they are situated within (Elwood 2006; Pryor & Crossouard 2008), rather than an isolated information-processing mechanism by individual assessors. Considering the sociocultural nature of 'learning', neither formative nor summative assessment cannot be fully understood without considering their social dimensions (Shepard, 2000). In this sense, formative assessment is a socially and culturally situated activity (Wenger, 1998), whereby the teacher and students co-construct meanings through participation in the community. Similarly, summative assessment is also bound and shaped by culture. In assessment theories informed by sociocultural theory, the meaning of test scores is constructed collaboratively by testers and test-takers rather than 'objectively' defined by external authorities (Bonner, 2013).

This perspective has implications for the design and interpretation of assessments. Consider the validity of collected information: whether an assessment is valid depends on the social situation where it is undertaken and reflects the interests of different stakeholders such as teachers, students, parents, school administrators and policymakers (Bonner, 2013). With these considerations in mind, this chapter provides some cultural, historical and political background to Korean education and the curriculum in which this study was situated. The school that I visited for this thesis was, on several fronts, a microcosm of Korean high school education, where many teacher beliefs and actions (as well as those of students) are influenced by assessment-related concerns (Kwon, Lee & Shin, 2017; Sorenson, 1994). After considering the social context and assessment policy, I will turn to the curriculum context where efforts are being made to include more NOS-related content in the national science curriculum.

3.2 The ‘SKY Castle’: the context of accountability in Korea

The highly successful 2018–19 Korean television drama *SKY Castle* was a story of four elite Korean families living in a wealthy neighbourhood, portraying the lives of children who were in the same high school and whose parents strived to send their child to the most prestigious universities—from which the title ‘SKY’ (i.e., Seoul National, Korea and Yonsei Universities) came from. It described how the characters’ greed and obsession could go as far as destroying other people’s lives to ensure their children’s success, which starts from entering those universities (Yu, Jo & Kim, 2019). Once aired, the drama drew considerable attention and sparked social controversies. It revealed the dark and furtive sides of Korean education and gained much popularity in China where a similar culture of extreme competition exists (Shin, 2019). *SKY Castle* is fiction but is not merely fictional. It vividly pictured the uncomfortable reality of meritocracy and the ‘education fever’ underlying the high educational achievement in Korea (Im, 2019; Sorenson, 1994). As Martin et al. observe, ‘a strong sense of meritocracy continues to permeate Korean society with the belief that if a student, male or female, rich or poor, performs well on the CSAT [College Scholastic Ability Test], he or she will have the opportunity to succeed at the top national universities’. (p. 244)

The system of assessment is closely linked with such competitive culture of education in Korea. It has been argued that teachers’ ideas about and practices of classroom assessment inevitably relies on the culture and system in which the school system operates (Black, 2013), ranging from practical issues such as the school, district and national-level grading policies to more societal and cultural issues such as university admission criteria and the impact of higher education on one’s socioeconomic position. It is therefore necessary to consider the Korean context regarding the culture and system of accountability for a fuller understanding of science assessment in classrooms. In a recent review of educational assessment in Korea, Kwon, Lee and Shin (2017) described Korean education as based on the tradition of ‘testocracy’, which historically originated from the Confucian philosophy that has prevailed in East Asia since the Middle Ages. At least since the end of the 1960s, external high-stake examinations administered by the government have mostly determined whether a student could enter prestigious universities or not, which subsequently decided their job opportunities and income level in later life. Such heavy

	Age	Year	School	
Early Childhood Education	3-6		Kindergarten (3 Years)	
Primary Education	6-7	Year 1	Elementary School (6 Years)	
	7-8	Year 2		
	8-9	Year 3		
	9-10	Year 4		
	10-11	Year 5		
	11-12	Year 6		
Secondary Education	12-13	Year 7	Middle School (3 Years)	
	13-14	Year 8		
	14-15	Year 9		
	15-16	Year 10	General High School (3 Years)	Vocational High School (3 Years)
	16-17	Year 11		
	17-18	Year 12		
			College Scholastic Ability Test (CSAT)	
			Comprehensive School Record	
Higher Education	18-19	Year 13	University (4 Years) or College (2-3 Years)	
	19-20	Year 14		
	20-21	Year 15		
	21-22	Year 16		

Figure 3.1 Education system in Korea (based on KARIC, 2020; MOE, 2021)

reliance on the one-time, high-stake examination (e.g., CSAT) has produced serious harms in the Korean society, most notably the rise of cram schools offering ‘shadow’ education for ‘efficiently’ preparing students for the external exams (Byun, 2014); Classes in (regular) schools, considered less competent in test preparation, were neglected by students and parents, making teachers lose passion in their work.

In Korea, each student’s school-leaving grades (called ‘comprehensive school record’) entirely rely on the assessments planned and marked by teachers across all school subjects. When a student applies to universities, these school grades are complemented by national college entrance exam (CSAT) scores along with additional university-based tests such as interviews. Broadly speaking, the pathways to universities are divided into *early admission*[3] and *regular admission*. Although details vary by years and also by individual universities, early decisions admission tend to rely much on the comprehensive school records (i.e., school-based assessments), whereas the regular admissions consider CSAT scores as a primary factor. What has been changing drastically during the past 20 years is the proportion of early decision and regular admissions, which has been almost inverted during the period (Figure 3.1 and Figure 3.2). Such a shift largely stems from the government's effort to reform the system of accountability and university entrance from reliance on the CSAT to putting more value on school-based assessments by teachers. As a natural conclusion from classroom assessment now affecting students' entrance to higher education to a significantly greater extent than 20 years ago, there have been increasing expectations for teachers regarding selecting assessment methods and ensuring validity and fairness in their classroom assessment.

Such shift in the university entrance system soon met with resistance from students and parents, due to its misalignment with the traditional conception of fairness and accountability in Korean society. The deep-rooted idea that large-scale tests are the fairest and most objective, reliable method of assessment conflicted with the increasing emphasis on school assessment. As a response, in 2018, the newly elected Moon Jae-In administration launched the *Committee for Public Deliberation on the Revision of College Entrance System*, under the leadership of former Supreme Court Justice and then law professor Kim Young-ran. President Moon remarked that the new college admission system should be 'fair, viewed from the direct stakeholders such as students and parents, and simple, for every person to readily prepare for' (CPDRCES, 2018, p. 37). In their final report, it was revealed that there is little agreement among the public regarding the relative proportion of CSAT scores and comprehensive school records in making university admission decisions (CPDRCES, 2018). As of 2021, comprehensive school records remain to be a major determinant of university entrance.

---

[3] Early admission is close to the 'early decision' (rather than 'early action') in US college admission system, in that once offered admission under an early admission programme, the applicant is bound to that university and cannot withdraw to apply for regular admission to other universities.

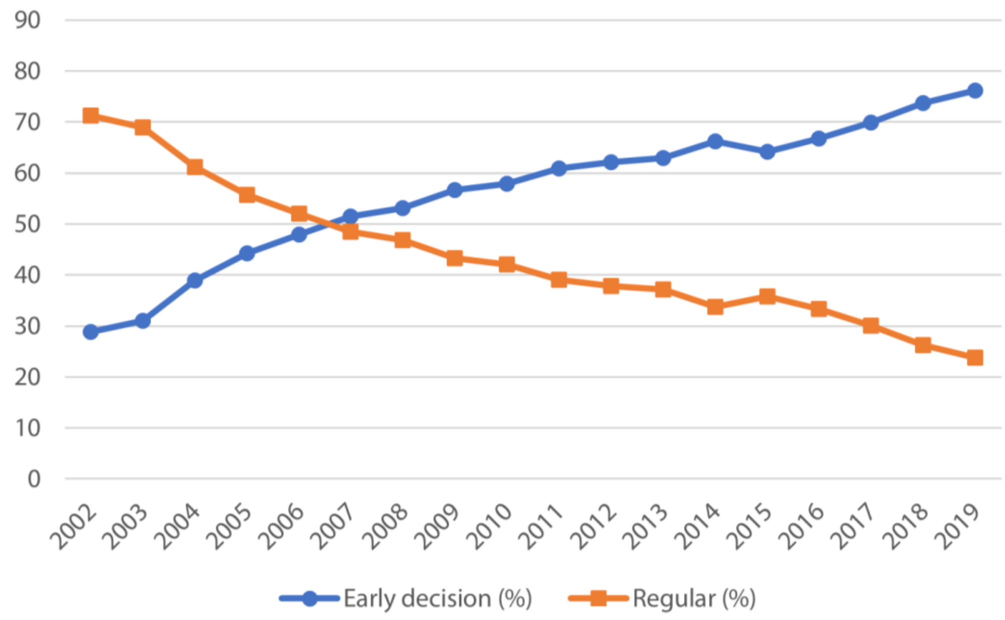
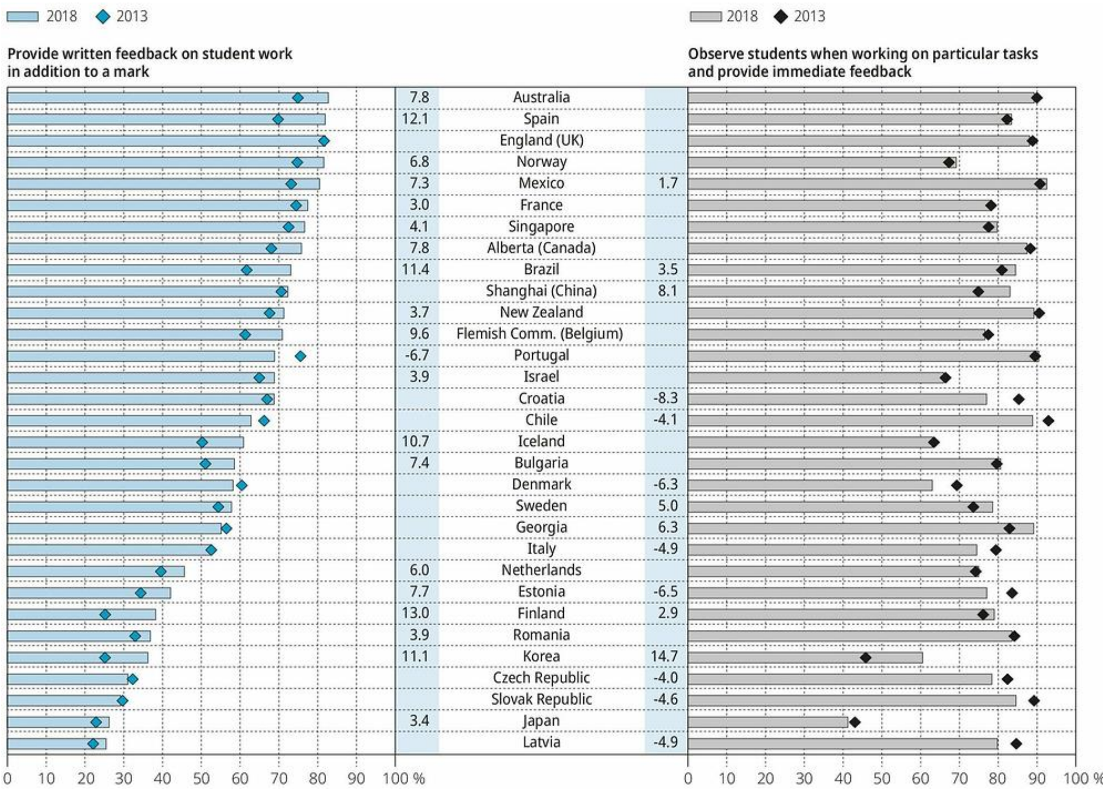


Figure 3.2 The proportion of early action and regular admissions in Korea, 2002–2019 (from Park, 2018)

Concurrent with the major reforms in the overall university admission system, the government has also been working to improve the quality of teachers' classroom assessment so that it can be linked more closely to students' learning. This, in turn, came hand in hand with the international emphasis on assessment for learning (AfL), where 'the term "assessment" refers to all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged' (Black & Wiliam, 1998, p. 8). In 2018, the Department of Education of Korea called for a paradigm shift in classroom assessment from *product-centred* assessment to *process-centred* assessment. Process-centred assessment is defined as 'an assessment that provides proper feedback that supports students' growth and development, by collecting various data about students' characteristics and change visible in the process of teaching and learning, according to the assessment plan where assessment is coupled with instruction based on the achievement standards in the national curriculum' (Korea Institute for Curriculum and Evaluation, 2019, p. 53), which coheres with the global shift from 'assessment of learning' to 'assessment for learning' and 'assessment as learning'. However, currently, there are conflicting expectations as to whether or not

Figure 3.3 Percentage of lower secondary teachers who report that they ‘frequently’ or ‘always’ use the following methods of assessing student learning in their class (OECD, 2019, p. 61)



this new policy will successfully reach the classrooms to initiate changes in teachers’ classroom assessment practices (Ban, Kim, Park & Kim, 2018).

3.3 Korean science teachers’ assessment practice

Compared to other developed nations, Korean teachers have been reported not to be very active in providing formative feedback in the classroom. According to the latest *Teaching and Learning International Survey* by the OECD (OECD, 2019), the proportion of Korean teachers who routinely assess students’ progress by observing them and providing instant feedback (60%) was lower than the OECD average (79%). The use of self-assessment for students’ evaluation of their own progress (36%) was also below average (41%). Figure 3.3 shows that the use of classroom feedback by Korean teachers is less frequent than in other countries. It is noteworthy, however, that both written feedback and immediate feedback in Korean classrooms have become more frequent in 2018 than in 2013, an increase rarely observed in other countries.

Turning to science teachers specifically, Noh, Lee, Kang and Kang's (2015) study of Korean science teachers' assessment beliefs and practices provides some further nuances. These researchers found that while Korean teachers use diverse strategies for assessing science learning in practice, their perceived purpose of assessment was mostly centred on the traditional purpose of summative assessment, i.e., evaluating students' learning. In addition, they found a discord between the teachers' *preferred* and *used* assessment strategies (Noh et al., 2015). They asked the teachers their own purposes and strategies for classroom assessment and compared this to their responses to three prototypical situations. Summaries of their findings in Table 3.1 and Table 3.2 indicate several important discrepancies. In terms of assessment purposes, teachers used assessment for a summative purpose (i.e., evaluate student learning) and other non-instructional purposes such as 'selecting' students to represent their school in science competitions or receive school awards. While teachers preferred to use formative assessments for providing learning opportunities and feedback, this function of assessment was less addressed in relation to their actual practice. Such discrepancy between teachers' beliefs about effective summative assessment and their actual classroom practices resonates with similar findings from other countries (Black et al., 2010; Rieg, 2007). Also noteworthy is that few teachers associated their assessment practice with preparation for high-stakes external examinations (Table 3.1), suggesting that the impact of external texts on classroom assessment is minimal despite the increased emphasis on the latter compared to the former in the university entrance process.

3.4 The introduction of 'scientific inquiry and experimentation' in the new curriculum

The 2015 National Curriculum of Korea, effective since March 2018, is organised around the 'core competencies' that students are expected to develop through learning subjects (MOE, 2015). In relation to the *science* curriculum specifically, there are five core competencies essential for learners to attain scientific literacy: scientific thinking,

Table 3.1 Korean science teachers' orientations to the purpose of classroom assessment and the actual purposes in their practice (from Noh et al., 2015, p. 729)

Assessment purposes		Orientation	Practice
Summative	Evaluate student learning	55%	53%
	Select students as representatives	30%	0%
	Identify students who need additional support	1%	0%
	Measure student contribution during classroom activities	1%	15%
	Reward students for their academic efforts	1%	1%
Formative	Provide students with opportunities for learning	9%	25%
	Provide students with descriptive feedback	0%	5%
Others	Familiarise students with the format and the content of external tests	2%	0%

Table 3.2 Korean science teachers' preferred versus used assessment strategies (from Noh et al., 2015, p. 730)

Assessment strategies	Used	Preferred
Multiple-choice	51%	1%
Open-ended	4%	1%
Experiment	15%	7%
Report	9%	31%
Essay	8%	0%
Presentation	5%	10%
Portfolio	2%	6%
Debate	1%	5%
Observation	2%	24%
Peer assessment	2%	11%
Interview	1%	2%
Self assessment	0%	2%

scientific inquiry ability, scientific problem-solving ability, scientific communication ability, and scientific participation and lifelong learning ability (MOE, 2015, p. 4). Based on these key competencies, the new curriculum attempts to incorporate new knowledge that is becoming increasingly important for scientific literacy, while trying to emphasise curricular content other than traditional scientific knowledge, such as, NOS as an important element of the science curriculum (MOE, 2015).

One notable initiative in the new curriculum is the introduction of a new subject SIE in an attempt to facilitate the students' understanding of NOS through inquiring how scientific knowledge is produced and applied (MOE, 2015). SIE was introduced as a compulsory subject for Grade 10 students in order to help them 'recognise the value of science as well as its impact on scientific inquiry and social, technological development' by providing opportunities to experience authentic science inquiry activities (MOE, 2015, p. 25). It deserves attention as the first curricular initiative in Korea—also the first around the globe—to teach NOS and scientific inquiry as a separate compulsory school subject. Research has provided evidence that Korean science teachers have high expectations about this new curricular initiative, along with the concerns about the radical shift in terms of assessment and teacher training (Son, 2016; Yoon & Kang, 2016). A recent study of seven SIE textbooks (Yang, Park, & Song, 2019) showed that textbooks address diverse epistemic and social NOS aspects, and more importantly, often in explicit manners with reflective prompts. This is contrary to most other analysis of NOS aspects in the written science curricula, which repeatedly reported naïve and implicit expressions of NOS with no opportunities to reflect on them (Abd-El-Khalick et al., 2017; BouJaoude et al., 2017; Leite, 2002; Li et al., 2018; McDonald, 2017; Wilkinson, 1999).

The uniqueness of SIE as a science curriculum reform stems from several aspects. First, in its process of development, the NOS-related learning expectations were selected and organised first, and then came the selection of relevant historical episodes, real-life applications and practical activities (MOE, 2015). For example, the curriculum developers elected 'understanding how science develops' as a learning expectation and then chose, among other examples, Galileo's thought experiment about free fall as a historical instance that can reveal it. Secondly, as a result, the 15 learning objectives in the SIE curriculum are dedicated to NOS, but not any scientific concepts (MOE, 2015). These characteristics support the view that NOS is addressed in a very explicit manner throughout the SIE, instead of as a high-level curriculum goal. Given that other recent curriculum reforms, for example, the *Next Generation Science Standards* (NGSS Lead States, 2013) in the US, are still criticised for the absence of explicit instructional aims for NOS (J. S. Lederman et al., 2019; Olson, 2018), SIE can certainly be considered a radical case with the potential to offer important information regarding teachers' responses to such curricular reforms. Below are the learning expectations of Unit 1, 'Scientific inquiry in history':

- Replicate the crucial experiments that motivated paradigm shifts in the history of science and explain how science develops.
- Conduct historical experiments that led to serendipitous discoveries and explain the NOS aspect found in them.
- Conduct an inquiry by direct observation and explain the inductive inquiry method.
- Conduct representative historical experiments that include making hypotheses and explain the features of the deductive inquiry method. (MOE, 2015, p. 114)

As seen in these learning goals, SIE highlights contextualising NOS within various settings (Allchin et al., 2013), and more importantly, foregrounds aspects of NOS such as how science develops, how discoveries are made in science and how deductive and inductive methods are used in science. Given the significance and novelty of SIE as a curricular initiative in teaching and assessing NOS and its explicit focus on NOS, this new curriculum can be considered an ideal context in which this thesis's research questions can be investigated.

3.5 The inclusion of NOS in science teacher education programmes in Korea

Most secondary science teachers in Korea are trained in four-year undergraduate programmes in Colleges of Education, in one of the four constituent disciplines of science: physics, chemistry, life sciences and earth science. During the four years, undergraduate preservice teachers take compulsory and elective courses that encompass the subject matter knowledge (e.g., classical mechanics, electrical chemistry, physical oceanography), educational theory (e.g., philosophy of education, curriculum theory, educational assessment, counselling), and curriculum and pedagogy (e.g., physics textbook development and evaluation, life sciences curriculum), and includes teaching practices in schools. To be employed as a science teacher at public

Table 3.3 Provision of courses related to history and philosophy of science (HPS) and NOS in preservice teacher education programmes in Korea (from Song & Joung, 2014, p. 2205)

University	HPS/NOS-related courses (credit)
A	Historical development of physics concepts (3)
	History of science for teachers (2)
	Philosophy of science for teachers (2)
B	History and philosophy of physics (3)
	Philosophy of science (3)
C	History of earth science and inquiry methods (3)
D	History of physics (3)
E	History of science (3)
F	History of the Earth (2)
G	History and philosophy of science (3)
	Education of earth history (3)
	History of chemistry (1)
H	History of chemistry and chemistry education (2)
	History of earth science and earth science education (3)
	Philosophy of science and science education (3)
I	History of physics and philosophy of science (3)
	History of biology and biology education (3)
J	History of physics (3)

secondary schools, the graduates need to pass the highly competitive teacher employment examination (Im, Yoon & Cha, 2016).

Song and Joung (2014) analysed the inclusion of history and philosophy of science and NOS-related courses in the curriculum of ten Korean universities' science teacher preparation programmes. They found that all ten programmes offered courses titled 'History of Science for Teachers', 'Philosophy of Science for Teachers', 'History of Chemistry and Chemistry Education' and the like, but half of them only provided only one course of two or three credits (Table 3.3). In addition, they also reported that NOS is rarely included in the in-service teacher training programmes offered by District Offices of Education (Song & Joung, 2014). Philosophy of science and NOS also appear in the Korean teacher employment examinations every year, and candidates reportedly spend a significant amount of time studying these topics (Kim, Cha, Kim &

Kim, 2010). Lee et al.'s (2013) analysis of the teacher employment examinations revealed that about 16 per cent of the science education theory questions were related to philosophy of science and NOS. These included questions asking the inductive, deductive and abductive methods in science, theories of scientific development proposed by Kuhn, Popper and Lakatos, and the theory-ladenness of observation (Lee et al., 2013).

Although Song and Joung interpreted these as a lack of teacher training in NOS, from a global point of view, at least the provision of any courses on NOS can be considered as encouraging, given that it is rarely the case in other countries (Metz, 2014; Wong, Wan & Cheng, 2014). Matthews (2014) lamented that despite the concerns and arguments around teachers' proper knowledge and skills in NOS, it had been rarely included in teacher education programmes in Australia, the USA and other countries. Instead, he criticised that teachers' knowledge of NOS and the epistemology of science mostly comes from their own science coursework, which is insufficient given the importance of NOS knowledge for science teachers (Matthews, 2014).

3.6 Chapter summary

This chapter examined the contexts of culture, policy and curriculum underlying this study. To summarise, the study is situated within a deep-rooted culture of competition, testing, and an obsession with high school education as a gateway to one's future. The traditional focus on the external high-stakes examination CSAT has recently moved to a focus on school-based grades, followed by a gradual shift of focus to teacher assessments in the classroom. In terms of NOS education, Korea has recently made several curricular and policy initiatives that are related to NOS. The introduction of a new compulsory subject SIE has been motivated by the increasing call for teaching NOS and scientific inquiry in school science. The inclusion of NOS and HPS in many current teacher education programmes also suggests that NOS is gaining attention in science education in Korea. This background helps to contextualise and interpret teachers' views and actions as a sociocultural product that reflects various aspects of science education in Korea that range from the broader culture of education to the trends in teachers' classroom assessment practice.

Chapter 4 Methods

This thesis addresses the question of classroom assessment of NOS through investigating science teachers' views and practices. Accordingly, it is a study of classroom assessment and of NOS at the same time. In the classroom assessment literature, teachers' assessment views and practices are often studied by using surveys with relatively large sample sizes to understand the frequencies of using different assessment methods. Some studies have used a combination of surveys and interviews to investigate teachers' perceptions of assessment, and assessment literacy, assessment expertise or assessment identity (see Section 2.2.5) and there are pre-designed and validated scales for describing teacher assessment views and practices (Brown, 2006; DeLuca, LaPointe-McEwan & Luhanga., 2016; Looney et al., 2018). However, instead of using these scales, this study adopted an in-depth qualitative design because most existing scales are targeted at domain-general aspects of classroom assessment. For example, the often-cited questionnaire for teacher conceptions of assessment by Brown (2006) includes statements such as 'assessment is assigning a grade or level to student work', 'assessment determines if students meet qualifications standards' and 'assessment helps students improve their learning'. Since these context-general questions are not sufficient for understanding the specificities of NOS assessment in the classroom (see Section 0), it was deemed more appropriate to take an exploratory approach to the research questions to discover emergent themes. Moreover, given the lack of observational research on NOS (Kelly et al., 1998), the case study will make a valuable contribution to NOS research methods.

4.1 Case study methodology

4.1.1 Rationales for case study

The case study methodology is used for various purposes in educational research (Thomas, 2013). Yin (2017) defines a case study as follows: (a) it 'investigates a contemporary phenomenon (the 'case') in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident'; (b) 'copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result; (c) benefits from the prior development of theoretical propositions to guide design, data collection, and analysis, and as another result; and (d) relies on multiple sources of evidence, with data needing to converge in a triangulating fashion' (p. 31). In the context of the present study, the case study approach was deemed suitable, provided that the aim was to investigate (a) what was happening at the schools at this moment, where (b) teachers' practices were shaped in the complex interplay of policies, personal knowledge, and curriculum and school contexts (Cohen, Raudenbush & Ball, 2003). As discussed earlier, my case study of teachers' NOS assessment practices (c) can be collectively informed by existing theories in NOS and those in classroom assessment, and (d) was based on diverse formats of data sources emerging from the planning and implementation of NOS assessment.

There are several ways in which case study research could be classified. Acknowledging that the type of a case study relates to the nature of questions that it aims to answer, Yin (2017) suggested that a case study can be descriptive, exploratory and explanatory. (p. 8) The three types are not mutually exclusive and, in many cases, coexist in a single study. When it comes to this thesis study, although I have provided a detailed *description* of the events, people and their actions, doing so will eventually answer exploratory and explanatory questions, too. The *exploratory* aim of this case study is clear in that it looks at the aspects of NOS assessment in which little previous research is available, in particular, the third research question about the formative assessment of NOS. My primary aim was to identify *what* types of teacher feedback classroom interactions occurred in the NOS-oriented lesson, which has not been studied previously. From this 'what' question, the goal is to 'develop pertinent hypotheses and propositions for further inquiry' (Yin, 2017, p. 9). In contrast, the first and second research questions on *how* the teachers assessed NOS summatively and *why* they did so are aimed at an *explanatory* account of the case (Yin, 2017). It goes beyond mere description or exploration of what happened but seeks why it happened that way, by investigating the various levels of contexts and the teachers' decision-making process about classroom assessment.

4.1.2 Single-case study design

A key task in case study research is to define and bound the 'case' in question (Yin, 2017). In this study, the phenomenon under study is the classroom assessment of NOS, and the case in focus is the event that the three teachers engaged in teaching and assessing NOS over the two observed years, more specifically two semesters (Spring 2019 and Spring 2020). This case definition concurrently sets the boundaries of my case. For example, I do not look at the other teachers in the same school or my participants' teaching and assessment activities for other school subjects. Likewise, I do not discuss the periods that were outside my fieldwork. However, these are not rigid boundaries, since they still matter as contexts that shape, influence and justify what was happening within the case, particularly from a sociocultural perspective on classroom assessment (see Section 3.2).

In this sense, this study employs a single-case design, which allows the readers 'to experience vicariously unique situations and unique individuals' (Donmoyer, 1990, p. 193). The decision to focus on one case rather than multiple cases can be justified for several reasons. First, the school could be seen as an *unusual* and *revelatory* case (Yin, 2017) in that NOS is not often formally assessed in classrooms (Abd-El-Khalick et al., 1998a). On the contrary, the three teachers in the school interpreted and implemented the curriculum in a way that NOS is taught and assessed as a learning outcome, which is a condition that was not previously accessible to science education researchers. This is even more so when one considers that most existing studies on classroom assessment of NOS are mostly situated in university-based professional development programmes, whereas there was no intervention in this thesis study (see Table 2.4).

4.1.3 The problem of generalisation

Despite its frequent use in educational research, the case study method is often criticised for the lack of generalisability, because it typically concerns a specific case of a small number (Yin, 2010). Recognising that all situations are unique in some respects, Pring (2000) argued that not every set of findings from a single situation can or should be generalisable in educational research. Rather, it is the concepts that emerge from the case that can inform other situations (Pring, 2000). Hence, it is worth considering to what extent and how this study can be meaningful as an educational research project. Qualitative methodologists have proposed the concept of 'analytic generalisation' and 'transferability', as opposed to the traditional 'statistical generalisation', to be the objective of case studies (Polit & Beck, 2010; Yin, 2017). In other words, qualitative research aims to generalise from particulars to broader constructs and/or theory, as well as translate the findings from a specific case to a different group of people or setting (Polit & Beck, 2010). Let us return to the three research questions: *First, how do teachers view and practise the summative assessment of NOS in the classroom? Second, how does a team of three science teachers engage in NOS summative assessment in a school? And what does their decision-making process reveal about the potential issues in the summative assessment of NOS? Third, what are the instructional strategies that allow for teachers' formative assessment NOS? Also, what are some potential issues and challenges involved in the formative assessment of NOS?*

For these three questions, the uniqueness of the case restricts making statistical generalisations in the sense that findings from this particular case will be replicable in other circumstances, or that the themes and issues identified from this case are exhaustive. Instead, the focus of the case study is on revealing the key issues of the classroom assessment of NOS through an in-depth investigation, which are situated within the case, but still can connect to theoretical issues in NOS research as well as provide insights for other similar phenomena. In this sense, although the context of this study is unique and thus is not representative of its kind, the findings can contribute to the improvement of science educators' views of NOS teaching and assessment, teacher assessment practice, and the intersections of these. It can provide answers to existing questions in these areas and raise new issues that have not been recognised by researchers. The actual generalised implications of the study's findings in terms of NOS and classroom assessment research will be presented in Chapter 8.

4.2 Reflections from the pilot study

It is often recommended for a case study researcher to conduct a pilot case study (Yin, 2017). It is a process of research that assists the researcher in organising research questions and design. My pilot study was conducted in March and April 2019. Its main aims were to get a sense of how the recent NOS-oriented science curriculum in Korea was being taught and assessed in the classroom, and to develop questions, ideas and methods for the main study. As with the main study, the pilot study had been approved by the Departmental Research Ethics Committee (DREC) prior to data collection. I did my pilot data collection at the school that later became the single case for this thesis. Accordingly, what I had collected as pilot data not only helped refine the research design and subsequent data collection but also constituted part of the ultimate dataset to answer the research questions. The pilot data (i.e., the first-year data) included four lesson videos and two interviews from two teachers, Dr Kim and Ms Nam, who were then teaching SIE to Grade 10 students at the school. Both teachers were teaching SIE for the first time that year.

The scope of the pilot inquiry is usually broader than the ultimate study since it aims to provide general insights into what is at issue and how it could be approached methodologically (Yin, 2017). During the pilot study, my attention was on teachers' interpretation and implementation of the NOS-centred curriculum, rather than specifically on their classroom assessment practices. Therefore, the interview questions were often broad about their knowledge and practice on the teaching of NOS (e.g., *'Tell me about your impressions when you first received the SIE curriculum'*, *'At this time, how would you define the nature of science in the context of SIE curriculum?'*, *'How do you assess SIE in your school? Do you think it is an appropriate way to assess it?'*) but were sometimes inspired by the lesson observations (e.g., *'How did you feel about this student group's presentation?'*, *'Why did you include this activity in your lesson?'*). Each interview was done in the school face-to-face and lasted between 40 and 60 minutes.

Thematic coding (Braun & Clarke, 2006, 2012) of the data, mainly interviews, led to four salient themes being identified about their ideas, expectations and practices about teaching NOS as part of the new curriculum. First, although both teachers considered NOS important to be taught to students, they were unsure about whether it should be taught as a separate subject and whether the curriculum would be successful. Second, teachers felt that teaching NOS is different from teaching science. NOS lessons, they said, use a lot of student-led activities and need more discussion and reflection, due to the controversial nature of the knowledge in question. Third, the teachers repeatedly mentioned that there were little support or guidance for teachers with regard to teaching the NOS-centred curriculum, which was a big change for them to deal with. Fourth, assessment and grading were key concerns that impact the teachers' planning and implementation of the NOS-centred curriculum. Also, since there was no clear guideline regarding the assessment of the NOS-centred science curriculum, teachers have much freedom in deciding what and how to assess and also have challenges stemming from the lack of information.

Although these themes will be revisited and elaborated throughout the findings chapters, it would be useful here to consider how these observations influenced the later stages of the thesis project. Most importantly, the classroom assessment of NOS was selected as the main focus of the thesis, on the basis of its perceived importance by teachers and the lack of research on the topic. Accordingly, research questions were organised around the assessment of NOS, particularly teacher' views and practices of formative and summative assessments. This was also considering that assessment can be a key element that mediates teachers' curriculum enactment (Black et al., 2010, 2011; Dolin et al., 2018). Methodologically, the pilot study allowed identifying several potentially important themes regarding NOS teaching, which were reflected in the interview protocols. Interview questions specifically aimed at the perceived difference between teaching (assessing) science and NOS, for example, were included to investigate the preliminary theme further. In addition, the experience of conducting the pilot study gave the impression that richer and more authentic information about teachers' practice can be obtained when an interview is done after a lesson observation. For this reason, in the subsequent data collection, I was particularly mindful of elaborating the observations from lessons through follow-up interviews, so that they can collectively lead to a deeper understanding of teachers' intentions, practices and reflections.

4.3 Data collection

4.3.1 Overview of data collection procedures and the impact of COVID-19

The fieldwork was carried out in April 2019 and between March and July 2020. The data from the first year was used to inform the design and planning of the second-year data collection. Although it was initially collected for such pilot purposes, this dataset was later also analysed together with the second-year dataset to develop the main findings of the thesis. Specifically, Chapter 5 and Chapter 6 are based on the entire two years' data, whereas Chapter 7 focuses on one teacher's lesson and interview data from the first year. Over the two years, I made five visits to the school (two in the first year, three in the second year). It should be noted that the second-year data collection was substantially influenced by an unexpected outbreak of the coronavirus disease-19 (COVID-19) pandemic. Here I briefly summarise how COVID-19 influenced my data collection procedure.

In the second year, data collection was originally planned to start in February 2020. When the initial interviews were set up and about to begin, COVID-19, which first broke out in Wuhan, China, landed in Korea and cases of infection increased by hundreds each day. Korea was the second most affected country after China in the early phase of the pandemic, causing the school openings in March postponed four times as the situation developed. It was delayed for a week (first delay), two weeks (second delay), another two weeks (third delay) and yet another week (fourth and final delay), resulting in an online reopening in early April, followed by a 'hybrid' reopening (see below) from early June. The first visit to the school was made towards the end of March, while the teachers were still waiting for students to return to school.

During March and April, I was in close contact with my main contact person, Dr Kim (pseudonym), and was informed of the updates on the school circumstance and contingency plans being made. According to the final semester plan which came as late as early April, students would come to school every other week starting from early June and attend online classes in the remaining weeks. For online class weeks, the school decided to offer asynchronous teaching where pre-recorded lesson videos (15–20 mins each week) and assignments to be completed by each student and uploaded on an online system. Meanwhile, the Department of Education and Seoul District Office of Education discouraged schools from making assessments in online environments until they have more certainty over the COVID situation. One consequence of this policy was that, for all subjects including SIE, teachers spent most of the face-to-face class time for assessment tasks, and the online class time for teaching the background knowledge necessary for doing the tasks.

In terms of the current study, these changes had a substantial impact on the contexts in which teachers perform NOS assessment. On the one side, such changes seemed to pose a serious challenge to understanding the continuity and longitudinal development of teachers' practices over the two years. However, on the positive side, the changes meant that I could understand the focal phenomenon (i.e., teachers' assessment of NOS and the factors shaping it) in a more diversified setting than before. For example, the group tasks that had accounted for most of the classroom assessment in 2019 were changed to individual tasks in compliance with social distancing, allowing me to consider both individual and group-based assessment of NOS. Also, since more than half of the face-to-face teaching was replaced by online video lessons, significant changes in the relationship between instruction and assessment practices were expected. This new setting, when combined with the relatively normal setting in 2019, enabled a much richer understanding of teachers' thoughts and practices. Key issues in classroom assessment such as the aims of assessing NOS, assessment criteria, ensuring fairness—which later formed the main themes in the study findings—could be illustrated more vividly as the teachers negotiated with the new school environment and the changing assessment policy under the pandemic situation.

4.3.2 The school

The school was located in an urban and wealthy region in Seoul, Korea, where a culture of competition was prevalent among many students to get into prestigious universities. It belonged to the so-called Gangnam School District 8, a group of 26 high schools where many students come from high-income and educated family backgrounds (Park, n.d.). Since the students' school-leaving GPA counts significantly towards the university entrance, teachers, parents and students were sensitive about how each subject is assessed in the school. The average income in the neighbourhood was top-five among the 226 local government units in the country. The school was surrounded by the Supreme, Appeal and District Courts, numerous law firm offices, major hospitals and clinics, which resulted in a significant portion of the students being from a high socioeconomic background and having highly educated parents. These contextual factors made this school a classic example of the competitive school culture in Korea—not in the sense that it is an average school, but rather that some aspects of the broader national culture were particularly visible in this school.

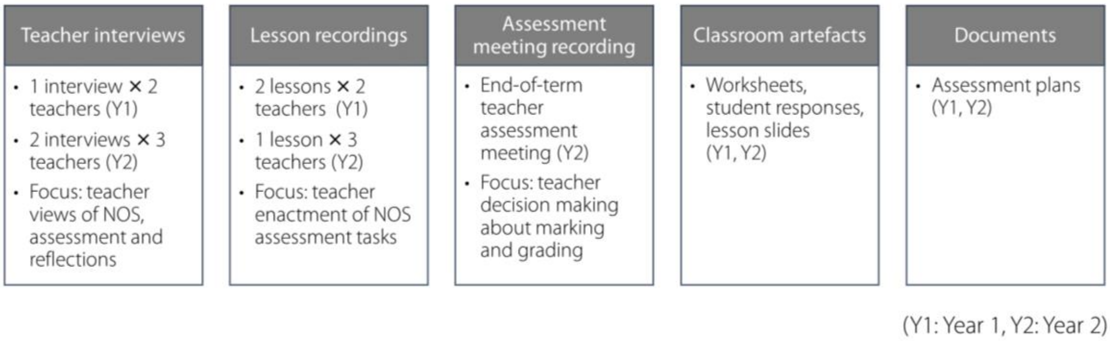
As per the national curriculum, SIE was taught to the entire Grade 10 students (age 15–17) in the school. In the first year, 11 classes (~290 students) were taught by Dr Kim (four classes) and Ms Nam (seven classes); in the second year, Ms You joined and the eleven classes were taught by Dr Kim (four classes), Ms Nam (two classes) and Ms You (five classes). Students attended the SIE class one hour per week during their first two semesters in high school and were graded twice, each at the end of the semester. This grade was supposed to contribute to each student's comprehensive school record that has a strong impact on their university admission, which would have made the school-based assessment of SIE high-stakes under normal circumstances. However, at the time of the fieldwork, the Ministry of Education had enacted a new guideline that required SIE to be graded on the scale of A, B and C rather than the usual Stanine system, which lowered the stakes associated with SIE[4]. The specific assessment methods to be used for grading were supposed to be decided by schools. In the case of the observed school, the teachers in the science department made a decision at the beginning of the semester not to do end-of-semester written exams for SIE but to grade with several group projects and individual papers. At the beginning of the semester, they met to discuss what activities to mark and consider for the end-of-semester grades, and decide on preliminary marking schemes. Each teacher was responsible for marking and grading her students, and the other teachers were not directly involved in this process. In the following, I describe in detail each type of data with a focus on my second-year data collection, given that the first-year data have already been outlined in Section 4.2.

4.3.3 Teacher interviews

Figure 4.1 summarises the types of data collected for this study. Interviews are a frequently used method for learning about people's lives, feelings and experiences in qualitative inquiry. The ability of humans to 'symbolise their experience through language' is essential to their existence (Seidman, 2006, p. 8). In qualitative research, interviewing provides a useful way of generating data about participants' lives and also about the social world they live in (Holstein & Gubrium, 2003). When interview

[4] The backgrounds to this shift and its impact will be discussed in Section 5.1.

Figure 4.1 Summary of collected data



questions are properly prepared and organised, it allows the researcher to focus directly on the topic of the case study, and is useful to answer the ‘hows’ and ‘whys’ of key events in a case study (Yin, 2017). Given that the focus of this study was on understanding the lived experiences of teachers in a given circumstance, interviews were used as a key source of understanding the teachers’ knowledge and practices associated with NOS classroom assessment.

During the second year of the study, two formal interviews (~30 minutes each) were conducted with each teacher based on the protocol in Appendix B, which was generated in advance to cover a range of ideas, practices and experiences of the teachers. Interviews conducted at the beginning of the semester were focused on the teachers’ usual beliefs and practices about NOS and assessment, and their expectations about teaching NOS. As the school semester proceeded and the teachers engaged in teaching and assessing NOS, interview questions became more specific and so did their answers. Case study interviews can sometimes be around 90 minutes (Schuman, 1982; Seidman, 2006), while some qualitative researchers prefer to have shorter interviews more frequently (Yin, 2017). Given that this study did not solely rely on the interview but also had other data sources, I used both long and short interviews. Formal interviews took place each year at the beginning and the end of the semester. These interviews generally followed a structured set of questions common to all three teachers, but I also used several impromptu and personalised questions, too. Shorter, 10–15 mins-long interviews that followed lesson observations were aimed at corroborating what I had noticed in the teachers’ practices and asking what I had wondered about the teacher’s or students’ behaviours. They were also used to understand each teacher’s own rendition of the key events (i.e., a lesson, an assessment meeting, etc.) that were useful for triangulation.

The fact that the interviews took place across a period of time had some implications for the study. For example, the multiple interviews allowed building relationships with the participants that generated trust and rapport (Grinyer & Thomas, 2012), as I met with them multiple times across years. This, together with my gradually increasing understanding of the school and classroom culture, and teachers' perspectives, enabled me to cultivate an insider's view of their practices (see Section 4.4). Most formal and informal interviews took place in the school during the teachers' spare time, but in the second year, two interviews (one with Dr Kim, the other with Ms Nam) were conducted over the telephone. This was to minimise the time I spent at the school in consideration of the COVID-related concerns.

4.3.4 Observations

4.3.4.1 Lessons

Patton (2015) suggested that the power of observational fieldwork lies in the fact that it provides the researcher 'the opportunity to see things that may routinely escape awareness among the people in the setting' (p. 332). Since most social systems involve routines, Patton explains, an observer can often recognise aspects of the system that are taken for granted by the participants and thus unnoticed (Patton, 2015). Observations also can reveal things that people are not willing to talk about during interviews (Patton, 2015). For these reasons, observations can be a valuable source of discovering things that are not apparent from interviewing. In this study, I used lesson observation mainly to analyse teachers' formative assessment practice (Chapter 7) and to understand the instructional context from which the teachers' summative assessment concerns arose (Chapter 5 and Chapter 6).

Before the semester started, teachers shared the course syllabus and assessment plans with me so that we could discuss which lessons to observe. The selection criteria were the expected level of classroom interactions in the lesson, teachers' convenience, and the researcher's schedule. In the second year of data collection, the COVID-19 related issues were also an important variable to consider (see Section 4.3.1). I attended the agreed lessons, mounted a video camera at the back of the classroom to capture the teacher's moves and student responses. After the first lesson observation, I sensed that students were not used to being videotaped especially when they were presenting something before classmates, so I switched to audio recording to minimise the observer effect. To record the aspects that could be fully captured in audio, I took detailed field notes describing my impressions about the lesson content and the behaviour of the teacher and students. A summary of observed lessons is provided in Table 4.1.

4.3.4.2 Teacher assessment meeting

In the second year, after all the assessment tasks for the semester had been completed, the three teachers met on Week 15 to discuss assessment criteria and moderate marking. Each teacher had previously skimmed through the worksheets collected from the classes they teach, so the meeting was for sharing the issues that arose from the initial look and collectively resolving them. The meeting lasted for about 60 mins, during my third and last visit to the school. As was for lesson observations, my role in this meeting was a nonparticipating observer who was an 'outsider of the group under study, watching and taking field notes from a distance' (Creswell, 2016). I thus joined the meeting while remaining silent, audiotaped the discussion and took notes of what sounded like important statements in light of the research questions. In the observation note, I captured the overall flow of the discussion, main issues, decision-making process, and the divergence among the teachers' views on some issues. Although my focus was on 'descriptive' notes about the conversation in black, I also used a blue pen to add 'reflective' notes (Creswell, 2016) at some places to record hunches, learnings, and possible links to assessment theory and/or my previously collected data. This meeting and notes have proved valuable because a number of issues around the instruction-assessment link, fairness, discrimination and so on became explicit as the teachers were talking and making assessment decisions, which will be elaborated in Section 6.4.

4.3.4.3 Observer influence and observer bias

I was conscious that the presence of an observer could cause changes in participants'

Table 4.1 Summary of observed lessons

	Year 1		**Year 2**	
Lesson topic	Paradigm shift in history	Make your own periodic table	Cloud observation and classification	Inductive, deductive & abductive inquiry
Duration	50m × 2 sessions	50m × 2 sessions	50m	50m
Teacher	Dr Kim	Ms Nam	Dr Kim, Ms Nam, Ms You	Dr Kim, Ms Nam, Ms You
Lesson type	Student group presentation + Q&A session	Brief teacher introduction + student group work + presentation	Brief teacher introduction + student individual work	Brief instruction + paper-and-pencil individual activity
Lesson synopsis	This lesson came after the previous lesson where the teacher taught the concepts of paradigm and theory change in science, and provided historical examples. Students were then broken into groups of four to find an example of a paradigm shift in history and prepare a 10-min PowerPoint presentation. These two lessons are videos of students making presentations, followed by a Q&A session by teachers and peers.	The teacher introduced students to the accidental discoveries in science to highlight that seemingly serendipitous discoveries are in fact rooted in previous efforts and accumulated knowledge. Then she briefly outlined the history of the periodic table. Students worked in groups of 5–6 to make their own version of the table based on regularities. Students then presented their work, and the teacher pointed out the different classification criteria and methods used across the groups.	This lesson followed an introductory online lesson on the nature of observation and classification in science, including a brief account of the 10 kinds of clouds. Students went outside to sketch the clouds seen from the playground (or from inside the classroom), filled out a worksheet with questions on the properties of clouds and other historical examples that shows how observation leads to the discovery of a classification system.	Teacher gave a short instruction for the task, and then students individually wrote their answers in the worksheet for the rest of the lesson. The worksheet consisted of three parts, each starting with a number of historical episodes and then a series of questions related to inquiry and NOS.
Summative assessment	Group presentation	Group presentation	Individual worksheets	Individual worksheets
Data collected	Video recording, audio recording, student-made PowerPoint slides	Audio recording, student-made periodic tables	Audio recording, student worksheets	Student worksheets, teachers' self-descriptions of the lesson during the interview

behaviours (Rosen & Underwood, 2012). For example, it is likely that teachers change their instruction to be more interactional and dialogical when they know that someone (e.g., headteacher, inspector or researcher) is watching them. Likewise, students might be influenced by a stranger sitting at the back of the classroom. These potential changes in behaviours pose threats to the validity of research and thus should be reduced. I discussed this with the teachers as I planned the first set of lesson observations and was told that students would be okay with my presence because students were used to being observed since several student teachers were at the school for teaching practice at the moment. As for the teachers, I had done interviews with them before the observation to help them feel comfortable about having me.

Regarding the observer influence, there was one moment during Dr Kim's paradigm shift lesson in the first year where the teacher encouraged me to ask a question during the Q&A session after a group of students had presented (see Section 7.3.2). It should be noted that the question that I asked, although it was not a 'new' question to the class (Similar questions had been asked earlier in the lesson), actually triggered peer feedback from classmates and eventually the teacher acting upon it. Except for this one incident, I remained a non-participating observer in all my observations. Another potential pitfall in observations may occur when the observer's recording of events is biased with prior knowledge and hypotheses about participants' behaviour (Rosen & Underwood, 2012). To avoid such bias that can stem from selective recording, in my observations, I recorded the lessons and interviews in video/audio formats as well as in the field notes, which enabling reviewing the files later and comparing them with transcripts.

4.3.5 Assessment plans and materials

The brief assessment plans that the teachers created before the semester were collected. In the second year, the initial assessment plan underwent two revisions, which were also collected. I was also provided with part of the lesson and assessment materials designed by the teacher. This included their lesson PowerPoint slides and worksheets that were filled in by students and to be marked by the teachers. The format and content of the materials differed depending on assessment tasks. In the first year, I collected the

PowerPoint slides that students made for the paradigm shift presentation, and samples of the periodic tables made by students. In the second year, teachers' lesson slides used for the clouds activity, and students' completed worksheets from the VNOS activity (N=27), the cloud activity (N=28) and the scientific methods activity (N=56) were collected. These student activity sheets were not analysed directly but used as supplementary data to contextualise the teachers' perceptions and decision making (see Section 6.3.4, for example). Some examples of classroom materials are provided in Appendix C.

4.4 Reflexivity and researcher positionality

In qualitative research such as this thesis study, the researcher is not wholly detached from the process in which knowledge is produced (Yin, 2017). It is therefore important to accept that all knowledge comes from positional perspectives, and positionality determines what is and is not possible to know (Haraway, 1988). Savin-Baden and Major (2013) explained that to acknowledge researcher positionality, a researcher needs to acknowledge their positions that can potentially influence the research, situate themselves in relation to the participants as well as in relation to the research context and process. By being transparent about the process of knowledge formation, reflexivity strengthens the rigour of the study and increases the validity (Hall & Callery, 2001). Although there are likely to be numerous ways that the multiple layers that consisted of my position, below I describe several major aspects of it—how my position might have influenced the research questions, analysis and interpretation, and how I addressed them in the study. These descriptions are primarily based on what I have recorded in my research journal and analytic memos throughout the research process.

4.4.1 My teaching experiences

The fact that I am not a teacher was something that I was conscious of throughout the research process. Although I completed my undergraduate and master's study in science (physics) education, with all required training including internship placement, I was never a full-time science teacher and thus do not identify myself primarily as a secondary school teacher. I taught several weeks-long programmes on science and related subjects at schools and a university extension school, which was helpful for improving my understanding of teaching and shaping my researcher identity. I had also been involved in several teacher training programmes in Korea and the UK, including undergraduate and master's level courses, and worked with science teachers, interviewed them and observed their lessons (Erduran, Guilfoyle & Park, in press; Guilfoyle, Erduran & Park, 2020; Park, Erduran & Guilfoyle, in review). Yet, I also recognise that such a background of mine does not qualify me to have a 'teacher's eye' when looking at how my research participants taught and assessed NOS. Knowing the potential limits of something that I am not familiar with, one of my key aim was to get as close to the insider's viewpoint as I could (Merriam et al., 2001). Through my two-year-long fieldwork and multiple interviews with each of the teachers, I could feel that I gradually have a better understanding of the teachers' practice and sympathise with them. For example, in the first-year interviews, when the teachers criticised the government's policy to assess SIE the same way as for other subjects, I could not see the problem of it clearly even after they explained it to me, so I made a memo in that transcript. However, when I revisited the same transcript a year and a half later, I could sympathise much better with their concerns based on my improved understanding of the school culture and how teachers' planning and enactment of classroom assessment are situated within the system.

There were other times that I found it helpful to have a researcher (rather than teacher) identity when studying the teachers. Once the teachers' talks, behaviours and decisions made sense to me through the development of empathy, analysing and interpreting the meaning of such data required a researcher's perspective. Dwyer and Buckle (2009) highlighted, quoting Fay (1996), that knowing an experience requires more than simply having it, and it is necessary to switch back and forth between the insider's and outsider's perspectives. They also pointed out that when the researcher is the member of the group being studied, there is a risk that participants 'make assumptions of similarity and therefore fail to explain their individual experience fully' (p. 58). In my interviews, I sometimes made it clear to them that I was not familiar with what they were speaking about so that they could explain it to me in full.

4.4.2 My geographical background

The geographical and cultural backgrounds of the researcher is a key element of reflexivity (Mauthner & Doucet, 2003). I attended school and university in Korea, in which this thesis fieldwork was conducted, before I moved to England for my doctoral study two years ago. In England, I worked with teachers and students in the Oxfordshire area as part of a funded research project. As I analysed the teacher interview and lesson video data from the project, I realised one core element in the two seemingly disparate education systems—the culture of tests and competition (see Section 3.2). The parallels between England and Korea with regard to the issues of classroom assessment and its relation to external high-stakes examinations were often the topic of discussion in my meetings with doctoral supervisors, both of whom were former schoolteachers and were aware of how these issues could be manifested in different contexts. For example, in one meeting, I was told about the problems when England had science coursework in GCSE—such as cheating and unreliability of teachers' marking (Crisp, 2008)—I felt these concerns resonating with the recent debates in Korea around increasing reliance on teacher assessment. As such, examining how the same educational issue unfolds differently across different systems helped to strengthen the analysis and arguments. It enabled me to discern what aspects of the teachers' practices and views could be specific to one system and what could be more generally transferable to other systems, which concurrently contributed to my thinking about the generalisation problem with regard to my study (Section 4.1.3). For example, when I found that the teachers prioritised fairness and discrimination in their assessment practice, this fact initially was considered relatively unimportant since it was nothing surprising in the Korean assessment system that I was used to. However, I later realised that this would still be a significant factor from a global viewpoint. In countries such as the UK, where teacher assessments are mostly low-stakes, these mundane practices of Korean teachers can be interesting, unique and meaningful and therefore offer insights through comparison. As such, throughout the research process, switching between the 'insider' and 'outsider' positions allowed a fuller and more nuanced analysis of the collected data.

4.4.3 My relationship with participants

I have indicated that I had known one of the three research participants, Dr Kim, before the study began. I first met her at a university graduate course for philosophy of science education, during my master's study in Korea. From the class discussions and personal communications throughout the semester, I knew that we had a shared interest in NOS. More than a year after she and I both left the university, I was looking for research participants. I thought of Dr Kim, contacted her, and was told that she came back to teaching and was teaching SIE that semester. She also introduced another teacher Ms Nam to me, so these two teachers became my participants, along with Ms You who joined later in the second year of the study.

In the first year, I made two visits to the schools, but I felt the interviews were not very smooth. This might have been due to the lack of rapport, but there were at least two additional sources of difficulties. The first was that both parties, participants and researcher, did not have good knowledge of what was going on. The teachers were not familiar with assessing or teaching NOS since it was their first time teaching the subject, and I was not familiar with understanding various school events from a teacher's viewpoint. A worsening factor was that questions related to NOS tend to require reflection and be difficult to answer. I knew from my experience that when teachers are asked questions such as *'What is NOS in your opinion?'*, a few seconds of silence is typical, sometimes followed by *'I haven't really thought about that'*. If I insisted, they would make something up, but the value of those made-up answers would be questionable. Further, repeatedly asking such challenging questions could have ruined my relationship with the teachers. From these moments, I felt it difficult to investigate deeply into their beliefs and reflections, so instead focused on understanding what was happening with the teachers.

In the second year, however, most of these problems were solved naturally. Dr Kim and Ms Nam, from the previous year's experience, were more comfortable speaking about their practices and sharing their views. They told me what they did last year, what they could have done better, what activity worked better than another and so on. Particularly I felt far more comfortable talking with Ms Nam after I learnt about her background and views and she about what I was looking at for the research. In Ms You's case, although she only joined in the second year, she was an exceptionally reflective and friendly person and seemed comfortable sharing her thoughts even about what I thought were hard questions, from the first interview with me.

Since the second year was under the influence of COVID-19, I was mindful of minimising the number and duration of visits and using the time efficiently, given that the teachers were already busier than usual with adjusting teaching plans, making online lesson videos and other extra duties related to the pandemic. There were also a couple of occasions that I was asked for some ideas and views about their teaching from the teachers, and I told them what I think, to the extent that it doesn't significantly affect their assessment practice (e.g., my feelings about their online lectures, ideas for activities to be used in the future semesters).

4.4.4 My academic and research backgrounds and epistemology

Another dimension of researcher positionality concerns a researcher's epistemological view on how knowledge is produced and evaluated, which in turn is shaped by my own ideas about research in general and my research topic in specific (Mauthner & Doucet, 2003). Qualitative research is a continual process of making methodological and analytical decisions among many possible options, and these decisions reflect my own views about research. Along with the other positionality factors described earlier, I was substantially influenced by my previous research experience in my decision-making process—selecting NOS assessment as the main research focus, employing an in-depth case study method, conducting fieldwork in Korea, and focusing on teachers rather than students, among many others. Here I briefly reflect on the first aspect about the selection of research focus.

By the time the fieldwork began, I had four years of research experience with NOS and other related topics. I had undertaken undergraduate and graduate-level courses on the history of science, philosophy of science, and science and technology studies, which constitute the foundation of NOS. I had also done research on textbook and curriculum representations of NOS in different countries, as well as theoretical projects on the nature of experimentation, pluralism in science education and the relationship of disasters and science education, all of which substantially drew on NOS research. My knowledge of NOS was significantly helpful in terms of access to the literature and conceptualising the research. In particular, since my research mainly used FRA as an analytical framework (Park, Yang & Song, 2019; Park, Yang & Song, 2020; Park, Wu & Erduran, 2019a; Park, Wu & Erduran, 2019b), where the connections and interactions between different NOS aspects are acknowledged (Erduran & Dagher, 2014a; Peters-Burton, Parrish & Mulvey, 2019), teacher practices in NOS could be interpreted flexibly and holistically, and as making connections between various elements of NOS (Section 7.4).

However, there were also moments when my knowledge of and familiarity with NOS research became a bias and made it hard to discover new things. For example, when I was observing Dr Kim's lesson on the paradigm shift in the first year, I thought that some of the examples that students thought were a paradigm shift were not actually an example of it, since I had familiarity with the concept of a paradigm from my postgraduate philosophy of science coursework and knew that there are certain conditions that a paradigm needs to satisfy, and there existed different interpretations about these conditions among philosophers. It was almost a year later that I found, while reviewing the recordings, that the teacher and her students were sharing their own, pedagogically adapted conception of a 'paradigm' and a 'paradigm shift', so the examples should first be interpreted in the context of their own culture, rather than the view of academic philosophy that I initially had (see Section 7.3.2). This epiphany deepened my understanding of what happened in the classroom.

Unlike NOS, (classroom) assessment had not been my primary area of specialisation as a science education researcher before my doctoral work. Although I was deeply aware of the importance of assessment in determining educational practice, I used to think of assessment as a primarily quantitative and statistical matter that concerns large-scale tests such as PISA and TIMSS. After I began to read about assessment, however, I figured that much of the data could be re-interpreted through the assessment lens. When I observed and recorded Dr Kim's paradigm shift lessons in the first year, they did not stand out to me at the moment. My post-observation summary read: *'Students present examples of a paradigm shift, the teacher makes some comments, and not much teaching and learning happens'*. When I revisited the lesson with the formative assessment lens (particularly informal formative assessment and assessment conversation) in the next year, however, I could see that the feedback, students' verbal and nonverbal reactions to the presentation, and the teacher's summary were providing ample opportunities for learning, and the teacher was continuously collecting information about student learning. This led me to interpret the classroom interactions that happened in her lesson as an ongoing, casual, informal formative assessment throughout which the teacher constantly looks for information about student learning and acts upon it to facilitate students' NOS understanding.

4.5 Ethical issues

It is essential for all empirical research involving humans to consider ethical concerns. The pilot and main studies for this thesis were conducted in full accordance with the research ethics standards set out by the Central University Research Ethics Committee (CUREC) of the University of Oxford, British Educational Research Association (BERA), and the Bioethics and Safety Act of Korea, National Research Foundation of Korea. The research proposal and ethical considerations had been approved by the DREC at the Department of Education, the University of Oxford before the pilot study, and an extension was granted for continuing the data collection in the second year. In this section, I describe some salient concerns that I faced throughout the research design, data collection, analysis and writing phases.

It is essential for classroom-based researchers to ensure that participants understand what is going on and give permission to the researcher. In my first visit each year, I met with the school's vice-principal to explain the data collection process and obtained consent for participation. To ensure participants' volunteerism, I obtained informed consent after explaining the nature of the study, the implications of their participation, and their rights to opt-out at any time. I knew that two of my teacher participants had prior engagements in research, both as a researcher and a participant. I gave them some time to read the information sheet, explained several major points including the number and timing of lesson observations and interviews to be made. For Ms You who was not familiar with research participation, I spent a longer time explain the documents line by line, especially confidentiality and anonymity. I described to her what the collected data would eventually look like in writing, using hypothetical examples (e.g., *'The paper would read that, Teacher X, a physics teacher in an urban high school, did this in her classroom because she thought it would … it would not say* *that this teacher named blah blah in what school taught it well or badly'*). This was to ensure that she understands that my study was *not* to evaluate the quality of their teaching or to compare across teachers and schools, since teachers may find it uncomfortable when they think that their practice is judged, scored and ranked by someone else.

This study was designed following the CUREC Approved Procedure 15 (*Research on Teachers and Teaching in Educational Settings*) where the focus is on the teacher, and students' words or work are used only to the extent they are produced in the normal course of teaching and learning. In accordance with the Approved Procedure, and as agreed with the teachers, I explained to the students that I would be sitting at the back of their classroom, just observing and recording to remember better for a research study, and they can tell the teacher or me anytime if they do not want me to use what they said or did. Information sheets and opt-out forms for parents/guardians were printed and sent out to homes. As with the teachers, students were fully anonymised and only referred by numbers in the thesis text. Since this study did not involve any form of intervention to the lesson, there were no alterations to the educational experiences provided to the students. However, since the presence of a stranger in the classroom may have some unwanted influence on the students, I discussed this with the teachers and made efforts to minimise the observer effect (see Section 4.3.4.3).

Finally, in a small-scale, in-depth case study such as this, there is often the danger of the participants being identifiable from the detailed description of their backgrounds and practices even when pseudonyms are used (Walford, 2005). This is an important ethical issue because sometimes participants might suffer harm or embarrassment as a result of published research, in various ways unexpected by the researcher (Punch, 1986). Although anonymity is a standard ethical practice today, qualitative researchers have raised questions about the possibility of anonymity. Walford (2005) pointed out that in qualitative case studies, it is inevitable that many people—the headteacher, other teachers, students and school staff, among others—become aware of the identity of the researcher as well as who are participating during the research process, so it becomes very difficult to hide the identity of schools or the individuals. Such difficulty of ensuring anonymity is even amplified in the information age where research spreads instantly to the world and anyone can easily search for individuals and organisations in seconds (Tilley & Woodthorpe, 2011). In my study, I took a slightly different approach to anonymity, by ensuring that my participants know what and how I write about them and are comfortable about the possibility of its publication, besides simply changing their names in the manuscript. Given that the point of anonymity is to protect participants from possible risks, it seemed reasonable to prioritise their opinions about how their ideas and experiences are publicised. Once the findings were outlined, I shared with the teachers a poster and a PowerPoint presentation summarising the content of the thesis, both originally produced for conferences. With this, I asked if they were fine with the way I wrote about their stories. This member checking process not only contributed to reliability but also enabled addressing possible ethical concerns around anonymity.

4.6 Data analysis

4.6.1 RQ1 (Chapter 5)

The analysis was guided by the research questions—*How do teachers view and practise the summative assessment of NOS in the classroom?*, *How does a team of three science teachers engage in NOS summative assessment in a school?*, and *What does their decision-making process reveal about the potential issues in the summative assessment of NOS?*. Interviews and observational data were transcribed verbatim for analysis. The first step to case study analysis was to familiarise myself with the data (Saldaña, 2013; Silverman, 2014). Since this study had multiple sources of information, it was necessary to develop an analytical approach that is both efficient and comprehensive. After importing the data into MAXQDA (VERBI Software, 2019), a qualitative analysis software, I read the data multiple times in varying ways. I first reviewed each teacher's data to get a sense of how they experienced the teaching and assessment of NOS. I then reread the transcripts and artefacts from all three teachers in chronological order, with a focus on the second year. In these analytical steps, I did not have specific focal questions in mind, but instead aimed for a holistic understanding of the data, focusing on what was in the data, where the different ideas were located and so on.

After this familiarising stage, I started coding the data, with the research questions in mind. The analysis for Chapter 5 was focused on the three teachers as individuals, so the main purpose of coding was to characterise each teacher's views and practices relating to NOS summative assessment and organise these codes around key themes. Coding is a process of 'condensing' the data that allows for deeper reflection and discovery of meanings by categorising similar data chunks (Miles, Humberman & Sandana, 2014). In this initial coding stage, a combination of deductive and inductive approaches was used to code the data (Miles et al., 2014). Several deductive, a priori codes came from the relevant literature, particularly classroom assessment theories (e.g., reliability, fairness, discrimination) and theories in NOS education (e.g., reflective question, relevant content knowledge, science vs. NOS). As the analysis progressed, empirically grounded codes were added to the list. When all the interview data were coded, I went on to search for themes that 'capture something important about the data in relation to the research question and represents some level of patterned response or meaning within the data set' (Braun & Clarke, 2006, p. 82). Each teacher's views were divided into three categories: their views on assessment, views on NOS and its teaching, and views on assessing NOS in the new curriculum, and themes for each category were developed from codes and coded segments. For example, the theme of 'a focus on the application and expression of NOS ideas' about Dr Kim's NOS assessment views was developed from her interview excerpts that were coded into 'application of NOS in life', 'talk in their own words' and 'relating NOS ideas'. Once emergent themes were constructed for each teacher, I then reviewed the whole data to draw cross-teacher conclusions (Yin, 2017). These comparisons were made in different manners: between the three teachers, between their views on NOS teaching and those on NOS assessment, between their views on science assessment and NOS assessment, to maximise the understanding about diverse aspects of the case in focus.

The analysis of their enactments of summative assessments followed a similar coding procedure, but in this case, the main data source was the evidence for each teachers' enactment and their reflections during the interview. There were three NOS assessment activities enacted by the teachers. One activity (the clouds activity) was video-recorded for each teacher, while for the other two activities, the analysis relied on lesson materials and teacher interviews. For each activity, I first focused on the teacher's practice before, during and after the activity, using all available data types. Before the activity, I looked at under what conditions the activity would be enacted, what instructions the teacher would give, and whether and how she would address contents from previous lessons. During the activity, the focus was on whether and how often the students ask questions to the teacher and how she responds to them. After the activity, I focused on the teacher's summary, feedback, comments and reflections shared to the students, as well as what she communicated about the next lesson, if any. In analysing the enactments of the activities, although the research focus was on the teacher, I also looked at student behaviours to gain a holistic understanding of the instruction. Similar to when analysing their views, here I also developed themes for each teacher and then compared them across the three teachers, and across the three activities. Once this was complete, I then revisited the analysis of their views to compare the teachers' views and practices, and if any disagreements, I sought the data for possible sources of it. This led to identifying the overarching patterns in the teachers' views and enactments of NOS assessment.

4.6.2 RQ2 (Chapter 6)

Having considered the individual teachers' views and practices in Chapter 5, my focus in Chapter 6 was to explore the issues arising from teachers' decision making about NOS summative assessment. The aim of Chapter 6 was to investigate the same three teachers' practices in the same timeframe as Chapter 5, but here I looked at the broader process in which the three teachers engaged in planning, enacting and marking the assessment activities as a team. To understand this process, various forms of data collected in the second year were analysed holistically to produce a flowchart that chronologically catalogued major events such as the tasks, actions, communication between the teachers, assessment-related decisions, and meetings that occurred throughout the semester. This allowed identifying five steps in the collaborative summative assessment, from outlining and communicating assessment plans early in the semester, and marking and awarding grades at the end of the semester (Table 6.1). I then characterised the degree to which each step was collective or individual, and if collective, what the aims and objects of the teacher collaboration were, using the framework of Allal and Mottier Lopez (2014) as detailed in Section 6.3.

To address *What does their decision-making process reveal about the potential issues in the summative assessment of NOS?*, I focused on the assessment meeting transcript collected at the end of the second year, where the teachers sat together to discuss the criteria for marking and grading criteria. To analyse the assessment meeting, I first divided the transcript into three large sections on the basis of the task being discussed (among the three assessment tasks) and sought any issues in need of resolution arising during the conversation. For example, while discussing the second activity (the clouds activity), the teachers noticed that they enacted it in different manners (one teacher outside the classroom and the other two in the classroom). After identifying episodes such as this, the conversations were coded based on the nature of the issue being discussed, using both theory-driven and descriptive codes. These codes were then aggregated to develop one or two themes that seemed central to each assessment task, revealing the challenges the teachers were faced with about the summative assessment of NOS.

The analysis that I have described did not occur at a single time but was a multi-cycle and iterative process. The coding was done at three time points: after the first-year data collection, after the first visit in the second year, and after the completion of data collection. It is common in qualitative research to conduct coding as fieldwork continues, which allows for collecting new data that fills in gaps, testing new ideas that emerge from analysis, and formulate rival hypotheses that can challenge researchers' initial assumptions (Miles et al., 2014). Codes and themes were reorganised, reassembled and renamed as the analysis progressed, to ensure the validity and reliability of analysis (Saldaña, 2013).

4.6.3 RQ3 (Chapter 7)

For RQ3, the analysis was focused on the two sessions taught by Dr Kim in the first year of the study, along with interviews with her on this activity and other classroom materials such as the PowerPoint slides made by students. Data were transcribed verbatim, imported into MAXQDA and then coded in three cycles. The analysis was guided by the research question on the use of formative assessment in the lesson: *What are the instructional strategies that allow for teachers' formative assessment NOS? What are some potential issues and challenges involved in the formative assessment of NOS?*

4.6.3.1 Deductive (ESRU) coding

For the first cycle of coding, Ruiz-Primo and Furtak's (2007) ESRU model (eliciting, student response, recognising, and using information) was used to identify the structure of the informal formative assessment occurring in the teacher-student, student-student interactions throughout the Q&A sessions. The ESRU model was proposed as a way to capturing differences in teachers' informal formative assessment practices (Ruiz-Primo & Furtak, 2007). It has been suggested that the traditional models such as the IRF model (initiation, response, and follow-up; Sinclair & Coulthard, 1975) and the IRE model (initiation, response and evaluation; Mehan, 1979) are more suitable for lessons where teacher talk is predominant and the 'correct response' is pursued (Ruiz-Primo, 2011) than for lessons with frequent participation of students and open-ended lesson topics. The ESRU framework was initially developed to analyse formative assessment conversations in the context of scientific inquiry (Ruiz-Primo & Furtak, 2007), and acknowledges the importance of peer feedback and the possibility of incomplete assessment conversation cycles (Ruiz-Primo, 2011). These characteristics of the ESRU model made it a suitable framework to analyse Dr Kim's paradigm shift lesson, which had as much student talk as teacher talk, and where the lesson goal was to understand NOS.

Figure 4.2 shows an example of ESRU cycle coding. This episode followed Group 1's presentation on the discovery of the double helix structure of DNA as an example of paradigm shift. She starts by asking an **eliciting** question to the presenters about why the double helix case shows a paradigm shift. A silence of 10 seconds followed, which was coded as a **student response**. Coding silence as a form of response was considered reasonable given that it provides the teacher with information about their understandings or feelings. Then the presenters verbally gave **another response** to Dr Kim's question based on the utility of the discovery. Dr Kim then briefly **recognised** their response with 'Yes', and **used** the group's insufficient response to reiterate what paradigm is and provide a better answer to her own question. When the group agreed to her interpretation, she **recognised** their selection of the example and moved on to initiate another elicitation.

This first cycle coding enabled preliminary sorting of the assessment conversations, and also the identification of the episodes. Given the nature of the assessment

1	Dr Kim:	… So today's presentation was meant to be about finding and introducing a historical inquiry that led to a paradigm shift. So I'll ask a question first. It applies to all groups. Why do you think this is a historical case that shows a paradigm shift? I'd like anyone from the group to answer.	ELICITING
2	Presenters:	[Silence for 10 secs] When we know the helical structure of DNA, it makes opportunities for human cloning, and tomato bearing potatoes, and the like, so we can say it's a paradigm.	STUDENT RESPONSE
3	Dr Kim:	Yes. A paradigm is, easily speaking, a system of thinking that prevailed people's thoughts in a certain era. Then you mean that there was a huge change in biology before and after discovering the double helix structure?	RECOGNISING USING ELICITING
4	Presenters:	Yes.	STUDENT RESPONSE
5	Dr Kim:	Good. In that regard, you selected a really good topic. …	RECOGNISING

Figure 4.2 ESRU coding example

conversations that came after a presentation, an episode usually started with a student or the teacher asking the presenters to clarify, explain or elaborate what they said, and ended with the teacher using the information to take actions. A total of eight formative assessment episodes were identified in the transcripts. Given the exploratory nature of the analysis, employing the ESRU cycle was intended as a preliminary step to discovering meanings, rather than quantifying the cycles or their patterns as was the case for several other studies (Nieminen et al., 2016; Rached & Grangeat, 2020; Ruiz-Promo & Furtak, 2006).

4.6.3.2 Inductive coding

Although the ESRU cycle is useful in characterising the dialogic structure of whole-class conversations and the role of each statement within that structure, it provides limited information as to how the statement relates to the teacher's instructional goals and strategies in the given context. Therefore, another cycle of coding was necessary to connect the episodes with the teacher's practices. In this second-cycle coding, conversations in each episode were coded inductively, with a focus on their semantic meanings. For each coded segment, I asked myself questions such as *'What is the intention of this talk?', 'How does it connect to what comes before and after?' 'What purpose does this talk serve for the teacher, the presenters and other students?'*. This

1	Dr Kim:	… So today's presentation was meant to be about finding and introducing a historical inquiry that led to a paradigm shift. So I'll ask a question first. It applies to all groups. Why do you think this is a historical case that shows a paradigm shift? I'd like anyone from the group to answer.	STATES LEARNING GOAL
2	Presenters:	[Silence for 10 secs] When we know the helical structure of DNA, it makes opportunities for human cloning, and tomato bearing potatoes, and the like, so we can say it's a paradigm.	HELPS TO RELATE THE EXAMPLE WITH NOS
3	Dr Kim:	Yes. A paradigm is, easily speaking, a system of thinking that prevailed people's thoughts in a certain era. Then you mean that there was a huge change in biology before and after discovering the double helix structure?	EXPLAINS A NOS CONCEPT
4	Presenters:	Yes.	
5	Dr Kim:	Good. In that regard, you selected a really good topic. …	MAKES A COMPLIMENT

Figure 4.3 Inductive coding example

helped me characterise the nature of each talk. When the context and intention of the talks were not evident from the text, I returned to the recording and watched it again to gain a more nuanced understanding of the meaning.

In addition to assigning the codes to segments, I also made analytic memos to record emergent ideas on the data itself as well as the analysis that could be elaborated and confirmed (Birks, Chapman & Francis, 2008). These memos multiplied as I read and coded the data, which helped to create connections between different portions of data, different data types, and different participants. Since I coded the data inductively in multiple iterations, after each round, I recorded the major changes I made using the logbook function in MAXQDA. As such, I kept track of how the analysis evolved and how each coding and categorisation decision was justified, which contributed to within-coder reliability (O'Kane, Smith & Lerman, 2019).

An illustrative example of inductive coding is provided in Figure 4.3. As compared to Figure 4.2, the codes here are more representative of the purposes of the talk made by the teacher. Since the lesson started directly from the presentation without introductory remarks by the teacher, she starts by **clarifying the learning goal,** which was to learn about the historical examples that show a paradigm shift. Her next talk included **explaining the NOS concept** of interest (i.e., what paradigm is) and **helping students connect their example to the target NOS aspect**. Then she **made a**

compliment about their selection of the example. As demonstrated by this example, the second-cycle coding allowed contextualising each statement within the paradigm shift and characterising the nature of instructional strategies. This analysis added nuances to the structural analysis based on the ESRU cycle, and the second-cycle codes were later aggregated and compared across all episodes to derive broader themes across the entire lesson.

The interview data was also read through and coded in parallel with the lesson transcripts. I first distinguished instances of the interviews in which she talked about this activity. This interview data included both the post-observation interview in the first year (i.e., when the activity was enacted) and Dr Kim's recollection of the activity in the later interviews and informal communications. I sometimes asked her for a general reflection on the activity, while at other times asked questions about specific moments that interested me during the observation. The initial code list was then merged, divided, renamed and grouped into categories to better represent the data and address the research questions. Table 4.2 shows the final list of inductive codes and what codes were found in the formative assessment conversations following each group.

4.6.4 Rigour and trustworthiness

As a qualitative research project, it is crucial to ensure that the analysis is a rigorous and plausible representation of reality (Corbin & Strauss, 2015). To ensure the rigour and trustworthiness of the interpretation, techniques such as testing rival explanations, thick description, examining negative cases and triangulation across data sources (Creswell, 2013) were employed throughout the analysis for RQ1 and RQ2. Triangulation entailed comparing the interview data with lesson observation data, comparing one teacher's description of an event with that of the other teachers, and comparing the initial assessment plans with actual practice (Creswell, 2013; Patton, 1999). In order to maintain consistency of the coding, a codebook that consisted of categories, codes, coding criteria and illustrative examples was generated and revised as the analysis proceeded. To ensure interpersonal reliability, the preliminary and final coding results were shared with two supervisors of this project during our regular supervision meetings and via email and were checked for accuracy and consistency.

Table 4.2 List of inductive codes from Dr Kim’s paradigm lesson

Dr Kim’s instructional strategies			
Theme	Code	Description	Example
Multicontextualing NOS*	Uses multiple historical episodes to teach paradigm shift	Teacher designs and enacts the assessment activity where students present diverse historical episodes to show paradigm shift.	N/A*
	Clarifies learning goal	Teacher states the goal of learning paradigm shift from multiple contexts.	Teacher: Good. Well done. … In the end, you had an opportunity for thinking about the (content) knowledge and nature of science about at least six examples. I hope you could learn something from them.
Mapping examples onto NOS	Clarifies learning goal	Teacher reminds students of the learning goal of understanding historical cases that led to a paradigm shift.	Teacher: So today’s presentation was meant to be about finding and introducing a historical inquiry that led to a paradigm shift.
	Explains a NOS concept	Teacher explains what a paradigm is.	Teacher: Yes. A paradigm is, easily speaking, a system of thinking that prevailed people’s thoughts in a certain era.
	Helps relating the example to the paradigm shift	Teacher asks or tells how the example shows a paradigm shift.	Teacher: So I’ll ask a question first. It applies to all groups. Why do you think this is a historical case that shows a paradigm shift? I’d like anyone from the group to answer.
	Gives a compliment	Teacher acknowledges the presenters for linking examples to paradigm shift.	Teacher: So well done on presenting the continental drift. Then we have Q&A. I can’t give you much time, so just one minute, let’s get one question. You explained really well about the paradigm shift, and connected it with continental drift so I want to clap.
Focusing the discussion on NOS	Delivers relevant content knowledge	Teacher provides scientific knowledge necessary for understanding NOS.	Teacher: So Einstein’s strong theory declines, and cosmology is, it is what uncovers the origin and evolution of the universe. In so far as the origin, and the evolution process, people have made remarkable development since. So we know the Big Bang theory is reliable among cosmological theories, and how the universe expands, with acceleration or otherwise, people are doing research about that. There are some conclusions out there. In that regard, it is clearly an example of a paradigm shift, right?

* The multicontextualisation of NOS occurred during the design and selection of the assessment task and during her concluding remark (see Section 7.3.1), so is not represented in this table.

Table 4.2 (continued)

Dr Kim’s instructional strategies			
Theme	Code	Description	Example
Focusing the discussion on NOS	Clarifies learning goal	Teacher states or clarifies the NOS focus of the lesson.	Teacher: So now, we are not focusing on the scientific knowledge to learn about Kepler’s third law or specifics of Newton’s law, so we can move on. But what’s important is, when you present, it’s important that you summarise and present what you can digest, right? In that regard, although it’s not a knowledge-focused presentation, you need to be able to introduce it [relevant scientific knowledge] briefly. The part you are talking about.
	States expectations about content knowledge	Teacher tells students to what extent she expects content knowledge to be presented.	Teacher: … It would’ve been nice if you focused on that and presented a bit easier. All groups alike. Let me tell you. Only the things you guys can understand. You don’t need to bring all the evidence. Only one of the three, but it’s good if you digest it and express it in your own language. Anyway I know that you prepared a lot …
	Shuts down detailed content questions	Teacher discourages or shuts down when students ask questions on scientific details.	Student: … Could you explain what Kepler’s planets are about? Teacher: So now, we are not focusing on the scientific knowledge to learn about Kepler’s third law or specifics of Newton’s law, so we can move on.
Exploring diverse NOS aspects	Explains a NOS concept	Teacher explains additional NOS aspects that she wants to address.	Teacher: Then in relation to tentativeness, the nature of science is that, science is technology-dependent. Or one development in science and technology can bring about another development. You can say these things.
	Helps relating the example to other NOS aspects	Teacher provides scaffolds to make the link easier for students.	Teacher: Any questions related to that? (Silence for a sec) Is it too difficult? Do you need another summary? When we studied with the electricity example, don’t you recall anything? Something leads to something. The blank filling that you like. Hint, the development of science … (Students silent) (Laughter). So the limitation of plate tectonics, will be able to be overcome with a new theory, through the development of the exploration techniques. Right? (Yes.)
	Connects to previous learning	Teacher refers to previous learning about content, NOS and history.	Teacher: When we studied with the electricity example, don’t you recall anything? Something leads to something. The blank filling that you like. Hint, the development of science …

We discussed what each of us saw in the transcripts, what emerged from our different perspectives and any alternative interpretations of the data (Morse, 1997). Member checking was also done by sharing the analysis with participants and checking for accuracy and resonance with their own experience. Finally, I describe my positionality as a researcher to enable readers to understand how my positions could have possibly influenced the analysis (Section 4.4).

For RQ3, there were several specific considerations for ensuring reliability suitable for assessment conversation analysis. Peräkylä (2004) argued that ensuring the reliability of conversation analysis begins from the 'selection of what is recorded, the technical quality of recordings and the adequacy of transcripts' (p. 288). This is because not everything from the field can be recorded and transcribed due to the constraints of time and resources. For lesson recordings, given that the Q&A sessions after presentations were of focus, I only transcribed these sessions and only made summaries of actual group presentations. Indistinct sections in the recordings were resolved through repeated listening. The adequacy of transcripts also needed to be ensured, particularly considering that the original data was in the Korean language. Although there is no perfect way to translate transcripts from one language to another while preserving the subtle nuances, it is still important to generate the best possible translation for the accurate representation and analysis of data (Chenn & Boore, 2010; Twinn, 1997). I, a native Korean speaker and fluent English speaker, first transcribed the recordings verbatim in Korean and then translated these Korean transcripts into English. The original and English versions were then examined by an external reviewer who was bilingual in English and Korean and had attended schools taught in both languages—so was familiar with day-to-day classroom language. To increase the transparency of analysis, I also present the actual transcripts in the thesis as block quotes so that the reader can compare the data with my analysis (Peräkylä, 2004; Richards & Seedhouse, 2005).

4.7 Chapter summary

As a small-scale, in-depth qualitative study, this thesis draws on various types of data and uses multiple analytical methods to reveal diverse aspects of the teachers' practices as fully as possible. The aim of this chapter was to describe and justify the case study design for addressing the research questions, the evolution of research design since the pilot study (i.e., the first-year data collection), methods for data collection and analysis, and my positionality as a researcher in relation to this research. The single case research design allowed an in-depth investigation of how three teachers in one school dealt with the instruction and assessment of NOS and what challenges arose in the process. Data sources such as teacher interviews, lesson videos, assessment meeting recordings, classroom artefacts and other documents were collected over two years and analysed qualitatively. Thematic coding and conversation analysis techniques were used to address the research questions about teacher views and practices in NOS assessment. Various aspects of researcher positionality played a role in the design and interpretation of the collected data. In the following three chapters, I describe and discuss the findings of the study.

Chapter 5 Science teachers' views about and enactments of summative assessment of NOS

How teachers view NOS and its teaching has been suggested as a key determinant of successful NOS learning (Abd-El-Khalick, 1998; Bell et al., 2016; Erduran et al., 2020; Hanuscin, 2013). Studies have found that teachers' good knowledge of NOS ideas does not automatically translate into effective NOS teaching practice (Lederman, 1995; Lederman & Zeidler, 1987), while many teachers find it difficult to rationalise how NOS fits into their instruction within high-stakes testing regimes (Bell et al., 2000). When it comes to the *assessment* of NOS, such difficulties of rationalising are even more evident (Hanuscin et al., 2011), which is unsurprising given that assessment by teachers is significantly affected by the external, high-stakes examination system (Black et al, 2010) that are typically focused on content knowledge of science. The increasing awareness that summative assessment should be seen as a learning process rather than being separate from learning (Black et al., 2010; NCME, 2018) poses the question of how NOS assessment in classrooms can create meaningful learning opportunities whilst ensuring the validity and fairness of the assessment. Despite the abundance of research on teachers' NOS understandings (Abell & Smith, 1994; Brickhouse, 1990; Lederman, 1996; Waters-Adams, 2006), the interplay of teachers' NOS teaching intentions, views of assessment, and their assessment practices has yet to be explored in current NOS research (Hanuscin et al., 2020).

To address such a gap, this chapter examines the case of science teachers in Korea, who take greater responsibilities for planning, conducting and marking summative assessments than in countries where summative assessment is concentrated on external tests. A Korean teacher takes charge of the summative assessment for the subjects she teaches each semester. If she is the only teacher for *Physics I* to a year group, she is solely responsible for its assessment; If there are other teachers for the subject, she will work with them to set up the plan for summative assessment that will apply uniformly to the entire year group.[5] The situation I describe in this chapter involves one Korean

[5] In the Korean timetabling system, decisions about who will teach which subject are the

high school with three teachers who were assigned to teach SIE. Each of them taught SIE to different Grade 10 classes (Dr Kim to four classes, Ms Nam to two classes, and Ms You to five classes, each with about 30 students), but since all eleven classes of students were supposed to be summatively assessed on the same basis, the teachers had to collaborate for the summative assessment. Such a system provides an opportunity to closely examine how a team of teachers practices NOS assessment within the school, as well as the potential differences in individual teachers' views and practices related to NOS assessment. The aim is to provide a detailed description of how the assessment of NOS was perceived, planned and practised by the three teachers with a focus on the second year of my research. The research question is: *How do teachers view and practise the summative assessment of NOS in the classroom?* More specifically, this chapter considers (a) how the teachers' views about NOS assessment differ and (b) what variation exists in teachers' enactment of NOS assessment activities, and (c) how each teacher's enactment aligns with their NOS assessment view.

In Section 5.1, I describe each teacher's background and prior experiences with a focus on NOS, and how they usually approach assessment in their science lessons based on the interviews. Given the high responsibilities that Korean teachers take in school-based summative assessment, these descriptions will be instrumental in understanding and interpreting their practices about the assessment of NOS that follow. In Section 5.2, I present how each teacher enacted the three NOS assessment tasks in their classroom. These tasks, aimed at teaching and assessing different aspects of NOS, were enacted in diverging manners by the teachers, which was partly reflective of their differing views and expectations. In Section 5.3, I provide several analytical themes that emerge from the data that show the interrelation between teachers' views and practices in the context of NOS assessment.

5.1 Teachers' background and views about assessment and NOS

The three teachers, Dr Kim, Ms Nam and Ms You, were from different subject

---

responsibility of the head of year who considers individual teachers' characteristics and preferences, and other school factors.

Table 5.1 Teacher views of assessment and NOS

	Dr Kim	**Ms Nam**	**Ms You**
Subject background	Earth science	Chemistry	Physics
Teaching experience	15–20 years[6]	10–15 years	15–20 years
SIE teaching experience	Yes	Yes	No
Science assessment	Prefers open-ended, free-response assessment Focus on the problem-solving process	Prefers free-response questions Focus on the problem-solving process	Focus on end products of student work to ensure fairness
NOS teaching motivation	NOS can promote decision-making abilities and a critical mind.	NOS would be beneficial to students' lives.	NOS is important for all students, enabling effective communication based on science.
NOS assessment	Prioritises the application of NOS understandings Focus on explaining NOS ideas in students' own words	Unsure about the value of assessing NOS Focus on participation and attitudes than end products.	Unsure about the value of assessing NOS Focus on feedback and collective learning

backgrounds and held different views about assessment in science, and especially about the assessment of NOS. Table 5.1 summarises the three teachers' subject backgrounds, teaching experience, and views of assessment in science and NOS. Each teacher's case is detailed below.

5.1.1 Dr Kim

5.1.1.1 Background and views on assessment

Dr Kim was an earth science teacher with a doctorate in science education, for which she undertook research on preservice teachers' NOS teaching aspirations. From such a background, she was familiar with the basic principles of NOS instruction such as the NOS aspects in the consensus view and the VNOS questionnaire, and FRA as an alternative approach to NOS. As an earth science teacher for 15–20 years, she had

[6] Unspecified intentionally for anonymity

knowledge in all subdisciplines—geology, astronomy, oceanography and meteorology—but had a particular enthusiasm about astronomy since her undergraduate years. As part of her graduate coursework, she had experience developing a couple of teaching materials related to NOS, such as activities on galaxy classification and models of the universe, and teaching these to students several times in an informal setting as well as to preservice teachers in an undergraduate course.

According to Dr Kim, in her assessment of science subjects, she tried to include as many open-ended, free-response assessment items as possible, because she knew that ‘kids who don’t understand rarely write anything’. She also preferred to have assessments that can show problem-solving processes, such as the ones asking to set up and solve equations. She emphasised assessing whether a student ‘knows the theory and can apply it, and then remember the equation, and using it to draw the answer’, although the task need not involve complex mathematical operations. She did not favour assessing by short-answer questions, drawing tasks, making things with hands and so on, because, in her view, they do not measure students’ abilities to think. In addition, she was particularly against the idea of group-based marking because it obscures individuals’ abilities and understandings:

Dr Kim: I really don’t like [assessing by] group activities. Because when you do it, I know that a few kids lead everything, and if they do it wrong, the whole group is messed up … So I like individual activities. In the individual task reports, I like to assess things including knowledge, in a sense. (Interview, June 2020)

As a long-time high school teacher, she was conscious of the ‘external pressure’ about accountability when planning assessment. Since science subjects are usually of high stakes, assessment tasks needed to be able to discriminate among students’ levels of knowledge of the subject. She commented that the need for discrimination and accountability at the high school level was often a crucial consideration in the design of the assessment:

Dr Kim: I work really hard to select problems, choices and adjusting difficulty, for one reason. **To clearly discriminate among the top grades [on a**

> **Stanine scale].** [7] **This is external pressure,** so that there's no dissatisfaction. It's needed for students, so I try to do that … **It's particularly important in this school** … So it's too bad for students, but the exams need to discriminate among their abilities. I change the questions and choices many times because of that.[8]

This remark also shows that such tendency was particularly visible in her school, which was located in Gangnam School District 8, where the competition and testocracy in the country's education system culminated (see Section 3.2).

5.1.1.2 Views on teaching NOS

Dr Kim characterised NOS as the knowledge about something scientists do and the features of the products seen from within and outside the scientific community. She believed that understanding NOS became increasingly important in contemporary society. She felt it problematic that the traditional image of science as being special and absolute has plagued the education of scientists as well as science teachers for a long time. However, in her view, having such a view of science is not helpful in today's world where 'information is everywhere' and 'proper decision-making requires the capacity to assess given piece of information from various angles and with critical mind' (Interview, July 2020). She said that such an ability is difficult to cultivate while teaching the (content) knowledge-focused curriculum, so it should be promoted through teaching NOS. In this regard, she thought that the introduction of SIE in the new curriculum was a good start, although the way SIE is taught in schools 'would eventually depend on how it is assessed, and especially how it would be reflected in

---

[7] Stanine (STAndard NINE) is a type of standardised test score used to compare a single score to a distribution of scores, on a scale of one to nine (Jesson, 2018).

Stanine	1	2	3	4	5	6	7	8	9
Proportion	4%	7%	12%	17%	20%	17%	12%	7%	4%

[8] When it was first introduced in 2018, SIE had been assessed on a Stanine scale as with other subjects. However, in 2019, the Department of Education amended the policy to assess SIE by three-tier system (high, fair and low). Consequently, SIE became one of the few subjects (mostly arts and sports) that are assessed as such.

students' university entrance' (Interview, April 2019).

5.1.1.3 Views on assessing NOS in SIE

From the one-year-long experience of teaching and assessing SIE, Dr Kim had developed some clear ideas about the aims and methods of assessing NOS. When asked what should be assessed about NOS, she clearly articulated her focus:

> Dr Kim: How well one understands the NOS. That's it … **Not just saying it blah blah only in words, but being able to apply their understandings when doing something** like inquiry examples or historical examples. 'This [case] is indeed about tentative NOS, this part relates to what', if they can say like these. (Interview, June 2020)

With the emphasis on students' expressions, she wanted to use the interview method to assess students' NOS knowledge in SIE. Were it not for the lack of time, she said, she would have used an oral test as the best way to measure what she believed to be the aim of teaching NOS:

> **Dr Kim:** **I really wanted to do** an **oral test,** I told you last time. If I had enough time I would've done so. **Have them sit, let them talk about NOS, or what they presented.** And then record what they express and assess … Isn't it important **whether they can actually explain, talk in their own words?** (Interview, June 2020)

Dr Kim's usual science assessment practice that stressed the process of student thinking was also reflected in her views about assessing NOS. Although her remarks above are focused on the cognitive and functional understanding of NOS ideas (see Section 2.3.1), when she was talking about the fairness of assessment during an interview, she also referred to the assessment of student participation and attitude. This is shown in the following quote:

> Dr Kim: And in terms of ensuring fairness, it's really different when doing Stanine written tests [for other science subjects] and assessing SIE. In SIE I record very hard. In a sense, when a kid is not diligent when participating in the activity, I write the clues to give them B. (Interview, June 2020)

In summary, Dr Kim held a clear and strong view on the NOS classroom assessment that focused on applying and explaining of NOS understanding as well as participation, which led to her preference for interviews as an assessment method.

5.1.2 Ms Nam

5.1.2.1 Background and views on assessment

Ms Nam attended the Department of Chemistry Education at university, after which she took a teacher qualification examination to be placed as a teacher in a public school in the Seoul area. At the time of research, she had 10–15 years of teaching experience at the high school level and was teaching SIE to two of the eleven classes in the cohort. She described that her prior experience with history, philosophy and NOS, although not much, was mostly from her training at the university and preparation for the teacher appointment examination. She remembered several basic concepts of philosophy of science from undergraduate courses but had almost never related such knowledge to her teaching before she taught SIE.

When asked about her usual methods for checking for student understanding, she answered that she relies on students' responses in the activity sheets, and the questions they ask while she walks around the classroom. When she was in charge of the assessment plan, she usually included written tests. Among different approaches to written tests, she preferred free-response questions to multiple-choice questions, since the former allows her to 'better see how much they [students] know … [and] see the process, where they're stuck, or didn't get the concept itself … better' (Interview, March 2020). Aside from written tests, she also utilised her own memos about individual students where she records 'some notable things about students, like if they asked questions' (Interview, March 2020). Her note-taking practice became even more regular after the Department of Education introduced the 'subject-specific abilities and remarks' as a section in each student's comprehensive school record. This new section required the teacher to write a paragraph (of 3–5 sentences) on every student's achievement in the subject taught. To write this section, Ms Nam said, she needed to 'find what this student is excellent on, but it's not easy to find it for every student'. Since this assessment by teachers contributed significantly to students' university entrance, she was keen to collect evidence of individual students' knowledge and progress during the course of her lessons.

5.1.2.2 Views on teaching NOS

Ms Nam described NOS as the 'characteristics of science, which are unique to science, as opposed to other things' (Interview, July 2020). She believed that NOS would be helpful to students and allow them to understand that 'science is, rather than being absolute and unchangeable, it is influenced by society and culture' and that 'it can change upon a new piece of evidence' (Interview, July 2020). These various ways of thinking about science, she stated, would be beneficial to students' lives, she mentioned.

5.1.2.3 Views on assessing NOS in SIE

Although Ms Nam sympathised with the importance of teaching NOS for students' lives, she was less certain about the value of assessing NOS in the classroom. When she was asked about what could be assessed about NOS, she mentioned 'testing the initial ideas, and then how they change' (July 2020 interview). Then she thought that questionnaire-style assessment would be a suitable, and feasible, method for assessing NOS.

Like Dr Kim, Ms Nam thought that 'process' was the key to the assessment of NOS. She made this point by referring to the previous year's assessment task (i.e., a group presentation on examples of a paradigm shift; see Chapter 7):

> Ms Nam: It's [assessing NOS] more about the process, like looking at the inquiry processes. For example, when doing the presentation about the paradigm shift, [I look at] the selection of a topic that well shows the change that we asked, how they unfolded it, how they understand well and present, so those things about understanding and performing the task. (Interview, March 2020)

At the same time, however, she said that her assessment was more focused on student participation, rather than the quality of student outputs. She did recognise that discrimination among students was possible based on their performance on NOS tasks, but instead gave full marks as long as the students had been actively engaged in the task:

Ms Nam: … Normally for other subjects, there is a huge difference between the eleven classes in the year group, but [for SIE] there was not much difference between classes … There would have been a difference in the performance levels but not for the participation and process.

Wonyong: More on the process, if [a student] did the best in the process and did all they were asked to.

Ms Nam: Yes, that kind of mindset. There would certainly be a difference in the outputs, but if a kid did their best … They might have different levels of knowledge, but if they actively participated … [I gave a full mark].

Wonyong: That's interesting. Because normally, when you are marking chemistry, you don't give full marks since a kid had done their best? (Chuckle)

Ms Nam: Right. I think assessing the process is more central [to SIE]. (Interview, March 2020)

From this conversation, we can see that Ms Nam did recognise the possibility of 'different levels of knowledge' of NOS that can be discriminated, but she did not turn this thought into action in her assessment practice due to her focus on participation and attitudes. She also mentioned the impact of the grading system on her practice of assessing NOS in SIE. The change in the assessment policy for SIE had changed from a Stanine system to a three-tier system in 2019, which significantly lowered the accountability pressure for the subject. This effect of assessment policy was stated by Ms Nam in the following quote:

Ms Nam: SIE is not assessed on a Stanine grade [Grades 1 to 9], so there's less pressure with discriminating … It's more of a process assessment, so, I can give many A's and I don't have to divide B and C's by some percentage, it's more important to encourage student participation … If they're fairly actively and diligently participating, then I can give an A, so I tend to think about how they can participate better. (Interview, March 2020)

Here, she makes it clear that the new grading system and less demand for discrimination was the main reason that made her grading focused on participation rather than the product of student work. She uses the word ‘process’ again, but this time, to mean active and diligent class participation, based on which she awarded letter grades. In summary, Ms Nam’s views of the assessment of NOS mainly was centred on student participation rather than the level of their knowledge. This was made possible because of the state-level assessment policy that reduced the accountability pressure and made students ‘less sensitive about their SIE marks’, as she said in the interview.

5.1.3 Ms You

5.1.3.1 Background and views on assessment

Like the other two teachers, Ms You also started teaching after finishing an undergraduate physics teacher training programme at university. Her first job was at a middle school, and she had since spent 12–15 years at public middle schools in the Seoul area. She was then placed at the high school in 2018. Accordingly, I interviewed her in her third year of high school teaching. She frequently referred to her transition from middle school to high school when describing her expectations about and experience of teaching and assessing SIE.

During her time at middle schools, the focus of Ms You’s assessment practice used to be more on managing and guiding students’ behaviours than on their intellectual development, so she allocated more marks on participation in lessons and tasks. However, after moving to high school, she found it difficult to give marks based on attitude and participation, because students were very pressured about their participation being assessed in the high-stakes climate. Unlike in middle schools, she explained, marks in high schools have an impact on students’ university entrance and subsequently their future. This led her to avoid some assessment practices such as attitude-centred marking and group-based marking, since there are some risks involved due to the high stakes of high school assessments. Consequently, her marking relied more on the end products of individual students that could be easily and objectively discerned: for example, when assessing a practical task, the results of the practical work recorded in the worksheet, and answers to the reflection prompts at the end of the worksheet (Interview, March 2020).

5.1.3.2 Views on teaching NOS

Ms You believed that learning about NOS would help students cultivate the scientific mind, even to those who pursue a non-science career. She thought that the lack of training in NOS was in part responsible for the proliferating fake news as well as the policymakers 'who easily get convinced by word on the street and make important decisions on non-scientific grounds'. A good NOS understanding, she said, would enable more productive discussion and communication between people based on scientific evidence (Interview, July 2020). This shows that Ms You were aware of the potential functional benefits of NOS learning that help individuals' social lives (Allchin, 2013; Ryder, 2001).

Before starting to teach SIE, she had very little knowledge and had not received any previous training on NOS. When she joined Dr Kim and Ms Nam to teach SIE, she was provided with the teaching and assessment materials used by the other teachers, which seemed challenging to teach for her. She was not even certain whether her students would understand the questions in the materials in the first place. With some doubt, she asked the other two teachers how students had done, but they told her that students had answered well. She recalled:

Ms You:	… When I received last year's materials, I thought it wouldn't be easy to explain these to students, as a teacher. It would be quite difficult to let them **read** those long paragraphs. So I thought, after I explain a concept, I should mention it again with actual examples, so the kids could understand the **concept** enough. So only after receiving the materials, [I could imagine] how the class can be done. But I didn't study [about NOS] much.
Wonyong:	When you say taking examples of concepts, what kind of **concepts** do you mean? … I wonder what you mean by understanding concepts.
Ms You:	Like, students think science is always right, but when you see how it changed, it changes after new evidence refuting it is found. So it is changeable, things like that.
Wonyong:	The fact itself.
Ms You:	Yes, itself. (Interview, March 2020)

As elaborated in her response to my question, she used the word ‘concept’ to refer to the concepts of NOS rather than scientific concepts, implying that she clearly recognises the cognitive learning goals for NOS and the value of exemplification as a way of achieving such goals. In the above conversation, she also refers to the issue of reading involved in teaching and learning NOS. She mentioned students’ ability to read longer writings about science, and also about the terms in the writings:

> Ms You: … Kids might not be familiar with those terms … There were some terms in the survey I received [VNOS], I thought these terms could be awkward and difficult for kids to take without any explanations, without any examples. So before doing that survey, we would need some examples to answer it. When going ahead [with the examples] in the lesson, kids might better learn about the nature of science, naturally, not as some terms. (Interview, March 2020)

Although Ms You had not taught SIE before, she felt that NOS was addressed implicitly while she taught regular science and physics subjects. She thought, however, that while students get to know about NOS ‘as a background in the process of getting [content] knowledge’ in other science subjects, NOS became much more explicit in the new subject SIE.

5.1.3.3 Views on assessing NOS in SIE

In the interview at the beginning of the semester, Ms You had an expectation that her old practice of attitude-centred assessment would be more feasible (and more reasonable) to do in SIE compared to traditional science subjects, for several reasons:

> Ms You: … I think it’d be possible to assess more about attitudes, since it [SIE] is less reflected in the comprehensive school records … Also when it comes to the experiments, this [SIE curriculum] is **not the kind of experiments with precise results**. So **it has no answer**, each one can present [their ideas], that kind of lesson … In physics, as long as the data is precise, the conclusion should accordingly come up precisely … But this [SIE] is, an ideal answer might exist but could be broad. (Interview, March 2020)

Here, she is drawing on the lower stakes of SIE for school-leaving grades, the nature of the subject which, to her, lacks ‘precise results’ and the answer might be ‘broad’. When I asked her to elaborate on this point, she made it clear by comparing it to the assessment in a practical physics lesson.

Ms You: In the case of physics, when you have accurate data, the conclusion must turn out precise. Of course, you might get the wrong data from the experiment, but as long as you have the data, there is mostly an ideal answer for how to draw the conclusion. And I can give some extra marks if they do some further discussion on that. But in this [SIE], I might say there is an ideal answer but it’s very broad. So when I ask a student to find out whether he knows well, it influences not only him but the whole class, as you exchange talks. **So rather than having an ideal answer and getting good marks, it is like students learn together as we exchange feedback, and unless they’re not engaged, I think most students will be given good marks**. (Interview, March 2020)

In the latter half of these remarks, it becomes clear that Ms You placed more value on formative feedback and collective learning than the summative considerations such as grading and differentiation. This, again, was associated with the nature of NOS knowledge perceived by her and the lower exam pressure in which NOS was assessed. Later, she used the example of social and cultural embeddedness of science to illustrate her perceived unsuitability of ‘assessing’ NOS, but thought that some aspects, such as scientific methods, could be more suitable for assessment:

Wonyong: What do you think should be assessed about the nature of science?

Ms You: … About the nature of science? In terms of assessment, in fact some things might be assessed, but some for other things, I can just teach that we can think like this, and broaden their ways of thinking. Those things might not be assessed. Because, for example, when we say that [science is] influenced by society and culture, we can’t assess if it’s yes or no … So like that, we offer some food for thought, and the student makes a choice. So some [NOS ideas] cannot be assessed as right or wrong – instead, we can assess that it could be thought this way and that way. And if a student’s inquiry method is not logical … whether they can apply inquiry methods logically to various phenomena, I think that can be assessed. ‘What you’re doing now is not scientific’, for example.

(Interview, July 2020)

As seen in this quote, her hesitance to assess NOS such as the cultural embeddedness of science was connected to the impossibility of judging right or wrong' NOS ideas. In contrast, it reveals that she thought that knowledge and application of scientific methods could be right or wrong. This stands in contrast with Dr Kim's view, who believed that 'there certainly exists a high level of understanding [of NOS]' and 'too naïve ideas like "science is absolute" are not acceptable' (Interview, July 2020).

To summarise, although Ms You had not taught or assessed SIE before, she had an interesting view about assessing NOS. She had an idea of NOS as a cognitive learning goal, but still questioned whether it would be possible and necessary to assess NOS, given its special characteristics. Instead, she thought that it would be more valuable to discuss NOS in class and give feedback, which will naturally lead to learning NOS. In other words, she viewed that summative assessment of NOS could be left out in favour of formative learning experience in the classroom.

5.1.4 Cross-teacher comparison

The three teachers' views on assessment reveal several interesting points. First, due to the special socioeconomic backgrounds of students and the high impact of assessments on university entrance, the teachers were commonly mindful of certain elements of summative assessment such as discriminating power and fairness. This resulted in certain assessment practices such as Dr Kim and Ms Nam's note-taking to record individual students' performance and attitude, in order to keep clear evidence of marking in case students and parents complain about the grade. Also, the teachers acknowledged the issues with group-based assessment, especially in relation to student inquiry activities. Dr Kim was generally negative about group activities since it obscures an individual's contribution, but when she had to do group activities, she asked students to clearly indicate each member's contribution. Ms You also recognised the difficulty of discerning individuals in group activities, particularly as she moved from middle school to high school where students care more about grades. Hence, she usually allocated more marks to the individual reasoning and results than the collective work. These two common aspects of the teachers' assessment views clearly suggest that their primary concerns have to do with fulfilling the summative function of assessment, relative to its formative function that promotes learning.

Second, specifically regarding the assessment of NOS in SIE, the three teachers varied in their perceived aims of NOS assessment. Although all three teachers referred to both cognitive and attitudinal aspects in learning NOS, the relative importance of teaching and assessing these aspects was perceived differently by the teachers. Dr Kim believed that it is necessary and also possible to measure students' NOS ideas and divide these ideas into different levels. Ms Nam thought that it is possible to do so, but in practice, she put more value on participation and attitude. On the contrary, Ms You, having not experienced SIE and NOS before, questioned the necessity of formally assessing NOS in classrooms. Underlying these perceptions were factors such as the relatively lower stakes of SIE and the open-ended and controversial nature of NOS knowledge. I will turn to their enactments of NOS assessment in Chapter 6 and consider how and to what extent these diverging perceptions and expectations were reflected in their assessment practices.

Third, there was a discrepancy between teachers' NOS *teaching* intentions and NOS *assessment* intentions. Although teachers have been reported to place limited value on NOS as an instructional goal (Abd-El-Khalick & Lederman, 2000a; Akerson & Donnelly, 2008; Herman, Clough & Owens, 2017; Lee & Witz 2009; Wahbeh & Abd-El-Khalick, 2014), how their aspirations and intentions for NOS *teaching* relate to those for NOS *assessment* has not been discussed much. Given this, it is worthy of noting that all three teachers recognised the value of teaching NOS, but they diverged in terms of the value of assessing NOS. While Dr Kim put as much value on the assessment as the teaching of NOS and thus expressed a clear preference in certain assessment methods, Ms Nam and Ms You were unsure about the value of assessing NOS in the classroom. These views seemed to be associated with their perceived aim of NOS instruction where more focus was put on learning through classroom participation and attitudes than producing grades. These incongruences suggest that teachers' intention to teach NOS may not directly lead to their intention to assess NOS, which sheds light on the need for considering the two views separately. Meanwhile, whether and to what extent their perceptions and intentions lead to their assessment practice is yet another question, which will be further explored in Chapter 6.

The fourth point concerns the relation between teachers' general assessment views (i.e., those in the context of traditional science subjects) and NOS assessment views, which were not necessarily in alignment due to several reasons. Dr Kim's emphasis of 'know[ing] the theory and can apply it, and then remember the equation, and using it to draw the answer' (Interview, June 2020) in science assessment was coherent with her purpose of NOS assessment, which was 'being able to apply their understandings when doing something' and 'talk[ing] in their own words'. Ms Nam's assessment focus in science was on cognitive aspects of learning such as problem-solving processes, but when talking about her view on NOS assessment, 'process' was taken to be about students' participation rather than their cognitive understanding of NOS. Finally, Ms You was an interesting case. As a former middle school teacher who recently moved to high school, she was generally conscious of the high stakes of high school assessment, which made her seek objective criteria based on end products of student work. However, because NOS activities 'are not like experiments with precise results' and 'has no answer', she believed that NOS assessment could focus on participation and attitudes. Both Ms Nam and Ms You were partly drawing on the lower stakes associated with SIE due to the revised marking system, to support their view about assessing NOS. Overall, these differences point to the possible link between a teacher's science assessment view and NOS assessment view, which will need further investigation. Besides, while existing studies on the subject-dependence of classroom assessment views looked at teachers of different subjects (Black et al., 2003; Marso & Pigge, 1993; McMillan, 2001; see Section 0), the findings of this study provide evidence that even a single science teacher's assessment view can be multi-layered, depending on whether the views are about science or NOS.

5.2 Enactment of NOS summative assessments

The teachers used three tasks for NOS assessment, each of which followed a teacher lesson on the topic covered in the task. The general outline for each task was decided early in the semester, but the specific goals and content of the tasks to be included in the worksheets were decided near the enactment of each. Since the entire year group, of eleven classes, was expected to be assessed with the same set of activities and on the same scale, the worksheets served as a standard to give uniformity in assessment across the three teachers. Figure 5.1 describes the outline of lessons and assessment tasks implemented during the semester. How teachers worked together to design these lessons and tasks will be the interest of Chapter 6, so here I focus on describing in detail the three activities and worksheets, and how each teacher used these in their lessons to enact the NOS assessment.

5.2.1 Task 1: VNOS

VNOS is a widely used paper-and-pencil test for NOS knowledge developed by Lederman and colleagues (see Section 2.3.2). The teachers used VNOS-C (Abd-El-Khalick, 1998; Lederman et al., 2001; Appendix C) as the semester's first assessment task. The VNOS task was implemented in association with online video lectures highlighting different aspects of NOS. In the first half of the semester, the teachers created a series of lecture videos targeted at a number of NOS aspects. The contents of these online videos are outlined in Table 5.2 (see also Appendix C for example teaching materials), with the aspects of NOS that the teachers targeted as learning goals. Figure 5.2 shows an example slide from 'The life of Lavoisier' online lecture that was used after a short video clip about his life, where observation and inference are discussed in relation to Lavoisier's scientific activities.

These lectures were shared online via the EBS Online Class platform (https://oc.ebssw.kr/). Each week, a pre-recorded video (10–20 minutes each) was posted with a few reflection questions about the video content. Students watched the videos, wrote their answers to the questions and uploaded them on the system. These reflection questions were for self-assessment and were not assessed by the teachers. The system allowed the teachers to check each student's access to the video and submission of homework, so they could ensure that all students were keeping up with the lectures in an asynchronous learning setting.

After the online lessons, VNOS was implemented in class in an exam-like setting. Students sat in their place quietly and completed the VNOS in a limited time. Since the same set of online lessons had been delivered to all students, and the same questionnaire was used for all classes, the assessment task was enacted quite uniformly across the three teachers. The teachers briefly explained the purpose of the task and told the class

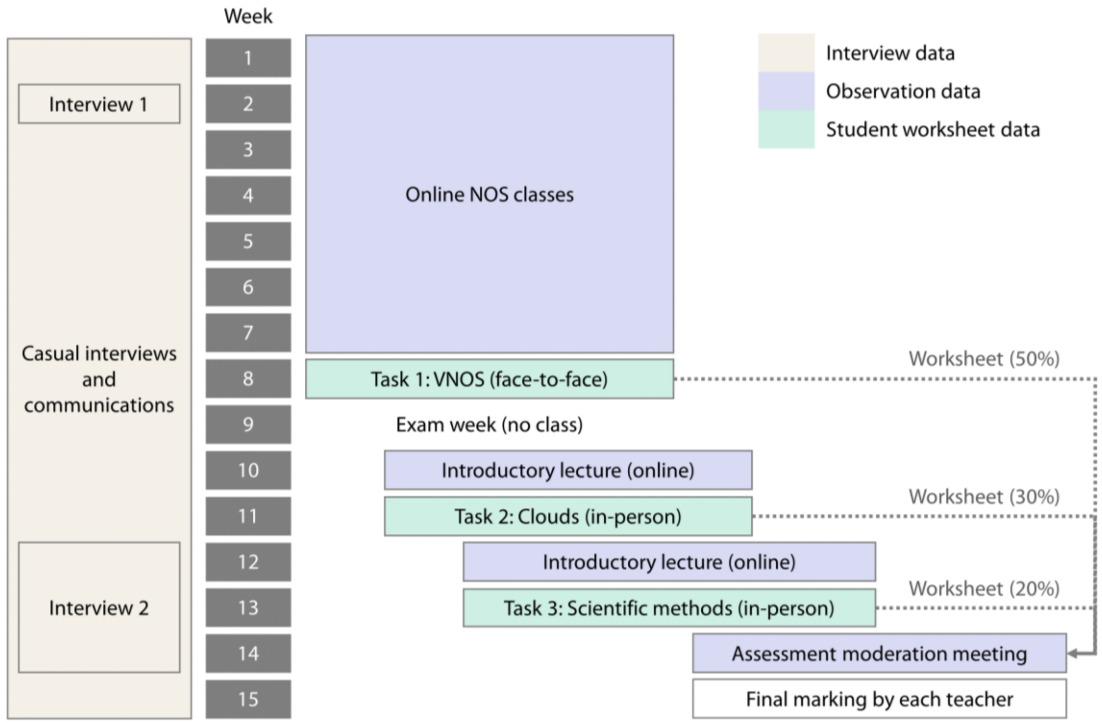


Figure 5.1 Outline of lessons and assessment tasks

Table 5.2 Summary of online lessons taught before the VNOS task

Week	Content	Featured NOS
2	The discovery of electricity by Galvani and Volta (I)	Observation and inference The evidence base of science Scepticism in science
3	The discovery of electricity by Galvani and Volta (II)	Change of scientific knowledge based on new evidence Replication in science Influence of technological advances on science
4	The life of Lavoisier	Change of scientific knowledge based on new evidence Explanation and prediction Observation and inference Subjective influences Sociocultural embeddedness
5	Cretaceous–Paleogene extinction event	Inductive methods
6	Louis Pasteur’s experiment on spontaneous generation	Deductive methods
7	The science of COVID-19	N/A

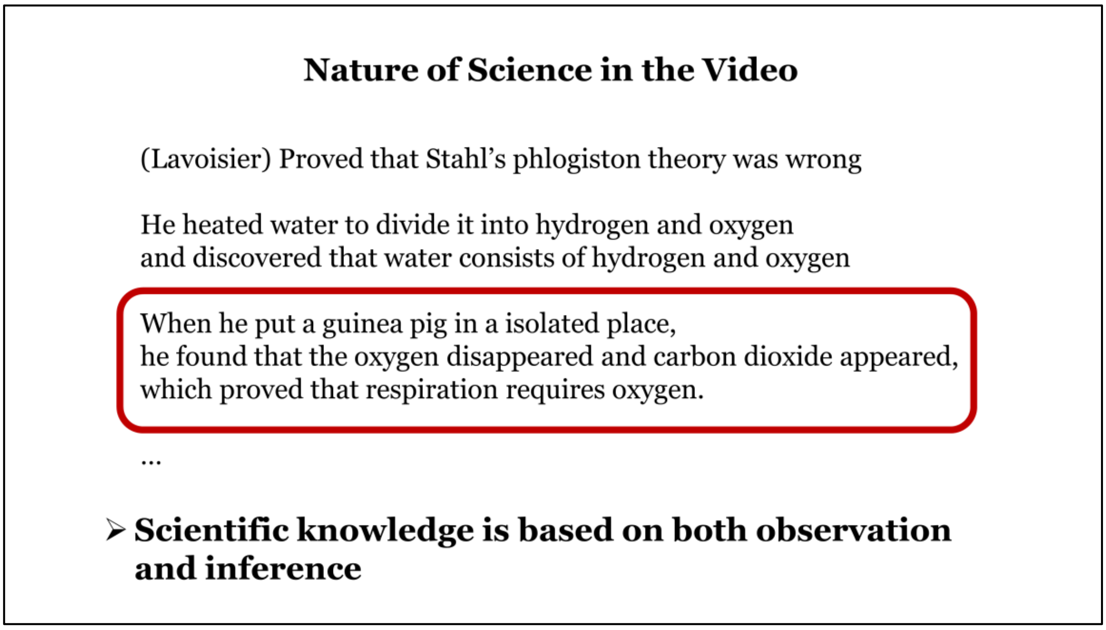


Figure 5.2 Slide from 'The life of Lavoisier' online video that explains the observation and inference in science (reproduced and translated from original). More slide excerpts are provided in Appendix C.

to write their ideas clearly as asked, before they gave students 40 minutes to fill out the VNOS. They refused to answer students' queries about the questions and instead told them to recall what they had learnt in the online lessons. One VNOS question, about a scientific law and a scientific theory, had not been introduced in the online lectures. When Ms Nam's students asked her for help, she just gave generic guidance by telling them to 'think if you know anything that ends with "law" or "theory" in science' (Interview, July 2020). Other than this episode, there was no notable variation in how each teacher enacted the assessment.

5.2.2 Task 2: Clouds

The NOS learning goal of the clouds activity was to understand how scientists use observation and classification to generate knowledge about nature. Students were engaged in the observation of the clouds around the school and asked to identify the types of clouds according to Luke Howard's classification system described in Figure 5.3. A brief 6-min introductory online video about clouds was provided prior to the implementation of this assessment task. In the video created by Dr Kim, she starts with

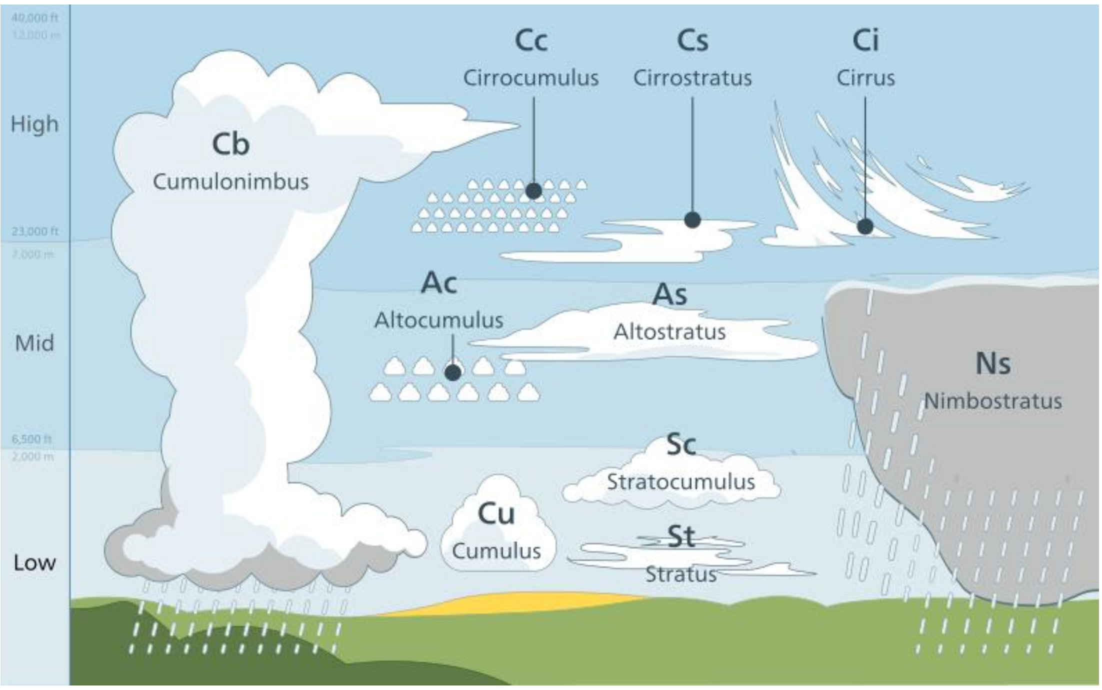


Figure 5.3 Cloud classification and nomenclature

(Image source: https://commons.wikimedia.org/wiki/File:Cloud_types_en.svg)

explaining observation and classification as examples of scientists' inquiry (Figure 5.4). After this generic account, the classification system for clouds based on their height and shape was introduced. She showed a number of cloud pictures and how each picture fits into one of the ten different clouds types. At the end of the video, she stated that 'I hope the day we meet will be a cloudy day'.

The activity worksheet was composed by Dr Kim, an earth science teacher. It consisted of a section to report observation in words and drawings, followed by a reflection question that asked students to find an example similar to the cloud classification in other areas of science. This activity was enacted by the three teachers differently, in several important respects.

Dr Kim brought her students outside for this activity. It was a cloudy day, with low, grey clouds covering all over the sky. At the beginning of the lesson, she said that they were lucky to have a lot of clouds in the sky that day. The student scattered around a certain area in the playground. While students were observing and sketching the clouds, Dr Kim walked around to see how the students were doing and answer questions. Students asked her many questions to fill out the worksheet, and Dr Kim mostly helped

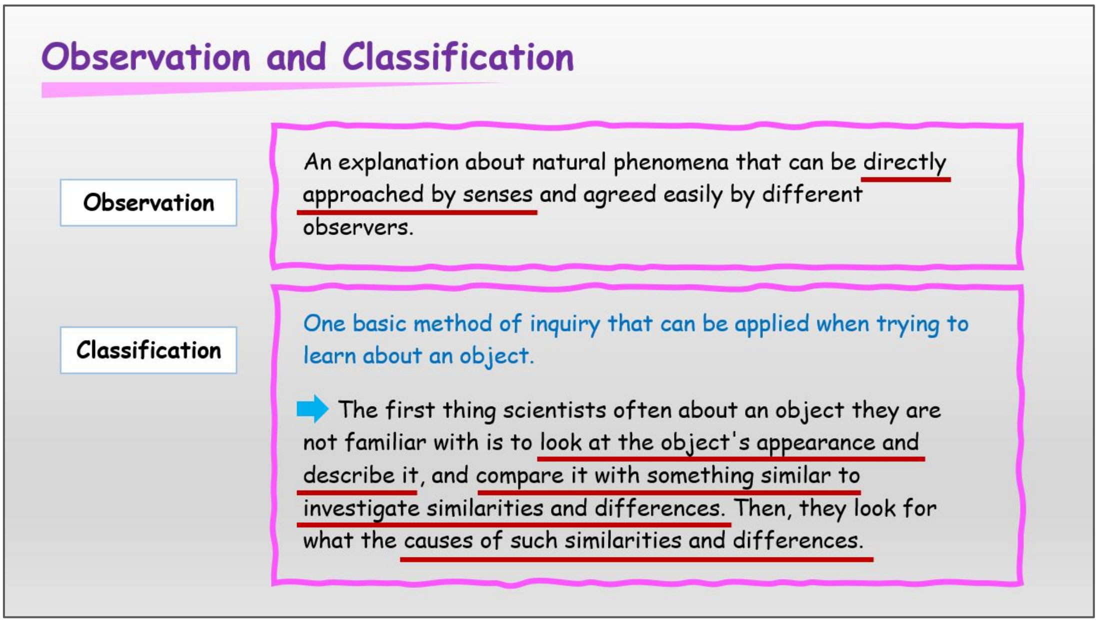


Figure 5.4 Slide from the clouds lesson explaining observation and classification (reproduced and translated from original)

them find answers by providing scaffolds. Since the students were scattered, her answers and feedback to students were only available to the handful of students who were near her. Her feedback was rich and formative, based on her subject expertise. For example, when one student asked whether the white part or the grey part of the sky was the clouds (see Figure 5.5), she answered,

Dr Kim: Both the white and grey parts are clouds. Why does it look grey? Because it blocks the sunlight much. It means the cloud is thick.

Student: If the cloud is thick, it's cumulonimbus, right?

Dr Kim: Yes, if the contrast is evident, a cumulonimbus is developing.

As such, she seized on students' questions to address more specific aspects of cloud classification and support students' reasoning. She also confirmed that the student's inference was correct. However, she refused to provide answers to 'direct' questions from students, such as 'Is that a cumulus?' or 'What is that cloud over there?'.

Dr Kim's approach as such remained consistent for the last worksheet question about finding a similar case where observation led to classification in science. This question was to assess whether a student understands classification as a scientific

Figure 5.5 The sky filled with clouds on the day of Dr Kim's clouds activity. Both white and grey parts can be observed in the clouds (picture by author)

practice that produces knowledge about nature. When students proposed an example, she asked why they think that was similar to how clouds were classified, sometimes giving them some examples. For example, when one student was asking if the discovery of electromagnetic induction could count because it relied on repeated observations,[9] she asked him to think if those observations amounted to classification so that the student could realise the difference between the two cases. Later in the interview, she said that giving hints to students was to prevent 'going to different directions, and finding answers in wrong places' (Interview, July 2020).

[9] The student was referring to Hans Christian Ørsted's discovery that an electric current creates circular magnetic fields around it. In comparison with the clouds activity, the electromagnetism example does have the observation element, but it lacks the establishment of a classification system as a result of the observations.

Similar to Dr Kim, Ms Nam brought her students outside for the task. She brought the students after spending 10 minutes in the classroom to review the online lesson about clouds and give instructions (Note that Dr Kim's lesson did not have this part but started outside straight away). In general, her feedback while students were working on the task stressed the process and reasoning more than getting the right answer. For example, when one student asked Ms Nam whether the cloud she was seeing was cumulus, below is how she responded:

Student 1: I've looked at this cloud, and I think it must be either of these two [types of clouds]. But it looked more floppy so I said it's cumulus.

Ms Nam: Right, right. **You can just write the reasons for your judgement.**

Student 1: Then is it a vertical cloud? Or a low cloud?

Ms Nam: Well … **You can just say a low cloud. You don't have to say it's a vertical cloud. As long as you wrote the position and the height.**

Student 2: I will say a vertical cloud.

Ms Nam: **It [low clouds] only means that it's down there.**

This exchange clearly suggests that Ms Nam's feedback was not focused on elaborating on student questions to create a 'teachable moment' for the science of clouds or NOS, but rather on the process skills such as inference and argumentation. Low clouds (e.g., cumulus) and vertically developing clouds (e.g., cumulonimbus) are two distinct categories, but instead of using Student 1's question to articulate their difference, Ms Nam implied that either way would be acceptable as long as the reasoning is sound. Then Student 2, who was also listening to the teacher, interjects to say that she would go for a vertical cloud, and Ms Nam did not object. This episode shows the different uses of feedback between Dr Cho's and Ms Nam's enactments.

When Ms You enacted the clouds task, it was drizzling outside so she decided to enact the task in the classroom, with limited teacher-student, student-student interaction. As with Ms Nam, she spent the first 10 minutes summarising the online lesson about observation and classification, after which she let her students observe the clouds through the classroom windows to fill in the worksheet. She also told the students that it was okay if they were poor at drawing and colouring, as long as their observations are well represented. She walked around the classroom during the task, answering some clarification questions about the task, but most of her answers did not go beyond repeats of her earlier instructions. When a student asked her a question, such as 'Which way is the north?', instead of telling the student privately, she talked to the whole class. Later she said that it was her way of making the task fair to all students (Assessment meeting, July 2020).

When she later found out that Dr Kim enacted this activity outside the classroom and actively gave her students feedback, she thought that it would have been nice if she had done that way, too:

> Ms You: … **it's my first time doing this subject** [SIE]. **So I probably cared more about assessing fairly than the other teachers,** just thinking about other subjects. If I do this outside … there will be more communication … and more feedback. But I really had students sit quietly, and didn't make any comments during the assessment. I regret that. (Interview, July 2020)

Ms You's regret can be associated with her inexperience in the subject, and her avoidance of risk in classroom assessment (Section 5.1.3.1) as someone new to high school assessment. Overall, there were some important variations among the three teachers in terms of the 10-minute introduction at the start of the lesson, location of the lesson, weather, and the level of guidance and scaffolding that the teacher provided in response to student questions.

5.2.3 Task 3: Scientific methods

The topic of the final NOS assessment task was scientific methods, particularly inductive, deductive and abductive methods. Like VNOS, this task was intended as a paper-and-pencil activity about NOS. The overall aim of the task was to assess if a student understands each scientific reasoning method and is capable of applying them to concrete examples in science. The implementation of this task was preceded by online lectures about deductive and inductive methods (Table 5.2). In the lectures, the two methods were introduced as major ways of scientific inquiry, and were each exemplified with historical examples: The Cretaceous–Paleogene extinction event as an example of the inductive method, and Pasteur's experiment about spontaneous generation as an example of the deductive method.

The task consisted of reading five stories about science in the worksheet, each followed by open-ended questions. For example, one question presented four different scientific discoveries and asked to identify which ones are inductive inquiries, and then explain their answer. Other questions asked students about specific processes in each scientific method, for example proposing a hypothesis about the observation or distinguishing between control and independent variables. Although the inductive and deductive methods had been defined and illustrated in the online videos before the enactment, the abductive method had not. The worksheet thus spent one page to define abductive reasoning as 'the process in which scientists infer a hypothesis that best explains the phenomenon in question', compare it with inductive and deductive methods, and exemplify abduction with how Kepler came up with the hypothesis of the elliptical orbit of the Mars based on observational data.

The teachers' enactments of the scientific methods task were close to those of the VNOS task, in a similar exam-like setting. The teachers were not actively answering students' questions about the task. Ms You, for example, said that she opened the activity by giving minimal guidance to students:

Ms You: I didn't talk much before I started the task, because … even if I had explained about the abduction [at the beginning], the kids wouldn't get the ideas immediately and answer the question. I said, 'Inductive and deductive methods are what we have learnt from the online lessons. On page 1, you can read and answer well to the questions … whereas, on page 2, it talks about a method called abduction, **you will read this carefully and answer the questions … but there is no one correct answer. You write what you think, and if that's logical, it's fine'.** (Interview, July 2020)

Here, we can see Ms You telling her students that the focus is on logic and reasoning rather than getting the answer 'right or wrong', which is reminiscent of how Ms Nam reacted to student questions during the clouds activity. For this last activity, the other two teachers' enactments were largely similar to Ms You's. Dr Kim just briefly told students that abduction is about 'finding patterns in the observations and making an explanation for those patterns' before starting the task. In other words, the teachers did not enact the task interactively, but still emphasised the open-ended nature of the task, by telling students to express their ideas in a logical way, based on what they had learnt and read about scientific methods.

5.3 Patterns in teachers' views and enactments of NOS summative assessment

The variation in how the teachers perceived NOS and its assessment and how they enacted the same set of assessment tasks sheds light on several important aspects of assessing NOS in classrooms. First, the degree of cross-teacher variation in enactment differed across the three assessment tasks. For the VNOS and scientific methods tasks, the teachers' enactments were largely similar in terms of time use and the amount of teacher feedback, partly because these two were paper-and-pencil tasks administered within the classroom. On the contrary, in the case of the clouds task, which was based on a student-led inquiry activity, a significant variation existed between the three teachers' enactments. The differing approaches to feedback during the cloud lesson were in part connected to each teacher's relative emphasis on facilitating learning versus ensuring fairness. Dr Kim's enactment prioritised the opportunities for learning that the task provides, so she actively used feedback to help students articulate their ideas about NOS (i.e., observation and classification) and content knowledge (i.e., clouds). In other words, she 'gave students lots of hints' so they 'don't have to seek answers in a wrong place' (Interview, July 2020). This stood in stark contrast to Ms You's enactment. Although Ms You thought that NOS activities could focus more on formative aspects such as exchanging feedback and encouraging participation, than marking-related concerns, in the actual enactment, she was very hesitant to give feedback or answering student questions, due to the fairness concerns.

Second, underlying such difference was the essential tension between formative and summative functions of classroom assessment (Black et al., 2010). The NOS tasks had the potential to serve both as an assessment and a learning experience, although the two were prioritised differently by the three teachers. From what has been discussed, one way to describe this difference would be that Dr Kim was primarily enacting *a* *lesson*, whereas Ms Nam and Ms You were enacting *an assessment*, during the clouds activity. Dr Kim was aware that the tasks would be marked for grades, but she still prioritised turning them into student learning about NOS and clouds; The other teachers did not exclude the learning element of the activity, but the overwhelming assessment pressure prevented them from actively helping students to develop ideas about NOS. In hindsight, an interesting contrast that signalled these differing enactments was in what the teachers had said to me before I attended their cloud activities: Dr Kim said that it would be ‘very exciting to watch’ for me; Ms You said that ‘there wouldn’t be much to see, since it’s an assessment’. (Field note, June 2020). Such a difference in enactment turned out to be more dramatic in the inquiry-based NOS assessment task (i.e., clouds) than in the paper-and-pencil tasks (i.e., VNOS and scientific methods).

Third, the teachers’ enactment of NOS assessment was not necessarily reflective of their perceptions of it, due to several reasons. The mismatch between teachers’ views and practices in classroom assessment has been reported constantly in the literature (Black et al., 2010; Noh et al., 2015; Rieg, 2007). The analysis of the three teachers in this study offers how such a mismatch can manifest in the context of teachers’ summative classroom assessment of NOS. The discordance was evident in the cases of Ms Nam and Ms You, who were unsure about summatively assessing NOS but whose enactments were still focused more on assessment than learning. On the contrary, Dr Kim recognised the possibility and value of assessing NOS summatively, but she was also active in providing scaffolds and turning the tasks into learning opportunities through on-the-fly assessment conversations with students. Despite such differences, Ms You and Ms Nam did wish that they had more time to exchange feedback with students to improve learning. After all three assessment tasks were completed, Ms You said that she could have enacted the VNOS and scientific methods tasks otherwise:

> Ms You: … If I had enough time, I would have shared on the screen some of the example answers that kids wrote in the worksheet, and assessed those answers together with them. If there’re any errors, how we can think about it, it would be nice if we could talk about these things together. But this year, kids just pump things out, no feedback is given at the moment. (Interview, July 2020)

Given that the class time was shorter than usual due to the COVID-19 reductions and some lessons had to be delivered online asynchronously, her reflection suggests the possibility that she could have provided more feedback and peer assessment through 'talking about things together' after the tasks had there not been such restrictions.

5.4 Chapter summary

Teachers' assessment practice is influenced by various factors including their personal conceptions of classroom assessment, the content being assessed, and the specific instructional and curricular settings in which the assessment is conducted (Brown, 2002; Wolf et al., 1991). Depending on these conditions, the same assessment task can be enacted in similar or different ways in classrooms. In this chapter, I explored how teachers' perceptions about NOS and its assessment aligned with their enactment of three NOS assessment tasks. Although the teachers used a common set of NOS activities for assessment, the findings suggested that teachers' enactment of NOS assessment is not uniform, nor is it simply a function of their views about NOS teaching. As will be illustrated in the next chapter, the classroom enactment of NOS tasks described in this chapter was principally a private practice; that is, how exactly the common tasks were enacted relied principally on each teacher's professional judgement. Given that this chapter has focused on the teachers at this individual level, a fuller understanding of their practices will require considering their summative NOS assessment as a collaborative effort, which will be addressed in Chapter 6. By investigating the teachers' concerns about the marking, as became evident in the assessment meeting at the end of the semester, the collective dimension of NOS assessment as a social process and what it reveals about the challenges of summative NOS assessment will be revealed.

Chapter 6 Collaborative summative assessment of NOS in a teacher team

6.1 Classroom assessment as a process of continual professional judgement

Classroom assessment, both summative and formative, is a complex event. As discussed earlier, recent perspectives on assessment consider it to occur throughout teachers' instructional activities, starting even before the lesson. The process of assessment begins as teachers plan their teaching and organise the learning experiences, particularly when designing the main activities that students are to complete (Klenowski & Wyatt-Smith, 2013). Anderson (2003) provided a good description of assessment as a continual process of gathering information and making decisions:

> Assessment is the process of gathering information to make informed decisions. Before anyone engages in assessment, he or she must know why the assessment is being made (the purpose), what information is needed to make the decision (the basis), when the information is needed (the timing) and how the information is best collected (the method). (p. xi)

These assessment decisions require a series of professional judgements of teachers. Allal (2013) compared teachers' assessment-related judgement to the clinical judgement of medical practitioners, in terms of connecting what they know about an individual (i.e., patient or student) with their professional knowledge and institutional norms, in order to produce the best course of action. As Allal and Mottier Lopez (2014) noted:

> Teacher judgement in assessment intervenes in a myriad of choices concerning the purpose of an assessment activity, the type of information to be collected, the method of collecting it, the sort of analysis and interpretation that is appropriate, the type of student involvement in assessment, the way of combining information and communicating outcomes, the consequences for the regulation of teaching and learning and/or for the fulfilment of reporting obligations and the appraisal of anticipated benefits and risks of decisions based on assessment. (p. 151)

Given the centrality of teacher decision making in summative assessment, looking into the process of teacher decision making can provide useful insights for supporting teachers' NOS summative assessment. Guided by the research question 'How does a team of three science teachers engage in NOS summative assessment in a school? What does their decision-making process reveal about the potential issues in the summative assessment of NOS?', the objectives of this chapter are to identify (a) the aims and objects of teacher collaboration for NOS summative assessment, (b) the process in which such collaboration occurred, and (c) the challenges of NOS summative assessment as revealed during the collaboration.

6.2 Teacher collaboration for summative classroom assessment

Decisions about summative assessment are often made by a single teacher, but they can also involve more than one teacher for several reasons. Common examples include when teachers co-teach a class, or when they teach the same subject to different classes in a year group. When such school-based teacher collaboration is required, teachers' assessment practices and decision making are *de-privatised* (Vanblaere & Devos, 2016). This makes an exception to the fact that teachers' professional practices are mostly private because 'the actual influences on and bases for judgement typically remain private' (Klenowski & Wyatt-Smith, 2013, pp. 12–13). In such a collaborative setting, teachers become engaged in a *community of practice*, which is 'a group of people who share a common concern, a set of problems, or an interest in a topic and who come together to fulfil both individual and group goals' (Wenger-Trayner & Wenger-Trayner, 2015). Viewed as a community of practice, participation in teacher assessment teams elicits continuously negotiating meanings (Wenger, 1998), for example, of the aims, methods, and fairness of assessment (Allal & Mottier Lopez, 2014). Assessment research has suggested that collaboration can have a positive impact both on the quality of assessment such as validity and consistency, and teachers' assessment skills (Black et al., 2010; Klenowski & Wyatt-Smith, 2014).

In NOS research, several studies have explored teachers' collaboration to teach NOS, most notably those of Valarie Akerson and her colleagues. In Akerson, Donnelly, Riggs and Eastwood's (2012) study, five preservice elementary teachers met with researchers regularly during their internships as a 'NOS community'. This led the preservice teachers to develop a shared repertoire of NOS teaching strategies, by sharing and discussing ideas, resources and practices. Similarly, Akerson, Cullen and Hanson (2009) also formed a community of practice consisting of 17 elementary teachers to improve their NOS views and teaching practices. They found that the community helped the teachers change their knowledge and practices by discussing the struggles within the community. Similarly, Bilican, Akerson and Nargund (2020) showed that a co-teaching experience with a university teacher can be an effective means to develop the NOS knowledge and pedagogy of an elementary teacher without NOS teaching experience. These studies provide initial evidence that teacher collaboration has the potential to improve classroom practices in NOS, but the actual process of collaboration on classroom assessment of NOS remains to be explored.

When viewed as a sociocultural activity, one key element of collaborative assessment is what Klenowski and Wyatt-Smith (2013) called social moderation of assessment. Social moderation is 'a process involving teachers in discussion and debate about their interpretations of the quality of assessed work' (Klenowski & Wyatt-Smith, p. 74). In social moderation meetings, teachers examine samples of the student work with an aim to 'reach shared judgements that are aligned with institutional references, such as curriculum objectives, standards of learning outcomes and criteria for assessing student work' (Allal & Mottier Lopez, 2014, p. 153). This process helps teachers develop a common understanding of standards and assure consistency and form a shared recognition of performances that correspond to those standards, which collectively contribute to the reliability and validity of marking judgements (Adie, Klenowski & Wyatt-Smith, 2012). Adie, Klenowski & Wyatt-Smith (2012) well illustrated how the interactions between teachers can collectively develop teachers' interpretation and use of standards through their participation in the moderation meeting:

> … prior to a moderation meeting, the teacher has interacted with the assessment task and the standards to develop an understanding of what denotes evidence of a particular standard. This event is connected to other historical practices and interactions in which the teacher has taken part that has informed a particular interpretation and use of the standards. Once this teacher meets with other teachers in the moderation meeting, this interpretation and use of the standards may be challenged or confirmed. New understandings develop through the negotiation of the standards with the other teachers involved in the meeting. (p. 225)

This account highlights the crucial role of moderation meetings that mediate multiple teachers' personal views and practices into a common understanding and assessment decisions based on it. In this chapter, I utilise the concepts of collaborative summative assessment and social moderation to investigate how the three teachers, in the second year of the study, planned, enacted and reflected on NOS assessment. Compared to assessing science, assessing NOS can involve many additional complexities, given the topic-specificity of assessment (Section 0). Moreover, as discussed in the earlier chapters, NOS knowledge is distinct from scientific content knowledge in several important ways. Assessment of NOS thus hinges on what it means to understand NOS, how to plan and design assessment tools that can best capture such understanding, and how to connect the assessment with students' meaningful NOS learning experiences (Hanuscin et al., 2020). In this view, a close investigation into teachers' decision-making process can reveal many subtle issues around assessing NOS specifically. When a teacher assesses NOS alone, her practices are likely to remain private; It is, therefore, possible that issues and challenges are kept to herself and not deliberated. However, when NOS is assessed by a group of teachers, the issues can be revealed more vividly as they are deprivatised through the process of social moderation that involves multiple teachers with a shared goal.

In the following, I first explain the broad and specific contexts in which the three teachers engaged in the collaborative assessment as a subject team and then describe the five steps of the teachers' summative assessment practices: (1) Outlining and communicating assessment plans, (2) planning assessment tasks, (3) enacting assessment tasks, (4) moderating marking, and (4) marking and grading (Section 6.3). I then provide an in-depth analysis of the end-of-semester assessment meeting and the decisions made in that meeting, which was to determine the end-of-term marks recorded in students' comprehensive school record (Section 6.4). As will be discussed, this moderation meeting revealed several crucial issues and challenges of the summative assessment of NOS, and many of these were not something that could be fully resolved with a single meeting.

6.3 Process of collaborative summative assessment of NOS

In Korean high schools, it is not unusual for teachers to plan their teaching and assessment as a team for a school subject. It is less common for advanced, elective science subjects such as Physics 1, Earth Science 1 and Chemistry 2, since these subjects are usually taught by one teacher due to the relatively small number of students taking them. On the contrary, such collective practices are more popular for compulsory subjects such as Integrated Science and SIE, according to recent surveys (Byun et al., 2019; Kwak, 2020).

When such collaboration is required, teachers set up outlines for teaching and assessment in a recurrent meeting called a 'subject team conference'. Semester plans are drafted and made public through the School Information System operated by the Ministry of Education. Teachers decide on the activities that will be aggregate for each student's end-of-term grade, although the specific content of the assessment such as the written test questions and tasks for performance assessment is usually decided at later stages. Such moderation was a routine practice for the three teachers in this study, whenever multiple teachers engage in teaching the same subject simultaneously. A conversation with Ms You in the following shows to what extent the teachers in the school used moderation for marking student works:

Wonyong: … When you mark the open-ended and essay questions, do teachers individually do it? Do they mark the students they are teaching?

Ms You: Right.

Wonyong: Then do you usually discuss the criteria ahead of time?

Ms You: Yes. Although we roughly set the criteria, after doing [marking] them, we meet again to revise … For physics, the number of students [taking the subject] is small so I mark them all. Physics 1, and Physics 2 alike, I do them alone. For Integrated Science, last year when three of us taught the subject, we had a meeting before the tasks, and then after the task, each of us did some markings and talk about how many marks for which item.

Wonyong: Don't you decide if, like, each teacher can give some percentage [of As] and some percentage [of Bs and so on]?

Ms You: We didn't' do that.

Table 6.1 Steps in the collaborative assessment of NOS

Steps	Description	Mode
Outlining and communicating assessment plans	Making a list of main NOS assessment tasks Decide how the marks will be aggregated for the final grade Communicate assessment plans to students	Collective
Planning assessment tasks	Design the NOS assessment task (worksheet, instruction, etc.) Design the lesson that precedes and/or follows the assessment task	Mostly private
Enacting assessment tasks	Implementing the NOS assessment task in the classroom to collect information	Private
Moderating marking	Discussing detailed marking criteria based on examples of student work	Collective
Marking and grading	Awarding marks to individual student work and aggregating to decide end-of-term grades	Private

Wonyong: Then it's still possible that one teacher gives better marks, and another teacher doesn't?

Ms You: Yes. So there was a huge difference between classes last year, but the criteria are the same. (Interview, March 2020)

Ms Nam felt that in her school, more emphasis was put on moderation compared to other schools where she had taught and that teachers were conscious about the issue of comparability. This was connected to ensuring fairness and minimising the potential of grade appeals (Interview, March 2020). When the three teachers were asked by the school to teach SIE to the Grade 10 cohort, they followed this usual practice to collaborate for planning, enacting and marking the assessments. Table 6.1 summarises the steps in which the team planned and conducted NOS assessment throughout the semester, constructed from the assessment and lesson plans, interviews and observations. The team's activities were broadly grouped into five steps based on their purposes, and each step was characterised in terms of the degree to which the assessment-related decisions were made collectively or privately.

Allal and Mottier Lopez (2014) argued that teacher collaboration for summative assessment can be classified with regard to and the aims of collaboration and the objects of collaboration. The collaboration reflected many levels of objects and aims, although they were not given equal significance. In their framework, teacher collaboration in summative assessment can have several aims. First, teachers may collaborate to increase the **consistency** and **comparability** of assessment through the use of common assessment procedures or well-specified standards. Second, collaboration can contribute to the quality of summative assessment by simultaneously increasing its **validity** and enhancing **students' learning experiences**. Third, teacher collaboration can increase the **coherence** of procedures in summative assessment, in terms of the format of the assessments and the methods for aggregation. Fourth, it can ensure the transparency of summative assessment practices and judgements.

They also identified four objects on which collaboration can focus. The first object is the **teachers' conceptions of summative assessment and related policies**. The second object is **the construction and selection of assessment methods and the use of them** in the classroom. The third object concerns **the assignment of grades** to each task that students completed. Finally, the fourth object is **the determination of grades** and other information to be recorded in student report cards and other official documents (Allal & Mottier Lopez, 2014). In the following, I describe the steps of collaborative summative assessment over the semester, with a focus on the aims and objects of collaboration that stood out in each step.

6.3.1 Outlining and communicating assessment plans

An outline of the three assessment tasks and how they would aggregate towards students' end-of-term grades were agreed upon early in the semester. The teachers put together a draft semester plan for teaching and assessment early in the semester.[10]

---

[10] The plan for each subject is required to be disclosed via the School information System (https://www.schoolinfo.go.kr/), following the Act on Special Cases Concerning the Disclosure of Information by Education-Related Institutions. Hence, teachers need to upload the draft plans with the week-by-week contents of the curriculum and assessment tasks (including how they aggregate into the final grade) early in the semester.

During this process, the team's decision on the selection of assessment tasks were principally based on the learning experiences that each task can afford, rather than the ease of marking for the teachers themselves, as illustrated by Dr Kim's statement:

Dr Kim: I think **we really didn't centre on the assessment. In fact, we didn't just use things [activities] that are good for assessment. We focused much on the fact that kids do something, in our school. Not that 'It's too hard to assess so let's not do it', but that 'It would be nice if kids did this'.**

Wonyong: So learning is before assessment. You decided on the content first and think about the assessment later?

Dr Kim: Yes. That's why we decide on assessment criteria afterwards. Although we did have some big picture at the start. (Interview, July 2020)

Her response shows that this first stage of collaboration served the aim of improving learning through summative assessment. This practice of 'big picture – enactment – moderation of criteria' was an unusual approach compared to the teachers' usual practice with other science subjects. The teachers said that for typical science subjects, which are mainly assessed through multiple-choice tests, they spend a great deal of time selecting the questions and choices so that the test can discriminate among students based on the Stanine scale (Interview with Dr Kim, June 2020). Instead, when assessing NOS, the teachers had significantly fewer concerns about accountability at the outset, as there was no obligation of norm-referenced assessment.[11]

Although further modifications to the initial plan were expected to suit the changing COVID-19 circumstances, the theme of each activity had been determined by this point and then announced to students via a 4-min video clip uploaded on the platform (Lesson video, June 2020). In this video created by Ms You and published to the entire year group, she explained that the three assessment tasks would be included for the marks, along with some details about dealing with student absence. Ms You not only explained the assessment plan but also justified some aspects of the plan. For

[11] This policy expectedly resulted in grade inflation in SIE in many schools, but it was not particularly problematic because most other subjects continued to be graded in the traditional (i.e., Stanine) norm-referenced system.

example, while explaining the first assessment task (VNOS), she said: ‘In fact, in this subject, we should be actually doing group projects and also do some labs … but we now have a guideline asking to reduce group activities because of COVID-19 and also have fewer practicals. So inevitably we’re doing an individualised activity’ (Lesson video, June 2020). This process shows an example of the teachers’ effort to make assessment transparent by using a video clip to communicate summative assessment criteria (Stiggins, Frisbie & Griswold, 1989). However, this video provided only an outline of the tasks rather than the criteria for marking each task that were not fully decided until the assessment moderation meeting at the end of the semester.

6.3.2 Planning the assessment tasks

Once the assessment outline was set and communicated to students, the teachers began to plan the content of each assessment task. This mainly included making the activity worksheet and planning the instruction that would precede or follow the activity. One key challenge, especially during that semester, was to plan instruction and assessment for the alternating online and in-person lessons during the pandemic. Considering this unusual format, the team decided that they would first provide an asynchronous online video to provide the background to the task and use the following in-person lesson for the assessment activity.

Common to all three tasks, the teachers worked largely independently of each other in the lesson planning and assessment design. Their collective decisions were limited to the issues of who would be in charge of which activity and what would be the focus of that activity. Further details were to be decided by the one teacher in charge, and no feedback was exchanged about the content of the lesson or the worksheet. Making introductory online lessons for the VNOS task were divided among the three teachers: Dr Kim produced three videos, Ms You two, and Ms Nam one video. In the case of VNOS, the team focused on addressing various NOS aspects in the lessons and finding good examples that can contextualise those NOS aspects. Since the task used an existing questionnaire, they had fewer concerns about designing the assessment itself. In contrast, for the clouds and scientific methods activities, the design of the learning experiences and assessments occurred iteratively. Nevertheless, the teachers still worked independently in preparing the lessons and worksheets. This shows the complexity in the degree to which the summative assessment was collective.

The task design process also was subject to the pandemic effect to a certain extent. Overall, two major impacts of COVID-19 on this stage stood out. First, since the changing pandemic situation increased teachers' workload, and much of what the teachers did the previous year could not be recycled in the hybrid teaching setting, the teachers felt that the whole process was rushed than usual. This issue of the workload at least in part explains the lack of feedback on the lesson and worksheet design between teachers. Second, the hybrid (of online and in-person) classes significantly reduced the class time that teachers could use for assessment. Their selection and design of the tasks thus were focused on finding 'something that could be done within an hour'[12] (Informal conversation, June 2020). Due to this time constraint, two of the activities, VNOS and scientific methods, were designed as paper-and-pencil tasks.

6.3.3 Enactment of assessment tasks

When the lesson and assessment were ready, the teacher who was in charge of that session recorded a video and posted it on the online platform so that students could watch it. In the following in-person teaching week, the 45-minute class time was used for students to work on the assessment task and fill in the worksheet individually. Each of the three teachers' enactment of the three NOS assessment tasks is described in detail in Chapter 5. As discussed there, the enactment was mostly a private practice, and as a consequence, the teachers did not enact the tasks in the same way. Similar to the previous step, this enactment step did not involve much collaboration among the teachers. The aims of each activity and the assessment criteria had not been fully developed before they were enacted in the classroom but only became clearer when the team met with completed student worksheets for moderation.

---

[12] She meant a 'class hour', which had been reduced from the usual 50 minutes down to 45 minutes due to the pandemic. In the previous year when there were no restrictions on class time, the teachers used to spend more than one class hour for an activity, as in the example of the paradigm shift activity that will be described in Chapter 7.

6.3.4 Moderation of marking: Defining and bounding acceptable answers

After all the classes were finished, the three teachers sat together to discuss how to mark the worksheets they had collected from students and how to convert these marks into end-of-term grades. Their eventual task was to convert the worksheets collected from each student to a final letter grade in A, B or C. Such moderation meetings were frequent in subjects that included activity-based assessment (as opposed to written examinations) as a major element. Ms You explained that it was common to have two assessment meetings: a pre-marking meeting to agree on the focus and criteria of assessment, and a post-marking meeting to develop detailed marking criteria after each teacher conducts a preliminary round of marking (Interview, March 2020). At the time of this study, however, only the pre-marking meeting was called due to the time limit.

One key function of collaborative summative assessment and social moderation is to set and delineate marking criteria (Klenowski & Wyatt-Smith, 2014). The team's moderation for marking occurred broadly in two steps. The first, relatively simple, process was to decide the distribution of marks across the questions in the worksheets. For example, since the assessment plan stated that 50 marks would be given to the VNOS task as a whole (Figure 5.1), the teachers had to decide how those 50 marks would be distributed within the task. In the meeting, Ms You pointed out that sometimes students did not 'fully' answer the question in the worksheet; For example, as for the question '*Describe what science is and how it is different from other subjects*', some students only answered the first half of the question but not the latter. The teachers then agreed that the completeness of student answers could provide one clear criterion for marking.

The second step of the moderation meeting was to define the 'correct' answers and 'acceptable' answers to each question in the tasks. Unlike judging the completeness of an answer, making differentiation within the completed answers involved more complicated procedures. For each assessment task, the teachers were faced with challenges that made it difficult to, borrowing Dr Kim's words, evaluate the 'quality' of individual student answers:

1	Dr Kim	… They wrote their own thoughts. **There is the right answer and there're reasons for that, but I think we shouldn't be marking based on that … If they wrote something, we perhaps shouldn't**

say it's wrong. We should give either a full or half mark. Then, if [the question] required two things and they wrote two, they get a full mark. And if wrote only half [*sic*: She meant one], they get a half mark … **This will make it quite discriminating.**

The reason why the teachers decided to avoid the 'right or wrong' judgement will be illustrated in Section 6.4, but aside from the reasons why, it is important to note from the above quote that the problem of marking was reduced to merely counting the number of completed questions. As becomes clear in her remark, discrimination was mainly achieved by checking whether a student answered everything rather than the level of understanding that their answer represents.

Although the team managed to circumvent the quality evaluation problem by focusing on completeness, problems remained as to how to deal with 'complete but absolutely wrong' answers. Ms Nam defined 'really wrong' to be irrelevant to and out of the question's context (Interview, July 2020). For the VNOS question about the difference between a scientific theory and a law, she thought it deserves no mark if the answer concerns neither a theory nor a law:

Wonyong: **Really wrong** ones, what kind of things [answers] are **really wrong?**

Ms Nam: For example, what did I have … I guess it was students who wrote **something unrelated to the content.**

Wonyong: So it's not about unclear answers, but answers that don't answer the question at all?

Ms Nam: It's not like they didn't answer. **They answered, but something that has nothing to do with a theory or a law.** Just those with answers that don't match the question. (Interview, July 2020)

In another example, the teachers had to decide the criteria for the final question about finding a similar historical example in the clouds task worksheet (*Can you think of an example where the result of observation led to the discovery of classification? Explain your answer*). In the assessment meeting, Ms You brought up what would be an 'absolutely wrong' answer:

1 Ms You: … So for this question, **if it's absolutely wrong** then no mark, but if it's hard to judge then give a mark, should we do it like that?

2 Dr Kim: Right. Absolutely wrong examples would be such as, if it [student answer] is completely about experimentation. Experimentation of course involves some observation, but since the question said we get some patterns. **Observation and classification**. It needs to have some classification, if it's off that line too much, then no marks. **When a kid asks [to appeal], we can surely say 'that's not about classification'.**

In Dr Kim's remark, she elaborates that the standard to discern 'absolutely wrong' answers would be whether the answer represents both observation and classification (which were the main learning goal), and such a strategy would lower the risk of student appeals against the grades. This conversation again shows that the focus of marking was moving towards giving a full mark unless the answer is incomplete or absolutely incorrect.

However, this agreed criterion turned out to be far less sufficient for each teacher to mark the hundreds of student worksheets. When I talked about this later with Ms Nam, she gave an example of the answers to which she gave a half mark. She gave a half mark to those students who wrote a *non-historical* example of classification (e.g., an everyday example), since the question asked specifically to name a *historical* example (Interview, July 2020). Similar to this, in marking the answers, the teachers tended to rely on criteria that they considered being straightforward and inferred directly from the question itself, but other than the 'observation and classification' criteria discussed in the moderation meeting, these personal criteria were not shared or deliberated with the other teachers.

6.3.5 Summary

Throughout the process of summative NOS assessment thus far, it should be noted that the degree of collaboration differed significantly across the four steps. Although the outlining and moderating processes were collaborative, the enactment of assessment activities and marking remained mostly solitary and private. Accordingly, different aims and objects of collaborative summative assessment were found in each step. In the outlining step, the teachers' collaboration was principally aimed at selecting NOS assessment tasks that could maximise student learning, rather than at assessment-related issues such as discrimination and the ease of grading. This step also involved developing a shared understanding of assessment, but this was focused on dealing with the fast-changing assessment policies during the early phase of the coronavirus, while substantive discussions around NOS-specific assessment concerns were limited.

Since the middle steps of designing and enacting assessments were independently done, their collaboration served a minimal role in these processes. The lack of exchanges in these steps later caused some issues in terms of comparability and fairness across teachers, particularly with the clouds activity where the teachers' enactments differed significantly (see Section 6.4.2 below).

The object of the end-of-term moderation meeting was primarily to determine the term grades based on collected worksheets, and the meeting served the aim of increased consistency and comparability of marking among the three teachers, to a certain extent. The main outcomes of the meeting were several criteria to discriminate among student answers and to prevent student appeals, but since the teachers were interested more in counting completed answers than assessing the quality of each answer, the discrimination was only coarsely made. In the following, I look more closely into the specific challenges of the summative assessment of NOS that unfolded in each of the three NOS assessment activities used during the semester, by analysing the teacher team's conversation during the moderation meeting.

6.4 Challenges of NOS summative assessment revealed during the social moderation

6.4.1 VNOS task: Out-of-rubric student answers, and the question of 'informed views vs more marks'

Of the three NOS assessment tasks, VNOS was the only task that had already existed, been tested for validity, and came with a scoring rubric (Lederman et al., 2002). Dr Kim had shared to the team a rubric with illustrative example responses representing informed and naïve NOS for each of the questions, before they met for moderation. Such ready-to-use rubrics provided to teachers are typically believed to facilitate moderation, and eventually increase the reliability of assessment, by 'making explicit the match between the evidence of the qualities in the work and the standards' (Klenowski & Wyatt-Smith, 2014, p. 78). However, in the team's moderation meeting, it was soon agreed that the rubric was not very effective or applicable for marking student worksheets. In the moderation meeting, Ms You started off by saying that 'the level classification [for naïve, mixed and informed NOS views] in the rubric is sort of different from what kids wrote'. Ms Nam, then, pointed out that conditions for perfect marks in each question should be discussed. Dr Kim said that even for her students, who she found were all hardworking and brilliant, not all of them wrote very good answers. Along with this dissatisfaction, the team also noticed that many students had left some questions unanswered. Moreover, there were many instances where questions were only partially answered; for example, some students only answered the first part in a two-part question.

Another issue with VNOS became apparent through the discussion about one student's answer from Ms You's class. Some of Ms You's students had responded that there is no difference between scientific theories and scientific laws, which was brought up during the moderation. After referring to these answers, Ms You said:

1	Ms You:	… These are individuals' thoughts. So as long as they wrote to some extent, a circle [full mark]. And those who said there's no difference between a scientific theory and a scientific law. Does it get no mark?
2	Dr Kim:	It gets no mark.
3	Ms You:	Just say it's no mark? I did say no mark for now. (chuckles)
4	Dr Kim:	Yes. VNOS is made not for this kind of purpose, anyway.
5	Ms You:	Because some kids said no [difference] and explained some reasons for that. About these … (Assessment meeting, July 2020)

The beginning part of Line 1 reveals how Ms You perceives the nature of VNOS questions: They concern 'individuals' thoughts'. Given such an interpretation of the questions, she was unsure how to mark the answers that should be categorised as naïve NOS views according to the rubric (i.e., no difference between theories and laws) but still did provide reasons for their answers. Here are some such answers:

> I think there is no difference. Because a theory can become a law, and a law can become a theory. (Student answer)

> I think there is no difference between a scientific theory and a scientific law. A theory that 'Gravity is directed towards Earth's inside' ends up becoming the law of gravity, so I think laws and theories are the same. (Student answer)

> I think people build on scientific theories to establish scientific laws. Therefore, I think laws can be included in scientific theories. If theories ⊂ laws[13], theories and laws cannot be always the same. In the future, we may observe some results that we haven't experienced yet, so I think they are not necessarily the same. (Student answer)

On the contrary, there existed many answers saying that they are different things, without corroborating why they are different, or on weak, incorrect grounds. Some examples included:

> I think there is a difference. First, a scientific theory is a confirmed thing. Fact. It talks about scientific truth, whereas a scientific theory is similar but has a scientist's **subjective** ideas more, so they are different in this respect. (Student answer)

> I guess there is a difference between a theory and a law. A theory may have some **errors**, but I think a law is the opposite. (Student answer)

> A scientific theory is a theoretical result of scientists' **experiments or research**, and a scientific law is just a basic and scientific thing without experiments or research. (Student answer)

Evidently, subjectivity, fallibility and experimental basis that these answers refer to are not the characteristics that distinguish a theory and a law, according to VNOS. Faced with these problematic answers, the team found the existing rubric unhelpful for deciding which of these responses would be given more marks. The simplified rubric could not discern every possible and complex answer to an open-ended question, nor did the teachers attempt to construct their own rubric from hundreds of responses.

Aside from the difficulty of marking VNOS, there was a sense of inappropriateness in using VNOS for summative assessment. Ms Nam characterised their agreed marking criteria as 'broad' and recognised that these criteria were not necessarily reflective of the VNOS rubrics. In the following conversation, she distinguishes between what she called the 'levels' of NOS understanding that individual students hold and the marks

---

[13] The actual student answer had a Venn Diagram describing this subset relationship.

they deserve from classroom assessment:

Ms Nam: So you can discriminate among the levels,[14] but I think it's really hard to turn those into marks. Assess? Although we can test it.

Wonyong: Is the problem with converting the levels into marks and using them for a student's school record?

Ms Nam: Yes. So the reason why we marked broadly … was since it could only be assessed in broad levels.

Wonyong: What are the concrete problems in converting those into marks?

Ms Nam: **When there are the low level and the high level, for example, we can't do it like, the low level gets fewer marks and the high level gets more marks.** So we just broadly treated them as the same [marks]. So that way it should work, but when we want to be more specific, that would be harder [to mark]. If we do the same thing next year, I would perhaps do it like this in a broad way.

Wonyong: What you mean by 'broad' is … to give a mark unless it's too wrong?

Ms Nam: **Right. Don't discriminate between low level and high level, but altogether, they will get full marks as long as they wrote the reason properly.** (Interview, July 2020)

In other words, Ms Nam's comments can be interpreted as prioritising reasoning 'skills' over NOS 'views' in summative assessment. This view coheres with Allchin's (2013) criticism that VNOS are aimed at characterising students' views rather than assessing students' knowledge, which calls into question its value as a classroom summative assessment. Along with Ms You's comment about 'individual's thoughts', Ms Nam's view here reveals that the experience of using VNOS for summative assessment led the teachers to recognise the problems with using a researcher-developed NOS assessment for summative classroom assessment (see Section 8.3 for further discussion).

---

[14] Here she is using 'levels' to refer to the naïve, mixed and informed views of NOS suggested in the VNOS rubric.

6.4.2 Clouds task: Fairness and different enactments of assessment

Student-led inquiry activities in the classroom are one common way to contextualise NOS instruction (Allchin et al., 2014). However, when a NOS task entails these activities, they are likely to add new complicating variables to assessment, since the enactment of activities itself hinges on many conditions to be done intendedly and uniformly. Such complications were observed during the team's discussion of the clouds activity during the moderation. While sharing the results from their preliminary marking, the team soon discovered that each teacher implemented the activity in different manners and in different physical conditions, which raised the question about fairness in assessment. In the following conversation, the teachers realise that they provided different levels of guidance to students during the activity:

1	Dr Kim:	… **I provided a lot of information about clouds during that [activity].** I gave answers to their questions. So **they could find a way to the answer.** Did you not say anything at all?
2	Ms Nam:	**No.**
3	Ms You:	**I almost didn't say anything**
4	Dr Kim:	[To Ms You] It rained on the day you did it, right?
5	Ms You:	Right. **So I couldn't do it outside and only did it in the classroom. I thought it would be unfair if I answered only to the kid who asked,** so [instead] explained again to the whole class. If someone asked, answered to the class.

As described in Section 5.2.2, Dr Kim frequently used formative feedback and scaffolds to students during the cloud activity to help them 'find the way to the answer', whereas the other two teachers were cautious to give feedback (Lines 1–3). Ms You states that it would have been 'unfair' if she had answered only to the student who asked a question (Line 5). Then Dr Kim said:

6	Dr Kim:	The [clouds] worksheet has lots of information, unlike VNOS. Also we had the PowerPoint slide with the picture on the screen [during the activity] … So wouldn't we have to count the wrong answers?

Here, Dr Kim emphasises that the worksheet itself had much information about the task available to students, so there is some degree of uniformity despite the different level of teacher help provided to students during the activity (Line 6). On this ground, she is suggesting that discrimination would still be possible. In the conversation that immediately followed, however, a related issue pertaining to fairness was brought up—that the three teachers enacted the task in different weather conditions.

7 Ms You: … It's hard to distinguish low and middle clouds.

8 Dr Kim: Right. But when it rains it's likely to be low clouds, as a matter of fact.

9 Ms Nam: So how do we to this? The cloud heights were various [when I enacted it].

10 Dr Kim: Were they?

11 Ms Nam: Yes. I had different kinds [of clouds].

12 Dr Kim: Ah, it was a clear day?

13 Ms Nam: **One day we did it in the classroom on a rainy day. Another day there were lots of types of clouds.**

14 Ms You: **For me, it was cloudy and drizzly, or about to drizzle, so mostly cumulus clouds when we observed.**

15 Ms Nam: It would be easy to mark, then.

16 Ms You: Right. Middle clouds and cumulus.

17 Dr Kim: **Yes. Then a kid who wrote high clouds and cirrus gets no mark.**

18 Ms You: Yes, I did say that [cirrus] was wrong, but I wasn't sure if it's wrong to say middle clouds. Because it's kind of not clear … [how to distinguish low and middle clouds with naked eyes] …

19 Dr Kim: Then we say middle and low clouds are correct answers, on rainy days? As a matter of fact, even we don't know for sure.

20 Ms You: I wasn't sure either.

21 Dr Kim: Okay, let's do that then. If there were indeed very high clouds, then it's correct [to say cirrus], and if it was raining but they said it's some cloud seen on clear days, then it's wrong. Since they don't understand it at all.

These questions about observing, drawing and identifying clouds were not to assess NOS understandings in themselves, but they included scientific practices from which the nature of observation and classification could be derived and learnt. Here again, the team is taking the strategy of 'deciding what is wrong' (in this case, saying cirrus on a rainy day) instead of 'deciding what is right' (Section 6.3.4). This way, all answers that are not 'wrong' can get a full mark, and consequently, the potential fairness issue from different weather conditions becomes less problematic, at the expense of finer discrimination.

Such an approach was observed again when the teachers were discussing another question at the end of the worksheet. Following the questions about drawing, describing and identifying the clouds that students observed, this last question asked for a specific example in the history of science where scientific observation led to the discovery of a classification system. After looking at the student answers in the collected worksheets, the teachers encountered some complicated problems in marking this question. Most frequently, students referred to the discovery of the periodic table. One student wrote that 'Scientists observed various elements and discovered that those with the same number of valence electrons have similar chemical properties'. The teachers agreed that this last question was important in that it concerns linking the earlier questions to NOS, but found it hard to decide whether to mark the periodic table example as right or wrong. Ms You mentioned that she was not sure if the periodicity of elements can be said to have been discovered *through observation or experimentation*.[15] When this issue was raised, instead of delving into it, the teachers quickly decided the criteria for 'clearly wrong' answers (see Section 6.3.4) and moved on to the next task. This suggests that the lack of knowledge about the philosophical aspect of science (i.e., observation and experimentation) can restrict teachers' decision making about summative assessment

[15] She did not elaborate why she was not unsure, nor did the other teachers add anything, but there are some possible reasons why she thought so. For observation, it can be problematic to say since the valence electrons were not observable when early periodic tables were proposed. For experimentation, the process in which the periodic table was discovered did not involve much of 'experiments' that typically entail manipulating and controlling variables. There is room for philosophical discussion about these issues, but it would have been impractical, if not impossible, to have such discussion in an assessment moderation meeting by teachers.

of NOS.

To summarise, for the clouds activity, the teachers had two major issues: one concerning the different enactments of the activity, and the other concerning the difficulty of judging the correctness of student answers. The former stemmed from the fact that this particular activity involved student-led, hands-on investigations, whereas the latter is essentially similar to the challenge they had with VNOS. The teachers handled these issues by defining wrong answers and allowing marks to grey-zone answers; in other words, by setting a generous marking scheme.

6.4.3 Scientific methods task: Reading literacy and content knowledge demands

Assessments should be presented in an understandable format, both linguistically and culturally (Lam, 1995). Meanwhile, reading scientific texts often poses additional challenges for students, due to their specific characteristics such as vocabulary demands, academic language, context-specific usages of words, and nominalisations of verb phrases (Osborne et al., 2016). When it comes to assessment, assessment tasks are 'saturated with literacy demands' although teachers are often not aware of them (Wyatt-Smith & Cumming, 2003). An extreme case of such literacy demand issue would be with students whose first language is not the language in which assessment is conducted, but literacy can still matter when students are assessed in their first language. In science instruction and assessment, the concept of literacy extends to include the ability to read not only texts, but also diverse modes of representation such as equations, tables, graphs, and figures (Krajcik & Sutherland, 2010). For these reasons, the literacy issue in assessment is closely associated with the fairness and validity of assessment tasks (McDermott, 1999; Stobart, 2008).

Along with reading literacy, many NOS assessments demand students to have prior knowledge of relevant content knowledge of science to varying degrees. For example, understanding the tentative NOS through the example of Newton's laws of motion cannot be adequately achieved without a reasonable degree of understanding of the laws themselves. Similarly, an understanding of how massive, transnational collaboration was crucial to the detection of gravitational waves can only be achieved when students have basic knowledge of general relativity and the characteristics of gravitational waves.

Such knowledge demand for NOS learning and assessment can be particularly problematic for certain categories of NOS knowledge, such as the nature of scientific methods and scientific knowledge. This is possibly one reason why textbooks often use very simple examples to illustrate NOS (Consider the classic 'black swan' anecdote used to teach about the inductive method) rather than authentic, complicated examples from science.

Throughout the three teachers' summative assessment process, literacy and content knowledge demands were the central problems in moderating assessment for the task about the inductive, deductive and abductive methods in science. Regarding literacy, it should be noted that the earlier two activities (VNOS and clouds) also required students' literacy for comprehending the questions. Still, this last assessment task was particularly demanding in terms of literacy in that it included reading a one-page article about abduction that students were not familiar with and answering questions based on the reading. The teachers said that the task was difficult to implement in the classroom.

1	Dr Kim:	… So the recent one [the scientific methods task] was quite difficult, wasn't it?
2	Ms Nam:	(Mimicking students in a high voice) **Miss, I have no idea what this means.**
3	Ms You:	**Like kids don't know the vocabulary.**
…		
8	Ms Nam:	This was perhaps too difficult. Particularly this Question 5, kids almost didn't understand it … They exploded. (chuckles) (mimicking students in a high voice) What on earth does this mean?
9	Dr Kim:	Then …
10	Ms Nam:	For Question 5, we allow them to think freely and write?
11	Ms You:	Unless they left it blank … (chuckles)

This episode sheds light on the centrality of literacy in text-based NOS assessment tasks. The teachers say that their students had difficulties comprehending the reading and the questions about abduction, which included many scientific jargons and historical accounts. Particularly problematic was Question 5, which was aimed at assessing whether a student (a) understands abduction from the earlier reading, (b) understands the reading here about Kepler's model, and then (c) can construct a hypothesis by combining (a) and (b). Later in the interview, Ms You raised a related, but slightly different, issue about the scientific methods task, relating to content knowledge:

Ms You: … For Questions 3, 4 and 5, I think **these would have been easier for students who had learnt them in advance.**

Wonyong: You mean, learnt what part?

Ms You: **For kids who learnt the science, like earth science and physics …** When a kid who knows Kepler's laws, and a kid who doesn't, get this question … [they would have different understandings of] what they're required to do. Kepler's laws were not taught in schools yet, but quite a few students learnt it [advanced science contents] from cram schools. Then they would know what Kepler's laws are, and that makes the problem easier**. In fact the question could be about summarising what they had heard of, instead of setting the hypothesis.**

Wonyong: In reality, it becomes not an assessment of NOS but an assessment of some other subjects.

Ms You: That's a possibility. (Interview, July 2020)

Ms You's remark points to a complicated problem in the NOS assessment. The assessment was to evaluate students' knowledge in abduction as a method of science and their ability to apply it to construct a hypothesis based on what they read. However, if a student already knows Kepler's laws, they may write the answer without necessarily understanding abduction, as Ms You suggests. Such a reduction of NOS task to content knowledge task, along with the earlier issue about literacy demand, made the teachers question the validity of their NOS assessments, but this issue was, to some extent, managed by broadening the range of 'acceptable' answers.

6.4.4 Assessment for learning improvement or assessment for accountability?

In the school-level summative assessment practice, the tension between assessment for improving learning and assessment for accountability requirements is commonly experienced by teachers (Adie, Addison & Lingard, 2021; Cumming et al. 2019). An assessment should be supportive of learning, but it should simultaneously produce valid and reliable evidence of student learning. In the context of this study, the lowered accountability pressure for SIE released the teachers from some of their usual practices in the highly competitive assessment culture of their school, such as double marking for open-ended questions (that is, each student work is assessed by two teachers) and fine-tuning of choices for multiple-choice questions (to eliminate possible alternative interpretations). The teachers took advantage of the low accountability pressure to prioritise learning over assessment and designed the assessment tasks that they thought could facilitate NOS learning best, rather than the ones that could be later marked easily by them. This suggests that the change in national and district-level assessment policies from the traditional Stanine system to the new system with fewer gradations contributed to the learning-focused summative assessment design for NOS.

While the low accountability pressure led the teachers to put the focus on student learning of NOS, this also meant that some limitations existed with validity, reliability and discrimination. These issues did not cause serious problems since the relative significance of SIE in the whole of the school curriculum was low; Most other subjects were still assessed in Stanine grades and sufficiently served the accountability purposes. At the end of the assessment meeting, Dr Kim expressed her discontent about how they decided to mark, by saying that 'We thought we used really good contents for assessment, but when grading … it's no different from doing a paper exam'. In a sense, such a reflection suggests that their attempt to use the three activities for assessing NOS was a partial success. In terms of the teachers' intention to use these summative assessments as meaningful learning experiences for NOS, there was some success particularly in Dr Kim's case who used frequent feedback, questioning, scaffolding and on-the-fly formative assessment during the enactment of the activities, although the amount of such interaction varied among the teachers. When it comes to the function of summative assessment to 'report the achievements of individuals and satisfy public demands for accountability' (p. 24), the NOS assessments were fairly limited as a result of the teachers' decision to mark generously instead of establishing detailed marking criteria that can discriminate among different levels of students' NOS understandings. As revealed from the teachers' disappointment about their own decision, they did not make such a decision because they considered it to be the best for the quality of assessment; it rather stemmed from pragmatic considerations that any decisions about classroom assessment should be practical in terms of the amount of time invested and the level of expertise needed for teachers (McComas, Clough & Nouri, 2020). Overall, the episodes presented in this section illustrate a tension between the quality of NOS assessment such as discrimination and the practicality of classroom implementation. The collaborative summative assessment, particularly the moderation process, contributed to bringing this issue to the teachers' attention, but the process played a limited role in balancing the quality and practicality of assessment.

6.5 Chapter Summary

This chapter investigated the three teachers' summative assessment of NOS over the semester as a collective practice with specific aims and objects of collaboration. I identified the steps of the teachers' collaborative assessment, the decisions made during the end-of-term moderation meeting, and the specific NOS-related challenges that emerged while the teachers moderated marking. It provided useful information about the potential issues of NOS summative assessment by teachers, when using a research instrument such as VNOS, when contextualising NOS within a student-led activity, and when NOS assessment involves reading about philosophical or historical issues. In so doing, the tension between the two functions of assessment, learning and accountability, was discussed. The challenges reported in this chapter can provide insights for future research and practice in NOS summative assessment, in both classroom and high-stakes examination contexts. Moreover, this chapter points to the benefit of collaborative summative assessment of NOS in bringing teachers' views and decisions into the public space. These findings about collaborative summative assessment experience have the potential to inform future professional development programmes to promote teachers' NOS summative assessment knowledge and skills.

Chapter 7 Instructional strategies that allow for the formative assessment of NOS

In the previous two chapters, I have illustrated how the three teachers in the school dealt with the assessment of NOS particularly focusing on its summative aspects such as accountability pressure, assessment criteria, grading, reliability, discriminating power and fairness. While these aspects offer insights into how the classroom assessment practice of NOS operates within the broader disciplinary, social, political and cultural contexts of education (Figure 2.4), understanding how assessment-related concerns can manifest *in the course of instruction* would further require an in-depth examination of actual lessons. As discussed earlier, evidence of student learning is not only collected and used at the end of a lesson, a unit or a semester to report to stakeholders, but these activities continuously occur while learning is occurring to inform the decisions made by the teacher and students minute by minute in the classroom (Section 2.2.2). Building on assessment theory as well as the literature on NOS instruction, this chapter sets out to shed light on such *formative* aspects of NOS assessment in the context of instruction. This is done through a close analysis of one lesson (over two 50-minute sessions) taught by Dr Kim in the first year of data collection. This chapter is guided by the research question: *What are the instructional strategies that allow for teachers' formative assessment NOS? What are some potential issues and challenges involved in the formative assessment of NOS?*

Using the lens of formative assessment practices to examine NOS instruction, I focus on how the teacher gathered information about students' NOS learning during daily whole-class classroom conversations. The teacher elicited information from students by giving a task, asking question and facilitating peer assessment, and recognised and used this information to give feedback and ask further questions to promote learning. What makes this lesson particularly interesting is that there arose various scientific, historical and philosophical topics across disciplines, which the teacher attempted to pull together to achieve the NOS learning goal. This aspect of the lesson allows shedding light on the role of the teacher's formative assessment practices in building connections between various elements of NOS and the challenges involved in such a process. In analysing the teacher's instructional strategies that allowed for formative assessment, I not only consider the content of classroom conversations but also the contextual factors such as the accountability system, classroom culture and students' prior learning in science for a holistic understanding of teacher practices.

In the following, I first provide some details on the teacher's backgrounds with regard to NOS and what the lesson looked like, and also who were the students attending the lesson. I then analyse the whole-class conversations that occurred during the lesson to characterise Dr Kim's instructional strategies to create opportunities for the ongoing and frequent assessment of NOS, with illustrative examples. In so doing, I consider both the affordances and challenges of the formative assessment of NOS that are observed in her lesson. In terms of the formality of formative assessment (Section 2.2.2), this chapter considers both formal and informal formative assessment by the teacher.[16] Formal formative assessment takes place when a teacher plans and prepares a specific set of tasks and/or questions that are targeted at the learning goal of the lesson in order to check for student understanding (Shavelson et al., 2008), and students produce some outputs that can be used for assessing and possibly grading (Bell, 2008). On the contrary, informal formative assessment is not wholly planned ahead of time but occurs spontaneously when the teacher notices a 'teachable moment' during classroom conversation and seizes the opportunity to support students in advancing learning (Nieminen, Hähkiöniemi, Viiri, 2020). In Dr Kim's lesson, she used a project-based assessment activity that was designed and planned to support student learning of NOS, which is an example of a formal formative assessment. The result of this activity was student presentations that were assessed by Dr Kim for both formative and summative assessment. Throughout the lesson, she also had to react spontaneously to the various issues that emerged or were raised by the students about the content of the lesson to transform these into learning opportunities, and also ask questions to check what students know and what they do not know. In this sense, her lesson included both formal and informal formative assessments that have the potential to improve learning by eliciting student ideas, responding to and acting upon the collected information

[16] Informal formative assessment is also referred to as interactive formative assessment (Bell & Cowie, 2001), assessment conversations (Duschi & Gitomer, 1997; Ruiz-Primo & Furtak, 2006) or on-the-fly formative assessment (Harrison et al., 2018; Shavelson et al., 2008).

(Ruiz-Primo & Furtak, 2007).

7.1 Introducing Dr Kim and her students

As an earth science teacher, she had become interested in the history of science, NOS and scientific methods during her mid-career stage. She recalled that, at some point in her career, the national curriculum started to require more student activities to be included in science lessons, and this made her think of inquiry- and NOS-related aspects in the earth science context and incorporate them into lessons. This was because she felt it to be very hard to do laboratory experiments in earth science which, by nature, deals with either historical or large-scale objects and phenomena, most of which cannot be brought to or replicated in classrooms. The increasing curricular emphasis on scientific methods, inquiry and student-oriented learning, however, led her to develop her own way of incorporating such change, one of which was the frequent use of history of science in her earth science lessons. For example, when she taught the cosmology part, she always referred to how Copernicus's heliocentric theory triggered the Scientific Revolution in the sixteenth and seventeenth centuries, and how it developed into small-scale to large-scale cosmological theories. She felt that her students really enjoy learning about earth science in a historical context such as this.

Although Dr Kim was certainly above average science teachers in terms of her knowledge and experience in NOS, she had not taught any NOS-oriented lessons as part of a formal school curriculum before she taught the paradigm shift lesson, because NOS had not been presented as a learning goal in the national curricula before SIE was introduced. In addition, since the paradigm shift lesson occurred early in the school year, and she had other duties in the school, she felt that she had not had sufficient time to prepare the lesson. For these reasons, she recalled that most of her practices in this lesson were more improvised than carefully planned ahead of time.

The students that Dr Kim teaches in this chapter were with high achievement and motivation relative to other classes in the same year cohort, across all subjects. This was a common evaluation shared by Dr Kim herself and her colleague Ms Lee. The school had a policy of dividing a year group into classes based on the elective subject a student chooses, and the student in the observed class took a subject that is usually selected by high-achieving students. In addition, since Dr Kim was the homeroom teacher for the observed class, she knew about her students' interests and achievement levels better than other teachers did. As a result, there was a friendly classroom atmosphere where students comfortably express their opinions. In such a classroom culture, peer feedback, criticism and emotional responses to each other's presentations and comments occurred naturally, as will be seen in the conversation excerpts throughout the chapter.

7.2 The paradigm lesson

Among the seven lesson sessions observed for the study, this chapter focuses on the two sessions taught by Dr Kim. At the beginning of the semester in the first year of research, Dr Kim enacted three sessions on the paradigm shift and the development of science. Understanding these concepts was stated in the national curriculum as an explicit learning goal, although it did not specify on a concrete level how it should be taught (MOE, 2015). Dr Kim came up with an activity about the paradigm shift (described below) as an assessment task, which was then agreed upon by the other teacher who taught the same subject that semester. The rationales for her choice of the paradigm shift activity will be presented later in this chapter (Section 7.3.1).

From a research point of view, I selected this lesson as a focal case to examine the formative assessment aspects of NOS instruction for at least two reasons. First, the lesson contained lots of information that was new to both students and the teacher, which required impromptu responses for both parties to respond to the presentation and keep the conversation going. All six groups presented on different topics ranging from Copernicus's heliocentric model to plate tectonics, many of which were not familiar to the students and sometimes even to the teacher, too (see Table 7.1). Second, the learning goal of the lesson was clearly pointed at NOS. In other words, Dr Kim had the aim of teaching about the paradigm shift and how it leads to scientific progress,[17] asked her

[17] Dr Kim developed this activity based on a learning expectation set out in the SIE curriculum, which stated 'Replicate the crucial experiments that motivated paradigm shifts in the history of

students to present on this aspect, and assessed whether they understood this aspect of NOS. These characteristics of the lesson seemed to provide an ideal context for investigating the enactment of her formative assessment practices specifically related to NOS.

Dr Kim taught the paradigm shift lesson over three 50-min sessions. The first session was a teacher-led lecture about paradigm shifts. She started the lesson by explaining that the term, paradigm, refers to a system of thinking that defines people's opinions and thoughts in a certain era. Then it was highlighted that while science sometimes develops incrementally and continuously, at other times the development occurs drastically following a shift of paradigm. She elaborated on this key NOS aspect by referring to the example of Galileo's thought experiment that came from the subject textbook. Students were asked to challenge Aristotle's old idea that heavier objects fall faster than lighter ones and to follow the steps of Galileo's thought experiment that refuted Aristotle's idea. Dr Kim then summarised the development of scientific ideas on motion from Aristotle to Galileo, and then highlighted three interrelated NOS aspects that could be drawn from the case: (1) There are critical experiments in the history of science that led to a paradigm shift; (2) Scientific development occurs through scientists' research and experimentation; and (3) An old theory can be replaced by a newly confirmed theory by repeated experimentation and verification.

After this introductory session came the two sessions that were observed and analysed for the current study. Students were divided into groups of four to find an example of the paradigm shift in history and prepare a 10-min PowerPoint presentation (Figure 7.1). A summary of each group's presentation is described in Table 7.1. After each group's presentation, Dr Kim initiated a whole-class discussion on the presentation and facilitated and orchestrated the assessment conversation. For the first few groups, she started the question or comments herself, but later she often let the presenters moderate the Q&A to encourage peer participation. Each round of discussion took between 10 and 18 minutes. Later she said that she could afford to spend enough

---

science and explain how science develops' (MOE, 2015, p. 114). This learning goal can mainly be related to the tentative NOS in the consensus list (Abd-El-Khalick, 2012) and the 'scientific knowledge' category in the FRA model (Erduran & Dagher, 2014a).

Figure 7.1 Group 1 presenting on the discovery of the double-helix structure of DNA. The portrait of biologist James Watson is shown on the slide.

time for discussions because of the less examination pressure with the subject and she could use class time more flexibly:

> Researcher: Then [how you teach] is not so much about the characteristics of the subject matter, as about assessment and other things …
>
> Dr Kim: Influence a lot. How it is assessed gives me freedom for teaching. The reason **I feel free to have long Q&As, is I don't have to cover all curriculum contents. It's not a CSAT [College Scholastic Ability Test] subject. There's no pressure like I'm not blamed only if I teach every unit and don't miss anything.** (Interview, April 2019)[18]

While the groups were giving their presentations, Dr Kim stood at the back of the classroom, taking notes of the presentations for feedback and grading purposes. In her notepad, she recorded summaries of the presentations, each group member's contributions, and how they dealt with the questions and comments during discussion. She wrote her general evaluation of each group as well as specific memos on individual students when necessary. In addition, she also observed and recorded whether the

[18] The data quoted in this chapter are from the first year of the study unless indicated otherwise.

students were actually presenting what they understood or merely reading the script without a good grasp of it, whether all group members contributed, the quality of the PowerPoint slides, how much effort was put in, and so on. She later used these notes to facilitate discussion during the Q&A, provide immediate feedback after the presentation, and also produce grades for students at the end of the semester (Interview, April 2019).

While the presentations were to be evaluated and graded for the summative assessment purpose, Dr Kim made it clear that she expected the whole process of this paradigm activity including the preparation, presentation and Q&A to be formative at the same time. She expected that through the engagement in the task and the discussion, students could reach an advanced understanding of not only the concept of the paradigm shift but also the historical and scientific aspects of it (Interview, July 2020). This was also evident in how she used the Q&A time to check for each group's understanding of the lesson goal and involve other classmates to ask questions and give real-time feedback which can be interpreted as informal formative assessment. As a facilitator and mediator of student learning, Dr Kim promoted student learning through various practices that allowed for ongoing formative assessment throughout her lesson, which are described in the following.

7.3 Instructional strategies that allowed for formative assessment during whole-class discussions on NOS

The analysis allowed identifying four top-level themes from Dr Kim's instructional strategies that allowed for the formative assessment of NOS throughout the lesson: multicontextualising NOS, mapping examples onto NOS, focusing the discussion on NOS, and exploring diverse NOS aspects. These strategies unfolded in the six Q&A sessions, sometimes initiated by Dr Kim's own moves (e.g., asking questions to the presenters, giving compliments) but at other times by the questions and comments from students, either voluntarily or after being called on by Dr Kim. In presenting the findings, I illustrate how these aspects are revealed in her interactions with the presenting students as well as the whole class, what kind of moves contributed to each of her strategies, and also describe what complexities and challenges of the teaching and formative assessment of NOS emerge from the episodes. Table 7.2 indicates the codes that appeared in each Q&A session after each group's presentation.

7.3.1 Multicontexualising NOS: Examining the paradigm shift in different contexts

Dr Kim's first strategy was evident in her selection and design of the activity, where she utilised multiple contexts to shed light on a single aspect of NOS. The primary reason for using this activity, she said, was to exemplify NOS in a range of historical contexts (Interview, April 2019). She believed that implementing this activity would also increase student participation as they work in groups. In the post-lesson interview, she described her intention by making multiple groups present on different topics:

> Dr Kim: After I taught NOS explicitly with the PowerPoint slides I had made, [I wanted to address] historical examples that bear those NOS aspects. **It'd be nice to cover one [example] in depth,** but then it would have been only about me [teaching it to students]. SIE has limitations in discussing [scientific content] knowledge. **So when they work together, they could encounter various topics of inquiry, various historical examples, so that's why I used the activity.** (Interview, April 2019)

The above quote suggests that the paradigm shift activity was possibly the result of Dr Kim's choice between the breadth and depth of student learning from the activity. She acknowledged that either approach has values, but she chose to prioritise the breadth for two reasons: to increase student participation and to allow for some learning of content knowledge of science alongside NOS. In other words, she was *multicontextualising* NOS by setting out to examine a single NOS idea in various contexts. In a study on instructional tasks for learning NOS, Park, Yang and Song (2020) suggested that examining a single NOS aspect in multiple contexts can be instrumental for a sophisticated and nuanced understanding of the NOS aspect. In this view, Dr Kim's multicontextualising strategy had the potential to maximise students' learning about paradigm shift compared to when they only learnt it within a single context. After all six group presentations were finished, Dr Kim opened the discussion by briefly reiterating her lesson aim of learning NOS through multiple examples:

Table 7.2 Dr Kim's instructional strategies allowing for the formative assessment of NOS

Dr Kim's instructional strategies			Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Theme	Code	Description	DNA	Plate tectonics	Heliocentric theory	Mpemba effect	Big bang theory	Vaccines
Multicontextualing NOS*	Uses multiple historical episodes to teach paradigm shift	Teacher designs and enacts the assessment activity where students present diverse historical episodes to show paradigm shift.	N/A*					
	Clarifies learning goal	Teacher states the goal of learning paradigm shift from multiple contexts.						
Mapping examples onto NOS	Clarifies learning goal	Teacher reminds students of the learning goal of understanding historical cases that led to a paradigm shift.	•					
	Explains a NOS concept	Teacher explains what a paradigm is.	•					
	Helps relating the example to the paradigm shift	Teacher asks or tells how the example shows a paradigm shift.	•			•	•	•
	Gives a compliment	Teacher acknowledges the presenters for linking examples to paradigm shift.	•	•				
Focusing the discussion on NOS	Delivers relevant content knowledge	Teacher provides scientific knowledge necessary for understanding NOS.			•		•	
	Clarifies learning goal	Teacher states or clarifies the NOS focus of the lesson.			•			
	States expectations about content knowledge	Teacher tells students to what extent she expects content knowledge to be presented.			•		•	
	Shuts down detailed content questions	Teacher discourages or shuts down when students ask questions on scientific details.			•		•	
Exploring diverse NOS aspects	Explains a NOS concept	Teacher explains additional NOS aspects that she wants to address.		•				
	Helps relating the example to other NOS aspects	Teacher provides scaffolds to make the link easier for students.	•	•	•	•		
	Connects to previous learning	Teacher refers to previous learning about content, NOS and/or history.		•				

* The multicontextualisation of NOS occurred during the design and selection of the assessment task and during her concluding remark (see Section 7.3.1), so is not represented in this table.

(The last group's presentation and Q&A has ended)

Dr Kim: So, after all, you had an opportunity to **think about the [content] knowledge and nature of science from at least six examples.** I hope you have learnt something from them.

Considering the diverse topics that her students presented (Table 7.1), Dr Kim's intention to examine the target NOS aspect in a range of different historical and disciplinary contexts seems to have worked in a way. In addition, as will be described below, there were several moments where some groups turned out to have chosen 'better' examples of a paradigm shift than other groups. These moments created potential opportunities to elaborate students' ideas of what paradigm is, had the teacher provided appropriate feedback in a timely manner (Park, Yang & Song, 2020). Nevertheless, during the two observed sessions, Dr Kim did not take any further actions to explicitly compare, contrast or link the six examples (or other similar examples) in terms of their alignment with the idea of the paradigm shift. Instead, her feedback and comments occurred within the individual example that each group presented.

7.3.2 Mapping historical examples onto NOS: More than just learning about what happened

Empirical studies on NOS teaching and learning strongly suggests that using stories from history of science is a necessary but not sufficient condition for students' NOS learning (Abd-El-Khalick et al., 1998; Abd-El- Khalick & Akerson, 2009; Akerson, Abd-El-Khalick, & Lederman, 2000; also see Section 2.1.2). In other words, NOS is not naturally learnt by simply learning about what people discovered in the past and how; The teacher needs to make conscious efforts to translate this into the target NOS. Dr Kim was aware of this fact from her graduate study. When she was asked what the purpose of assessing NOS would be, she emphasised making connections between the abstract NOS ideas and actual examples:

Dr Kim: [The purpose of assessment is] How well one understands the NOS. That's it … **Not just saying it blah blah only in words, but being able to apply what they understand when doing something like inquiry examples or historical examples.** 'This [case] is indeed about tentative

NOS, this part relates to what', if they can say like these. (Interview, April 2019)

When she gave the group presentation homework, Dr Kim stressed 'finding something that fits the topic of the paradigm shift, and presenting with an emphasis on that' (Interview, April 2019), which was further elaborated in her answer to my question:

Wonyong: When students were doing the presentations, how much did you expect of them? In terms of the content and paradigm aspects.

Dr Kim: So in a simplified way, some inquiry, some experiment. **In the past, most people thought and believed as such, but through this scientific experiment or inquiry, because of this, peoples' thoughts changed. Make this clear at a glance. That's what I required of them.** (Interview, April 2019)

Throughout the lesson, Dr Kim made several efforts to keep the students reminded of the lesson goal, which was to understand the historical cases that led to paradigm shifts. The discussion that occurred immediately after the first group's presentation on the discovery of the double helix structure of DNA illustrates this point. In their presentation, Group 1 used cartoon clips that featured four major figures in the story, James Watson, Maurice Wilkins, Francis Crick, and Rosalind Franklin, to illustrate the story of the double helix structure (Figure 7.2; see Appendix C for complete presentation). They then ended with drawing attention to the neglected contributions of the female scientist Franklin in the discovery process. In the following transcript, Dr Kim opened the Q&A by asking a question to the presenters:

1 Dr Kim: … **So today's presentation was meant to be about finding and explaining a historical inquiry that led to a paradigm shift**. So I'll ask a question first. It applies to all groups. **Why do you think this is a historical case that shows a paradigm shift? I'd like anyone from the group to answer.**

2 Presenters: [Silence for 10 secs] When we know the helical structure of DNA, it makes opportunities for human cloning, and tomato bearing potatoes, and the like, so we can say it's a paradigm.

3 Dr Kim: **Yes. A paradigm is, easily speaking, a system of thinking that prevailed people's thoughts in a certain era.** Then you mean that

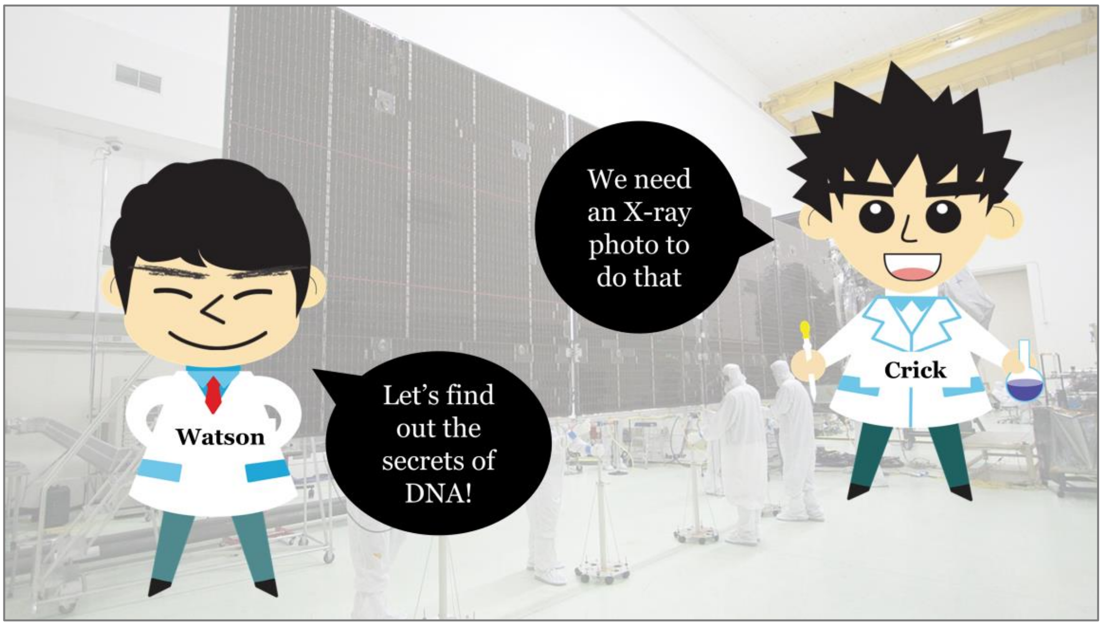


Figure 7.2 Slide from Group 1's presentation on the discovery of DNA double helix structure.

		there was **a huge change in biology** before and after discovering the double helix structure?
4	Presenters:	Yes.
5	Dr Kim:	Good. In that regard, you selected a really good topic. …

In this initial set of interactions, Dr Kim's intention to focus the discussion on the 'paradigm shift' is made clear. First, she clarifies what the given task was. Then she asks a question that encapsulates her intention: '*Why do you think this is a historical case that shows a paradigm shift?*', which she had prepared in advance to ask all groups (Line 1). It took the group, standing in front of the class, about 10 seconds until one of them started answering this question. During that pause, they turn their heads toward each other, one by one, as if they are waiting until any one of them will say anything, exchanging mixed facial expressions (Line 2).

From these 10 seconds, Dr Kim notices that the group is not prepared to answer her question and have not thought about it before. Their answer is focused on the practical benefits of the advances in genetics to human lives, which is not a defining characteristic of a paradigm (Line 2). Nevertheless, instead of criticising or correcting their answer, she quickly recognises them by saying 'Yes'. She then goes back to reminding the class of what a paradigm is, which is a prerequisite to connecting the example to the paradigm shift idea. She provides, in the form of a question, how the presented case about DNA can be interpreted as a paradigm shift. Although this interaction looks as if she is revoicing students' reasoning, particularly due to her use of 'you mean …', in fact it would be more plausible to infer that the group had not reasoned that way until Dr Kim did it for them (Line 3). The group then agrees to Dr Kim's reasoning (Line 4), and she makes a compliment by acknowledging their selection of the example (Line 5).

Although the above episode seemed to have resolved the issue of linking the DNA story to the paradigm shift, it soon turned out that it was not over. This became clear when one student, immediately after Dr Kim's feedback, asked an interesting question:

1	Student:	Since it's about the paradigm shift, then **what's the paradigm that existed before the double helix?**
2	(Class laughs lightly)	
3	Presenter:	I don't get the question.
4	Student:	So, when DNA double helix structure was discovered, your topic is that it is a paradigm shift. Then **what was the paradigm before it?**

This excerpt raises many interesting points, but before discussing these, it should be noted that whether the DNA example is an instance of a paradigm shift is in fact debatable. It is because there had already existed a specific problem waiting to be solved (i.e., determining the exact molecular structure of DNA), which was later solved by these scientists who had known that the problem existed and shared the knowledge base, methodological rules and evaluation criteria required to solve it. Hence, it may be problematic to call it a paradigm-shifting discovery through which anomalies unexplainable by old paradigms were recognised. The problem with this example is also apparent within the context of Dr Kim's lesson, where she used the example of the shift from Aristotelian to Galilean theories of motion (Section 7.2), which concerned more fundamental changes in people's beliefs about the world and methods of doing science. Such background helps us better understand the nature of the above episode—Since the discovery of DNA was not a paradigm-shifting example as evident as the Copernican revolution or the Galilean revolution, it was not easy for the presenters to name what were the old and new paradigms respectively and to contrast the two.

What is interesting is that the teacher did not react to this issue in the lesson although she allowed the conversation to keep going. Later I talked about it with Dr Kim, and we both agreed on a problem, that the way the concept of the paradigm shift had been described in the curriculum as well as in her lesson was rather broad and ambiguous. Following the national curriculum and the textbook, she defined paradigm in her lesson as ‘a system of thinking that prescribes the opinions and thoughts of people living in a certain time’ (Dr Kim’s lesson slides) and described it as a key mechanism through which science advances, but neither resource specifically addressed specific criteria as to what does or does not count as a shift of paradigm.[19][20] This vagueness was

---

[19] Philosopher of science McMullin (1993) noted that there were at least three different depths of the different changes that Kuhn referred to as a scientific revolution. According to his view, a Kuhnian scientific revolution (that is, a paradigm shift) can be deep, intermediate or shallow. In deep revolutions such as the Copernican revolution and the quantum revolution, the revolution causes fundamental changes in people’s beliefs about the order of the world, what constitutes valid evidence for scientific claims, and so on. An intermediate revolution occurs when a large-scale theory replaces an old theory, such as in the example of shift from the phlogiston theory to the oxygen theory of combustion in chemistry. These revolutions do involve the emergence of a new conceptual framework and new set of problems, but the fundamental epistemic and ontological beliefs remain largely unaltered (McMullin, 1993). Finally, there are shallow revolutions, such as the one caused by Wilhelm Röntgen’s discovery of X-rays in 1895. In these cases, no fundamental change in theory occurs, and no prior anomalies are noticed to signal the revolution to come. Instead, the changes mainly concerned the types of instrumentation and the experimental procedures for working with cathode-ray equipment (McMullin, 1993). Kuhn himself presented the X-ray case as a scientific revolution (Kuhn, 1962), but McMullin pointed out that labelling this as a paradigm shift seemingly contradicts what Kuhn had specified as the criteria for an impending revolution (i.e., awareness of anomalies, a resistance to change, etc.).

[20] In light of McMullin’s classification based on the depth of scientific revolutions, the DNA example seems to best fit with the shallow revolution, in that it primarily invoked methodological and technical innovations rather than theoretical changes or our conceptions of how the world is ordered. However, as discussed in Note 19, Nickles (2017) recently pointed

understandable, considering the inevitability of simplification as specialist knowledge goes through didactic transposition (Kampourakis, 2016).[21]

The issue of linking the paradigm shift concept to specific examples that were of focus in the presentations recurred multiple times in the later groups. In another Q&A conversation that followed Group 6's presentation on how vaccines were discovered by Edward Jenner and other scientists, a student from the audience asked:

1	Student:	Could you tell us about the paradigm shift?
2	Presenter:	The paradigm is, in the past, for treatment and prevention, it was antibacterial or the kind. [People thought diseases are] Because of hygiene so we need a hygienic environment, but here [in the vaccine example] injecting germs conversely creates tolerance [*sic: immunity*], so injecting germs on purpose. So this is a paradigm shift …
3	(Several clarification questions and responses are exchanged)	

From a paradigm shift perspective, this response generally sounds less objectionable than what was said about the DNA example. It contains all three key elements of typical 'shifts'—the original state, the new state and what changed between the two (Line 2). This example, as was for the DNA example, might not be a paradigm shift from a strict history of science perspective; It can be a shallow revolution at best. However, in the conversation of Dr Kim and her students, it is being considered a decent example that meets all the requirements to count as a paradigm shift. Unlike the DNA group's case,

out that it is unclear whether Watson and Crick's discovery is a Kuhnian scientific revolution, since it did not challenge the Darwinian paradigm although it had a significant impact on the development of molecular genetics as a field.

[21] The difficulty of deciding what a paradigm shift in biology has also been suggested by Weiss (1996). He argued that it is unclear if the major changes in our understanding of genes triggered by modern DNA technology, aside from their contributions, constitute 'paradigm shifts' in the Kuhnian sense. In the context of genetics and heredity, Oldham and Brouwer (1984) characterised the Mendelian genetics to be the first paradigm in the field, while they viewed that there had been no paradigm but only competing, partial explanatory systems before Mendel's.

this time, the presenter's explanation of the link to the paradigm shift was accepted among the students, and Dr Kim did not add anything to this discussion. In addition to the perceived suitability of the topic, another point to be noted when comparing Group 1's and Group 6's cases is that they were the first and last groups. Between these two presentations, Dr Kim made many strategic efforts to clarify the lesson goal and focus the discussion on NOS (Table 7.1), so the importance of the '*Why is this case a paradigm shift?'* question had been better shared among class members by the time of Group 6's presentation. Contrary to Group 1's case where ten seconds of silence came after Dr Kim's question, in Group 6's case, the same question was asked by a fellow student and was responded with no delay by the presenters.

Similarly, the Q&A after Group 4's presentation of the Mpemba effect provides an example of peer formative assessment against the shared learning goal, but in a more sophisticated form. The Mpemba effect refers to a phenomenon that hot water takes less time to freeze than cold water does, coined after a Tanzanian schoolboy Erasto Bartholomeo Mpemba who discovered it while making ice cream (Mpemba & Osborne, 1969). Group 4 also spent most (8:33 of 11:57) of their presentation time to provide a detailed scientific explanation of the Mpemba effect. The group started by explaining how the effect came to be known and has been verified and explained by scientists until recently. To do this, they explained the molecular structure of water, and how water molecules are tied together with hydrogen bonds.[22] They then explained that hot water molecules can release the energy stored in hydrogen bonds faster than cold water molecules do, hence freezing faster, and concluded the presentation.

Dr Kim asked the presenters to moderate the Q&A, and they asked if there were any questions, but no question was asked. Dr Kim then interjected and looked at me sitting at the back of the classroom, to ask if I had any. I recognised the group's interesting presentation and then asked for clarification on why the Mpemba effect example makes a case of paradigm shift. The answer they gave me then faced a series of questions from other students. This process is illustrated in the following episode:

---

[22] In the Korean curriculum, the molecular structure of water is included in Grade 11 Chemistry (an elective subject), so most students in the class had not learnt this in school unless they knew about it otherwise.

1	Dr Kim:	… (Asks the researcher) Do you have a question? We can have a critical question.
2	Wonyong:	[To presenters] I liked your presentation. I just missed what you said about **how this example relates to the paradigm shift, so could you explain that again?**
3	Presenter:	So, this is a paradigm. A paradigm means what people think over a time period, until now, for as long as 2,000 years. An average person would think so [that hot water would freeze faster]. But like in the 2003 discovery [theoretical verification of the Mpemba effect], breaking the stereotype, so breaking it and changing people's thinking so we thought it's a paradigm shift.
4	Student 1:	**So in the past, people took it for granted that hot water freezes faster than cold water?**
5	Presenter:	Yes, didn't they?
6	(Other students get confused and start to talk)	
7	Student 1:	They knew that since long ago?
8	Presenter:	Long time ago, Descartes and Aristotle mentioned such a thing. But they couldn't **prove** it and just thought that it is what it is. But through the 2003 discovery …
9	Student 2:	Then isn't it that they already knew it? That it freezes faster?
10	Student 1:	**So if they had not known it [hot water] freezes faster, it [the discovery of Mpemba effect] would have been a paradigm shift …**
11	Student 2:	**Well, they knew it. It's not like they came to learn about a new theory.**
12	Student 3:	**You mean the proof was added?**
13	Presenter:	Yes, because the proof was not added until …. That's right. **We think so. Anyone can make a hypothesis, and anyone can have thoughts. But we think that publishing [*sic: proving*] is important, so we presented with that focus about the paradigm shift**. (End of Q&A)

14 (Applause)

15 Dr Kim: I think connecting this example directly to a paradigm shift is a bit far-fetched …

The presenter's response to my question, which should have been no longer a surprise by this time, begins by referring to the definition of a paradigm. He then uses this criterion to claim that an old, common idea that had existed since Aristotle's time was changed by the 2003 discovery that explained the Mpemba effect, so it counts as a paradigm shift (Line 3). Here, he mentions that the effect is coherent with an average person's idea. Student 1 then asks a question to check if the presenter said it correctly about the average person's idea (Line 4), and the presenter confirms that it is what they meant (Line 5). Student 1 then repeats the question (Line 7), and now the presenter distinguishes the old state and the new state in his response based on whether the phenomenon is 'proved' (Line 8). Despite their answer, Student 2 is still not sure and says that the effect was already known from a long time ago (Line 9). Student 1 adds to her comment using a conditional statement, saying that the Mpemba effect must have been previously unknown to count as a paradigm shift (Line 10). He then uses this criterion to assess the presenter's answer—people knew the phenomenon was already known, so it is not a paradigm shift (Line 11).

These exchanges between the audience and the presenter exemplify how peer assessment can facilitate students' sophisticated understanding of NOS. Without any help from the teacher, they clarified the problems of linking the Mpemba effect to the paradigm shift. In the meantime, the condition of 'being previously unknown' was elaborated and sophisticated as they discussed the Mpemba effect case. These can be seen as incidents where not only the teacher but also learners become agents of formative assessment (Wiliam, 2010), and through such engagement, co-constructs knowledge about NOS as well as the historical example.

As it becomes gradually clear that the Mpemba effect is not a paradigm shift example, Student 3 picks up on the word 'prove' from the presenter's answer and asks if it was why they selected this example (Line 12). The presenter states that the addition of proof indeed was the paradigm-shifting part since it is what they thought to be the most important contribution (Line 13). Although the group's intention became less confusing by changing the focus from the phenomenon itself to the recent scientific verification of it, Dr Kim is clear in concluding that the Mpemba effect was not a good example of a paradigm shift (Line 15). Instead of shutting the case down, however, she then attempts to use the group's presentation in a constructive way, which will be illustrated later in Section 7.3.4.

Reflecting on Dr Kim's efforts to transform the historical examples into students' learning of NOS (i.e., the paradigm shifts) offers us a rich illustration of the affordances of her formative assessment strategies and how they can collaboratively keep the lesson focused on the target NOS. As the paradigm shift focus of the lesson became clear to the students, more detailed and productive feedback occurred among students, leading to an elaborated understanding of the paradigm shift. These processes made the lesson more interactive and participatory. Even the examples that were concluded to be not paradigm-shifting (i.e., DNA and the Mpemba effect) contributed to a sophisticated understanding of the paradigm shift by illustrating what it is *not*.

At the same time, however, the episodes showed that these opportunities are not without challenges. These challenges were epitomised by the '*what was the previous paradigm*' question in the context of the paradigm shift, but they can arise in other forms depending on the NOS being discussed. This question shows the complexity of providing feedback on NOS-related topics, especially when the lesson is student-focused and relies on student participation. NOS, as with other knowledge taught in schools, is the result of a 'didactic transposition' of academic knowledge (Kampourakis, 2016). Hence, interpreting actual examples for contextualising NOS in terms of NOS can often be very demanding to be done by students or teachers. Considering that Dr Kim had a high level of knowledge in NOS and its teaching, this issue could be far more challenging for teachers who are not familiar with and not trained for NOS.

7.3.3 Focusing the discussion on NOS: The dilemma of addressing content knowledge

In the previous section, I showed that Dr Kim tried to facilitate the whole-class discussion so that it could lead to learning of an aspect of NOS (i.e., the paradigm shift) instead of merely talking about people, events and stories from the past. Meanwhile,

Figure 7.3 During Group 3’s presentation, the presenter is using the whiteboard to explain the evidence of heliocentrism

her effort to keep the lesson focused on NOS was found in another related context as well, as she was trying to strike a balance between a NOS aspect and the scientific content knowledge from which the NOS aspect is derived. This strategy was manifested in Dr Kim’s talk as well as how she managed the formative assessment conversations during the Q&A session. To illustrate this strategy, in what follows, I draw on the analysis of two contrasting episodes from the discussion after Group 3’s and Group 5’s presentations.

Group 3 presented on the shift of paradigm from the geocentric to the heliocentric view of the universe. The group spent about one minute and a half explaining the backgrounds and contents of Ptolemy’s and Copernicus’s models. From 35:00 to 36:22, a member of the group picked up a marker started drawing the Sun, the Venus and the Earth on the whiteboard, to explain how the phase and size change of Venus supports heliocentrism over geocentrism (Figure 7.3). They then went on to briefly (< 30 secs) introduce that there were three major pieces of evidence that supported the heliocentric model: Kepler’s three laws, Galileo’s observation reports of Jupiter, and later studies by Newton and other scientists. The presentation ended with summarising the two systems and key aspects of the change, and sharing the group’s reflection on their collaboration.

In the whole-class discussion that followed, one student raised her hand and asked what she called a ‘specific’ question about the presentation, which is shown in the following excerpt:

1	Student:	Can I ask a **specific** thing, ask specifically? I can’t, right?
2	(Class laughs)	
3	Presenter:	Well, if I’m able to …
4	Student:	Frankly, it’s out of personal curiosity. So when I see your PPT, the evidence for heliocentric theory includes Kepler’s planets, Galileo’s observations and the like. **Could you explain what Kepler’s planets are about?**
5	Presenter:	First of all, Kepler …
6	Class:	(With excitement) Ooooooh
7	Presenter:	… argued for heliocentric, heliocentric theory. Then many scholars did various research to prove the heliocentric theory, and among them, there was Kepler’s third law of planetary motion … (Class laughs and boos) The content is quite hard. (Class laughs)
8	Dr Kim:	So now, **we are not focusing on the scientific knowledge** to learn about Kepler’s third law or specifics of Newton’s law, so we can move on. **But what’s important is, when you present, it’s important that you summarise and present what you can digest, right?** In that regard, although it’s not a knowledge-focused presentation, **you need to be able to introduce it [relevant scientific knowledge] briefly**. The part that you are talking about. So good, well done.

In this episode, the student gets curious about a specific part of the group’s presentation, namely Kepler’s laws, and asks if they could elaborate on it (Line 4). Before she asks this question, she checked if it would be okay to ask this kind of ‘specific’ question to the group, which caused the class laughter (Lines 1–2). The presenter agrees to answer if he is ‘able’ to do so (Line 3). At this moment, we can see that this Q&A session has turned into a test of the presenter’s ability to answer the asker’s hard question as if it were a kind of a game show on television. When one of the presenters, with no hesitation, started saying ‘First of all …’, the class gets excited and says ‘oooooh’, showing that their position has changed to the audience watching the game show (Lines 5–6). However, to the disappointment of his peers, the response that he made was merely a repetition of his earlier argument that Kepler’s laws were part of the evidence for the heliocentric theory, and not really answered what the laws actually are. The class recognises this and reacts with laughter and a friendly boo that shows their disapproval. The presenter admits that he did not answer the question and revealed the gap in his knowledge, by saying that the content was quite a difficult one (Line 8).

Dr Kim’s comment after this ‘game’ ended as a failure has several important implications. In particular, her first sentence can be read in several ways (Line 8). First, she is stressing that the point of their discussion is not the specifics of the evidence or any theories, thereby refocusing the lesson on the target NOS. Second, with ‘we can move on’, she intends to end the discussion on this ‘specific’ question about content knowledge without further elaboration. The third message is relatively implicit and is related to the fact that this presentation was also being marked by Dr Kim for summative purposes. With this context in mind, her statement can be seen as reassuring the presenters that failing to answer the content-related question would not damage their marks much, so making them less worried about the consequences of their unsatisfactory answer to the question.

In the rest of her turn, she tells the class that ‘scientific knowledge’ still matters although it is not the focus. She then states her expectation pertaining to the presentation of content knowledge—that students are expected to summarise and present the relevant content knowledge that they can ‘digest’ and mentioned in the presentation. She concludes this discussion by acknowledging the group’s effort. This episode, initiated by a ‘specific’ question from a peer, exemplifies the challenge in dealing with the relationship of content knowledge and NOS in lessons. A sophisticated understanding of how paradigm shift leads to the development of science would include making sense of the specifics of science itself, but in many cases, it is impossible to be achieved at the secondary school level. Moreover, it is not always necessary to grasp the scientific details to understand the meaning of science. For example, Park, Yang and Song (2019) suggested that there are many aspects of NOS found in general relativity theory and gravitational-wave physics that can be addressed to secondary students with minimal conceptual understanding requirements.

An interesting contrast to the above episode with regard to the content knowledge-related issues arose after Group 5's presentation. Group 5 presented how the steady-state model of the universe was replaced by the Big Bang theory. Contrary to Group 3 who spent minimal time on specific scientific information, Group 5 spent a substantial portion of their presentation time (9:34) to explain the contents of Einstein's steady-state model, the Friedmann model and the Big Bang theory, as well as the empirical evidence supporting each model. They mentioned theoretical constructs such as critical density and the cosmological constant, and evidence such as Olbers's paradox, cosmological redshift and the cosmic microwave background. The presentation was concluded with the significance of the Big Bang theory. The following transcript illustrates the Q&A on Group 5's presentation:

1	Presenter:	Thank you. Any questions?
2	(Silence: 5 secs)	
3	Student:	Can I ask about that, the critical value?
4	(Silence: 12 secs)	
5	Dr Kim:	Don't ask hard things like the critical value. (Class laughs) … So Einstein's strong theory declines, and cosmology is, it is what uncovers the origin and evolution of the universe. In so far as the origin, and the evolution process, people have made remarkable development since. So we know the Big Bang theory is reliable among cosmological theories, and how the universe expands, with acceleration or otherwise, people are doing research about that. There are some conclusions out there. In that regard, it is clearly an example of the paradigm shift, right? **It would've been nicer if you focused on that and presented it a bit easier. All groups alike. Let me tell you. Only the things you guys can understand. You don't need to bring all the evidence. Only one of the three, but it's good if you digest it and express it in your own language**. Anyway I can see that you prepared a lot …

At the beginning of this episode, the student checking if he could ask a question about the critical value (Line 3) is reminiscent of the one who asked that ‘specific’ question to Group 5. However, the episode unfolds differently here when Dr Kim interjected to suppress this content-related question. By doing this, she might have liked to prevent the discussion from turning too technical and losing the focus of the paradigm shift. After cutting the question, she decides to summarise the presentation herself in plain language and provide interpretation in terms of the paradigm shift, instead of waiting for it to happen from the students. After she points out that what the group presented would certainly count as a shift of paradigm, she expresses some disappointment. When she stated that the group could have focused on explaining how the Big Bang theory makes an example of the paradigm shift, the implication is that the group did not do so, since they spent too much time detailing the specifics of the theories and supporting evidence (Line 5). She then states her expectations about the presentations in terms of content knowledge, by emphasising to bring the content knowledge that they can ‘understand, digest and express in their own language’ (Line 5). This, together with her comments to Group 3 on content knowledge, shows her standards on how much and in what manner she expects the relevant scientific content to be delivered in the presentation.

At the follow-up interview one week later, when this episode was brought up, Dr Kim recalled that she felt that Group 5 did not have a good grasp of the activity’s objective, which was to interpret a historical example in terms of paradigm shift. Between my observation and the interview, she had taught the same unit to three other classes. Since the lesson I observed was her first-time implementation of the unit, she described the observed lesson as a ‘process of trial and error’. As she taught more classes, she felt improvements in her practice:

Dr Kim: For Class 1 [the observed class], **I trusted the kids and it became trial and error.** But as I did it [the lesson] for Class 2, Class 3 and Class 4, I regret you [researcher] didn’t see them, **it did get better a lot.** For example, encouraging the students’ questions to be more about the paradigm shift …

Wonyong: You did that?

Dr Kim: Yes. Or stopping them when they ask too difficult science-related questions. Explaining something that students can’t explain themselves

> rather clearly and easily. Using the figures in their PPT to explain [the content knowledge] and move on. Caring about connections to the general science subject. I could afford to do these things. (post-observation interview)

Taken together, the episodes of Group 3 and Group 5 show us how Dr Kim dealt with the dilemma of addressing content knowledge when formatively assessing her students' learning of the paradigm shift idea, and more generally, when teaching NOS. For Group 3, presenters only mentioned the relevant content knowledge and failed to elaborate even when this was asked by another student. For Group 5, presenters did so much on explaining the scientific details that the aspect of the paradigm shift was hidden. Dr Kim tackled such dilemmas by stating and elaborating her expectations about to what extent content knowledge should be included in the presentation, which was first proposed in Group 3's discussion and further conditioned in Group 5's discussion. These episodes illustrate that the challenge of balancing content knowledge and NOS knowledge is inherent in NOS teaching and thus inevitable (see Section 8.6 for further discussion on this issue), but it can be reduced and managed by teachers' effective use of feedback, classroom conversation and other formative assessment strategies as they accumulate NOS teaching experience.

7.3.4 Exploring diverse NOS aspects: Thinking beyond the paradigm shift

In the current NOS literature, historical case studies are most often used to teach theory change in science and tentativeness of science, among other NOS aspects (Allchin, 2014). This is because these aspects typically manifest over a longer time scale. However, Allchin et al. (2014) argued that historical cases can also be used to illustrate the critical roles of debate, the cultural relevance of science, funding, expertise and credibility, and ethical aspects of scientific research (see also Section 2.1.3). Throughout the paradigm shift lesson, there were three instances that Dr Kim elicited her students to think of NOS aspects besides paradigm shift from the examples that the groups presented. Particularly for Group 4, since Dr Kim was explicit to say that the Mpemba effect should not be considered a paradigm shift example, she attempted to invoke some other NOS aspects that could be learnt from the story. However, none of her elicitation attempts resulted in students' responses, so she ended up providing the additional NOS aspects on her own.

One such effort was observed after Group 4's Mpemba effect presentation. After the peer assessment conversations where the paradigm shift aspect of the story was debated (see Section 7.3.2), Dr Kim shut down the discussion momentarily by telling the class that it would be 'a bit far-fetched' to call it a paradigm shift. However, she did not shut down the conversation simply by denying the Group's example. Instead, she attempted to use their example to discuss other NOS aspects than the paradigm shift:

(After the peer discussion on whether the Mpemba effect shows a paradigm shift)

1	Dr Kim:	I think connecting this example directly to the paradigm shift is a bit far-fetched. But it's still a meaningful example in terms of the nature of science. [Student name] want to talk about it? What aspects of nature of science can we see in the example?
2	(Silence)	
3	Dr Kim:	Or [another student name] or [yet another student name]? [Student name] talked a lot.
4	(Laughter)	
5	Dr Kim:	**So first, we can see, scientific knowledge is confirmed by evidence. Through an example. Next, they developed an explanatory system. Inquiry techniques had developed, so that [Mpemba effect] was proved, right? As we know more, more things we can prove, and science develops. Science is based on, relies on existing theories. We can say these things.**

As shown, she starts by recognising the value of the example for learning NOS (Line 1). Then she calls on a student in the audience to ask what other aspects of NOS he noticed, but this elicitation does not lead to an answer (Line 2). She then calls on two other students, which also was not successful in eliciting responses (Lines 3–4). Dr Kim then decides to tell the students what other NOS aspects could be drawn from the story, such as the relationship of scientific knowledge and evidence, the advancement of experimental techniques leading to new discoveries, and the cumulative nature of scientific development (Line 5). In this episode, what we see is Dr Kim's attempt to utilise the presentation material to illuminate *other* aspects of NOS, although it had turned out to be irrelevant to the original NOS learning goal (i.e., paradigm shift). However, the fact that no students, even when called on, suggests that inferring another NOS aspect from a scientific story was not an easy task for them.

In another episode, after Group 2 presented on plate tectonics and continental drift theory, Dr Kim made another attempt to discuss other NOS aspects that could be said about the story. They presented what plate tectonics is, how the idea first appeared and became a paradigm, and what are the benefits and limitations of plate tectonics as an explanatory concept. At the end of their presentation, they revisited the definition of a paradigm shift from the previous lesson, and that plate tectonics meets the definition. When the presentation ended, Dr Kim first recognised how the group connected the example and the paradigm shift, and then directs the class's attention toward NOS, specifically the tentative nature of science:

1	Dr Kim:	Alright, well done to Bin's group who presented on continental drift. Next, let's have Q&A again. Seems like I can't give you much time. Just a minute. Let's have one question. You really explained well on the paradigm shift, and connected it well with continental drift theory, so I want to clap for that. **When we say the nature of science, we tend to talk about the tentative nature as a representative example. Anyone has a question related to that?**
2	(Silence for 7 secs)	
3	Dr Kim:	Is it difficult? Do you need another summary? When we studied with electricity earlier, don't you recall anything? Something leads to something. (Silence, 3 secs) The fill-in-the-blank that you like. All blanks are for filling. Hint, the development of science and technology …
4	(Silence for 5 secs)	
5	(Class laughs)	
6	Dr Kim:	So the limitations of plate tectonics, could be replaced [sic: overcome] through the advances in exploration techniques, right?
7	Class:	Yes.
8	Dr Kim:	**So, relating to the tentative nature of science, science relies on**

> **technology, and one advance in science and technology can lead to another. We can say these things**. Guys, I'm really sorry for doing this summary myself. Well done.

She recognises that the presentation was clear on linking the case to the paradigm shift and praises the presenters for that. Instead of adding anything on the paradigm shift aspect, she tries to use the group's presentation as a basis to discuss other aspects of NOS. She then refers to the tentative NOS and asks if anyone has thoughts about that aspect (Line 1). Seven seconds' silence followed (Line 2), from which Dr Kim knows that her question was difficult for the students. Then she invokes what students learnt from the previous lesson by referring to 'electricity', and spontaneously changes her question into a less difficult, fill-in-the-blank form ('*Something leads to something'*), which still did not work. She then gives another clue to make it easier by giving the first 'something' (Line 3). After these three attempts, students still were unable to respond for five seconds (Line 4). At this point, Dr Kim simplifies her question even further into a yes-no question (Line 6). She first starts from the fact that the current version of plate tectonics might not be final and could change as the exploration techniques develop (Line 6), and then draws a general NOS aspect on the interdependence of science and technology. She tells students that she feels 'really sorry' about having to tell students these interpretations instead of letting them do so, admitting that her attempts at discussing other NOS aspects did not go as intended (Line 8). Note that, in light of the ESRU model of informal formative assessment, the above quote includes incomplete cycles since her questions kept failing to elicit student responses. This contrasts with earlier exchanges described in the chapter, where she could seize on the student responses to create productive cycles that made the conversation dialogic and contributed to the NOS lesson goal.

In the Q&A for Group 3's presentation on the heliocentric theory, Dr Kim referred to the bravery of Copernicus to write a book that contradicts the then prevalent religious worldview. She told the students about how brave one should be to challenge the geocentric paradigm that survived for over a thousand years, and how hard it would have been for the medieval people to accept Copernicus's theory. She then referred to how the old paradigm was gradually replaced by the heliocentric system, as the geocentric system failed to produce accurate predictions about the position of celestial bodies, change of weather, and so on.

Collectively, her practices to use the presentations to shed light on NOS aspects other than paradigm shift were in the form of questioning, but none of these attempts led to her students answering her question. She gave hints and changed the question into the blank filling to make the question easier for students, but these efforts still did not work as intended. She ended up providing the additional NOS aspects on her own and touched on various epistemic and social aspects ranging from the tentativeness of science to the ethos of scientists to challenge the religious worldview. Again, this feedback by Dr Kim was mostly impromptu and on-the-fly rather than planned, as she did not know the presentation contents in advance and had not taught the same lesson before. In such a setting, to instantly discover additional NOS aspects from a story and deliver it to students, a teacher would need to have the ability to do so, as in Dr Kim's case. Such efforts of her to link the example to multiple NOS aspects are encouraging in that it is coherent with the recent emphasis on the interconnections between NOS aspects rather than treating NOS ideas discretely (Peters-Burton, Parrish & Mulvey, 2019), suggesting the potential of using similar activities to foster a rich and interconnected NOS understanding.

7.4 Dr Kim's NOS instruction as building connections through formative assessment

Considering that formative assessment is a process where 'evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers' (Black & Wiliam, 2009, p. 9), Dr Kim's practices in the paradigm shift lesson can be seen as creating a space for formative assessment of students' NOS learning. Figure 7.4 is a graphical representation constructed from the analysis of Dr Kim's instructional strategies that allowed for the formative assessment of NOS. As shown, her practices unfolded as she mediated the central NOS learning goal with four related elements: content knowledge, history of science, multiple historical examples, and other NOS aspects. Although connection building has been emphasised by NOS researchers, the focus of the connections was usually between the inquiry activity and the NOS that it targeted (Hanuscin et al., 2011) or between different aspects of NOS (Mulvey et al.,

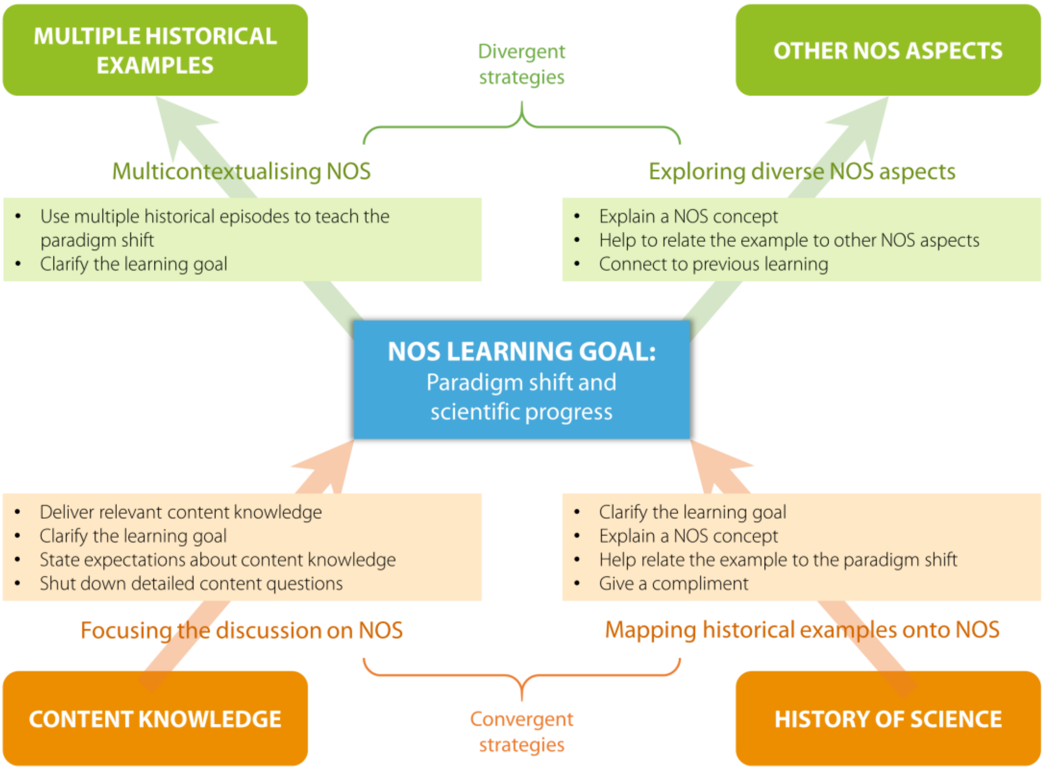


Figure 7.4 Dr Kim's instructional strategies that allowed for formative assessment

2021; Peters-Burton et al., 2019). In this sense, Dr Kim's lesson allows for seeing various types of connections illustrated in Figure 7.4, providing a more holistic and complex account of connection building that can be targeted in NOS instruction.

One way to characterise the four themes that emerged from her formative assessment practice is to understand them in terms of what might be called *convergent* and *divergent* strategies for NOS learning. Convergent strategies are the teacher's conscious efforts to direct the focus to the target NOS idea instead of scientific and historical details while making connections with these components. On the contrary, divergent strategies are used to promote learning a single NOS in multiple contexts or learning multiple NOS aspects from a single historical case. The ways she dealt with the background content knowledge (Section 7.3.3) and the historical details (Section 7.3.2) had one common aim: keeping the discussion focused on the key NOS learning goal, and preventing these contributing elements from becoming the focus of the lesson. However, since she believed that NOS should be understood in context rather than isolated statements, her feedback was concentrated on keeping the content knowledge and the historical context to an appropriate level. In this sense, these formative assessment practices represent Dr Kim's effort to make the lesson converge on the learning goal and not be centred around the content knowledge or the historical details.

The other two themes, multicontextualising (Section 7.3.1) and linking the example to other NOS aspects (Section 7.3.4), are divergent in nature, in the sense that they have the potential to offer broader NOS understanding compared to when only one NOS is delivered using a single context. Considering paradigm shift in multiple contexts and considering other NOS aspects of the same story requires students to have not only a sound NOS understanding but also creative thinking skills that connect their content knowledge, historical episodes and NOS aspects. This can pose challenges to both students and the teacher, particularly when the lesson heavily relies on student activities and participation. When the teacher attempted a divergent discussion about NOS, the ESRU cycles were mostly incomplete due to the silences after teacher questions, possibly because the students did not have enough background knowledge that was necessary to build external links to NOS aspects or other historical examples other than what had been already covered in the lesson. It is therefore unsurprising that students had difficulties in these respects (e.g., presenting examples that were not paradigm shifts, and failing to respond to Dr Kim's elicitation) despite the teacher's efforts. On the whole, Dr Kim's formative assessment practices can be seen as a process of coordinating and balancing the convergent and divergent strategies to promote students' NOS learning.

The connection building in her lesson was also supported by various contextual factors. In light of Ruiz-Primo's (2011) five conditions for effective formative assessment conversations (Section 2.2.4), Dr Kim's practices illustrate how these conditions can be fulfilled in the context of NOS instruction. First, her lesson was guided by a clear learning goal about the paradigm shift, which she reminded her students of through her feedback and questioning. Second, by using a planned formative assessment that included whole-class Q&A, she made the lesson dialogic and interactive, not only between herself and the students, but also among the students. Third, she managed to support the students to achieve the NOS learning goal (for example, by delivering relevant content knowledge, providing scaffolds, and

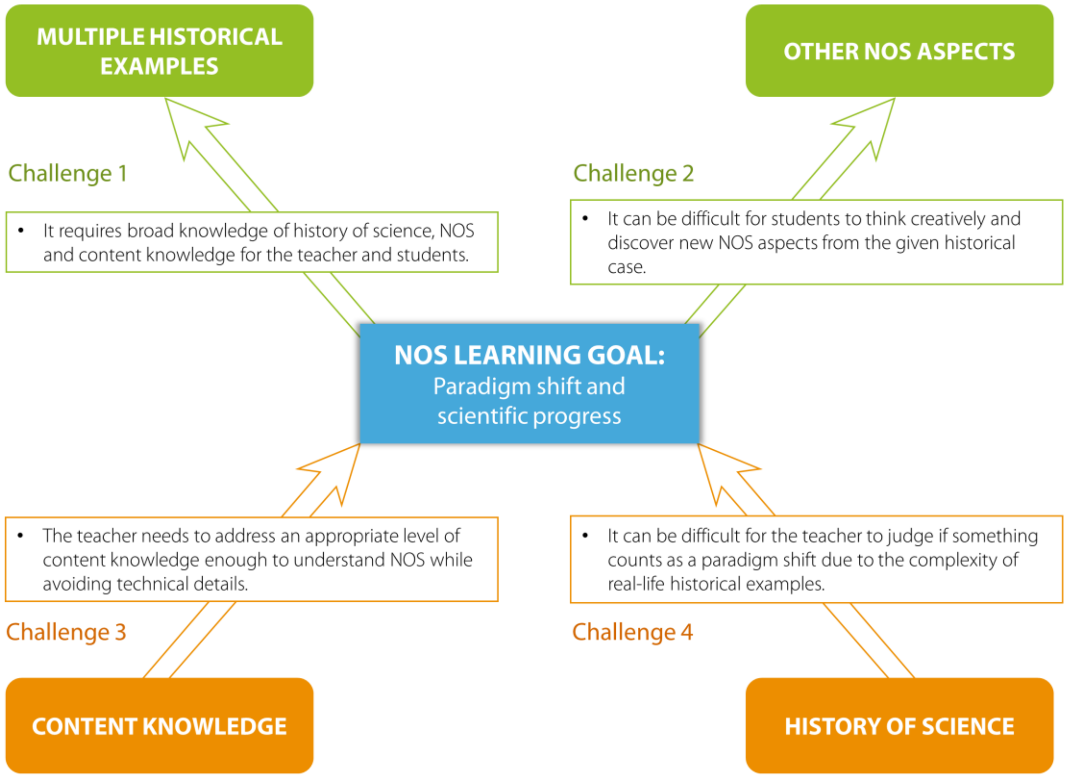


Figure 7.5 Potential challenges of formative assessment of NOS observed in Dr Kim's paradigm shift lesson

connecting to previous learning) and challenge them (for example, by asking questions, and discouraging detailed content questions). Fourth, the fact that Dr Kim was the homeroom teacher contributed to the friendly classroom environment, which enabled the students to actively participate, express opinions and give feedback. Fifth, the assessment conversations in the Q&A sessions encouraged the students to engage in argumentative practices that reflect the culture of the scientific community.

Meanwhile, her classroom episodes also revealed the several major challenges that can arise for teachers' formative assessment as they teach NOS (Figure 7.5). Although studies have previously suggested a range of practical difficulties that can arise when teaching NOS (Allchin et al., 2013), the formative assessment conversations in Dr Kim's classroom are meaningful in showing vividly how those challenges can unfold in different historical cases, how they are mutually related, and how the teacher reacted to those challenges. For example, episodes that showed the tension between 'too much' and ‘too little’ content knowledge suggest the difficulty of dealing with the content knowledge demand for effective learning of NOS. All four practices Dr Kim used to facilitate students’ NOS learning were not without challenges, some of which stemmed from pedagogical considerations (e.g., the level of content knowledge required to understand NOS), and some others inherent in NOS (e.g., the complexity of matching historical examples to NOS).

These difficulties observed from Dr Kim’s lesson suggest that issues often claimed to be too sophisticated and difficult and thus irrelevant on the school level sometimes may not be really avoidable to arise in classroom practice. In a sense, the development of theories in NOS education over the past decades can be seen as an effort to transform the complex and controversial ideas about science into a simpler, less contentious, and tidier form of knowledge that could be effectively learnt by young students. For example, the debate around realism and instrumentalism about the world is downplayed to highlight the empirical NOS (Matthews, 2012), and different perspectives about the extent to which science is tentative or durable are also obscured by the statement about tentativeness. Such a trend is epitomised in the consensus view and the paper-and-pencil tests of NOS grounded on this view, where debates about NOS are hidden for their ‘pragmatic irrelevance’ (Abd-El-Khalick, 2012). Underlying this view is an assumption that key aspects of NOS can be taught without worrying about the controversies in history, philosophy and sociology of science.

However, Dr Kim’s lesson seems to raise questions about whether the controversial aspects of NOS are indeed not relevant when NOS is taught and discussed in the classroom. Consider the following questions suggested from the lesson: *What makes a scientific discovery paradigm-shifting or not? Should there have already existed a paradigm?* Although these critical questions were not something Dr Kim had as lesson goals in the first place, as the groups presented several different historical cases, these questions naturally emerged to her students and herself so were needed to be resolved. This suggests that there can be occasions where such specific questions about NOS are not completely irrelevant to teaching and assessing NOS and thus cannot be just ignored. This is not to say that teachers and students need to engage in the sophisticated philosophy of science debates, but to shed light on a counterexample to the irrelevant argument and initiate further discussion on addressing controversies in classroom practice (see Section 8.8 for further discussion).

Given that the lesson occurred over two sessions, it is worth noting that the quality of peer assessment and peer feedback on the presentations improved in the second week, and played a salient role in increasing student agency and creating learning opportunities. Consider the moment where one student asked about what the paradigm shift was in Group 6's vaccine story during the first week's session (Section 7.3.2). As Dr Kim repeatedly emphasised the learning goal, students started asking the '*why is this a paradigm shift*' question to the presenters by themselves, which implies that they acknowledged the importance of this question and thus shared the learning goal. Here, peer feedback contributes to forming a shared learning goal and students' commitment to attaining them (Hattie & Timperley, 2007). In another peer assessment episode following Group 4's presentation on the Mpemba effect, it became clear that the conversation between the presenters and the audience is leading to an enhanced understanding of the case itself as well as why the presenters thought it counts as a paradigm shift (Section 7.3.2). As evidenced by these episodes, the second week's feedback was more student-centred, occurred in multiple turns and was generally more constructive, compared to that of the first week. In other words, there were more complete ESRU cycles through which the conversation allowed the teacher to elicit and respond to evidence of students' NOS learning. In this regard, Dr Kim's earlier practices to focus the lesson on the learning goal can be viewed as have had some effect in making the peer feedback beneficial to both assessors (i.e., peers) and assesses (i.e., presenters) (Topping, 2013). This suggests that clear goal-setting could be an important prerequisite for effective peer feedback to occur in NOS classrooms.

Several episodes suggest that the assessment conversation did not go as imagined by the teacher, which required her to make instantaneous adaptations and responses. For example, when it turned out the Mpemba effect example did not meet the condition of a paradigm shift, she did not close down the discussion by simply denying the case. Instead, she links the story to other NOS features such as confirmation of hypotheses by evidence and the reliance of new discoveries on the existing knowledge. The teacher's ability to be adaptive about their lesson plan is crucial in this paradigm shift lesson for two reasons: first, because the lesson was led by students making presentations, and second, because it was about NOS, which is usually unfamiliar to both the teacher and students. This suggests that flexibility is the key when NOS is discussed in whole-class conversations with various examples. In Dr Kim's case, she mentioned that her flexibility improved as she taught the lesson to more classes, but the prerequisite for such development through experience would be the acquisition of broad prior knowledge in NOS itself.

From the broader interest of this thesis, it is also noteworthy that there were several moments where her formative assessment practices were interacting with her considerations for summative assessment. Her choice of the assessment task was motivated by a summative factor that the assessment policy for the subject allowed flexibility to teachers, and so she could use more time for whole-class discussions which allowed for a deeper understanding of the presentations. Conversely, her feedback practices during the discussions sometimes had the intent to deemphasise the summative component (e.g., grading) of the assessment task but emphasise the learning component. This was shown when she reassured Group 3 about the impact of their failure to answer the detailed scientific question on grades. Taken together, Dr Kim's paradigm shift lesson shows examples of the diverse relations between summative assessment, formative assessment and instruction (Black & Wiliam, 2018; Dolin et al., 2018) in the context of NOS. Further investigations on these relations will be essential to improve NOS instruction in classrooms by strategic use of formative assessment, in conjunction and alignment with summative assessment for accountability (see Section 8.7 for further discussion).

7.5 Chapter summary

This chapter highlighted the instructional strategies found in Dr Kim's lesson on the paradigm shift and the development of science, using the lesson videos along with her reflections during the follow-up interviews. Specifically, I focused on how her convergent and divergent instructional strategies allowed ongoing formative assessment of students' NOS learning as she planned and orchestrated whole-class discussions following students' group presentations. As a result, I identified four categories of instructional strategies that allowed the teacher for the formative assessment of NOS, where she asked students questions, noticed students' level of understanding, and provided feedback and create spaces for NOS learning. I discussed the affordances as well as challenges in the teaching and formative assessment of NOS. From these findings, I asserted that, in a NOS classroom, it is crucial for teachers to use formative assessment practices appropriately to keep the classroom conversation focused on the target NOS and simultaneously create opportunities to broaden the NOS learning to other related contexts.

Chapter 8 Conclusions and discussion

So far, this thesis has focused on the classroom assessment of NOS as being key to effective and meaningful NOS instruction in secondary school science. Different aspects of teacher views and practices presented in the previous three chapters have revealed how the teachers enacted their NOS assessments in the contexts of their instructional beliefs, curriculum requirements, and the broader sociocultural settings in which the instruction and assessments were situated. Specifically, Chapters 5 and 6 considered the summative assessment of NOS, with the former focusing on the individual teachers' ideas and enactments of NOS assessment and the latter on issues arising from the collaborative summative assessment over a semester. Chapter 7 looked into Dr Kim's lesson to illustrate how questioning and feedback strategies can be used to facilitate NOS learning during the course of instruction. These chapters have shed light on how NOS can be assessed in the classroom as well as the difficulties of it. Given the scarcity of studies on the classroom assessment of NOS (Hanuscin et al., 2020), the findings have much potential to inform NOS research and practice, particularly considering that teacher assessment is gaining increasing importance in many parts of the world (e.g., HM Government, 2021).

This thesis has originated primarily from my concern as a science educator that little research exists to inform practice in the classroom assessment of NOS. During the two years and a half whilst it was being written, however, the importance of NOS and classroom assessment have both soared with the continuing impact of COVID-19. The spread of misinformation, distrust of science owing to scientists' misconduct and errors, and proliferating vaccine hesitancy—and all these being exacerbated by social media—have called for increased awareness of what science is and how it works (Reiss, 2020). The current crisis in public trust in science should be taken as an opportunity to cultivate a critical understanding of NOS as opposed to rote learning of isolated ideas about science (Yacoubian, 2015). Meanwhile, debates on classroom assessment have never been as intense, especially in those countries where national high-stakes examinations were cancelled, such as the UK (Benn, 2020). As teacher assessment will replace GCSEs, AS- and A-levels, it has become a keen interest of all stakeholders how to 'ensure these [teacher-assessed grades] awarded fairly and consistently across the country' (Whittaker & Booth, 2021), as put by the UK Education Secretary. The recent exam crisis in the UK has also shed light on the complexities around what is a valid, reliable, fair and equitable assessment and fuelled the debate about the current exam-based assessment system (Woolcock, 2021). Taken together, these social changes in these years lend urgency to researching the classroom assessment of NOS as a way to respond to the changing needs of society.

In this chapter, I first highlight four key contributions of the thesis. I then revisit the literature to reflect on my findings and address several core questions around NOS and classroom assessment. In doing this, I start with broader questions about the classroom assessment of NOS such as whether it would be possible and how it compares to assessing scientific content knowledge, and then move to specific issues that were raised in the previous chapters as well as some implications and limitations of the findings. Given the exploratory nature of the current study, the findings may not provide many definite answers to these questions; Instead, my intent is that they shed light on some salient aspects of NOS assessment that were previously unnoticed or unevidenced. In particular, comparing the literature base and the empirical data (i.e., teachers' views and practices) will provide further insights into the potential discrepancies between teachers' beliefs and practices as well as how to resolve them. By organising the discussion in terms of key questions, the contributions of this thesis and its implications for research and practice will become clear.

8.1 Key contributions of the thesis

8.1.1 Shedding light on classroom assessment as a key element of NOS instruction

This thesis started with problematising the fact that classroom assessment has not been the focus of NOS research, although it was suggested as one key component of NOS learning (Bell, 2008; Ryder, 2009). Acknowledging the role of classroom assessment in NOS instruction, this study initiates further discussion on how NOS should be assessed in the classroom, both formatively and summatively, to facilitate explicit instruction of NOS (Abd-El-Khalick & Akerson, 2009; Akerson et al., 2000). I have highlighted the role of assessing NOS in the classroom in providing teachers with evidence of student learning that they can act upon. Drawing on the domain-specificity of classroom assessment (Black et al., 2004), I have focused on the significance of classroom assessment as opposed to NOS assessment for research purposes, on the grounds that the former has unique aims and roles for NOS learning that are distinguished from those of educational research and are dependent on the instructional context. Given the increasing need for separate consideration of NOS classroom assessment by teachers (Hanuscin et al., 2020; Ryder, 2009), the themes and further questions developed in this thesis will be useful to guide future research on the classroom assessment of NOS.

8.1.2 Linking NOS with classroom assessment theory

While there exists a strong research tradition in the classroom assessment of science (Black & Wiliam, 1998; Ruiz-Primo & Furtak, 2007), how the issues of classroom assessment can unfold when the lesson goal is NOS has gained minimal attention. When considering summative assessment, I have discussed how the context of accountability interacted with the open-ended nature of NOS questions and the teachers' views of classroom assessment. For formative assessment, the case study illustrated the teacher's use of a planned formative assessment activity and formative assessment conversations to make students' NOS learning visible in a whole-class discussion setting, utilising the ESRU framework as well as inductive analysis methods. Episodes have illustrated how concepts such as validity, reliability and discrimination can be related to NOS summative assessment, and also what teachers' use of questioning, noticing, feedback and peer assessment may look like in the context of NOS. In view of the domain-specificity of classroom assessment (Black et al., 2004) and that of teacher knowledge (Hanuscin et al., 2011), these findings provide insights into the features of assessing NOS compared to assessing science, which have been little discussed in the NOS literature. These demonstrate the value of classroom assessment as a useful lens to improve research in NOS instruction and an important avenue for future research. From the perspective of classroom assessment, the study offers a unique case study of teachers' classroom assessment practices in an area that they are not familiar with teaching or assessing. In this sense, the case study of teachers also contributes to the literature on how teachers respond to the challenges with classroom assessment in the face of a new curriculum reform (e.g., Suurtamm, Koch & Arden, 2010; Willis, Adie & Klenowski, 2013; Yung, 2006).

8.1.3 Investigating teachers' NOS classroom assessment practices empirically

Given that NOS is rarely assessed in the science classroom (Abd-El-Khalick et al., 2013), the three teachers' NOS assessment practices documented in this thesis would be a rare and useful contribution to NOS research. Characterising and detailing the teachers' practices related to NOS assessment can help teachers, teacher educators and researchers make sense of what is observed in the classroom. I have looked into teacher practices at three different levels, each with a focus on individual teachers (Chapter 5), a team of three teachers (Chapter 6), and a single teacher (Chapter 7). The first two were focused on the summative assessment, and the last on the formative assessment. Through empirical investigation of these multiple aspects of NOS classroom assessment, I have shed light on teachers' practices as well as nuanced their views and decision-making process, which carries implications for improving NOS assessment practices in the classroom. For example, the discrepancies between teachers' NOS teaching intention and NOS assessment intention, and between their NOS assessment views and science assessment views, suggest the need for differentiating these constructs in research and supporting teachers to develop coherence in their views. Chapter 7 contributes to the knowledge of how meanings about NOS are co-constructed through everyday classroom interactions, which is a valuable addition to the NOS literature that is mainly focused on surveys and interviews (Kelly et al., 1998). By considering the context NOS, the findings expand the existing discussions on formative assessment in the context of science content (e.g., Bell & Cowie, 2002; Nieminen et al., 2020) and scientific inquiry (e.g., Cowie & Harrison, 2020; Ruiz-Primo & Furtak, 2007). Moreover, the characterisation of NOS instruction as 'connection building' between different elements of NOS (Chapter 7) coheres with the recent rise of the epistemic network approach to NOS learning (Mulvey et al., 2021; Peters-Burton et al., 2019) and shows how such approach can be implemented through teachers' formative assessment practices.

8.1.4 Identifying the challenges of NOS classroom assessment

A closer look into teacher practices not only allowed understanding what they did but also provided the practical challenges involved in NOS classroom assessment, which stemmed from diverse reasons. By revealing the challenges that teachers were faced with in a new, NOS-oriented curriculum, this thesis contributes to understanding the potential barriers to effective NOS instruction in the high school classroom. Given that most existing studies are focused on the challenges of NOS teaching rather than its assessment (Höttecke & Silva, 2011; Lederman, 1999; Wong, Kwan, Hodson, & Yung, 2009), this thesis's novelty is in providing initial empirical evidence specifically on the challenges of NOS assessment. The challenges identified in the study in turn have the potential to inform the future actions to be taken to best support the teachers' NOS assessment in the classroom.

8.2 How can NOS be assessed in the classroom?

Although research has profoundly shown that NOS can be *taught* to students across all grade levels, evidence is scarce as to whether and how NOS can be *assessed* in the classroom by teachers. As a recent review of literature by Hanuscin et al. (2020) noted, questions about the classroom assessment of NOS have been little discussed or investigated in the context of classroom instruction. Given the broad agreement that NOS is a form of knowledge and a cognitive outcome (e.g., Abd-El-Khalick, 1998), it seems reasonable to assess in the classroom whether and to what level the cognitive learning goals about NOS have been achieved. The findings of this study indicated, however, from the teacher's perspective, there are a number of practical issues related to the possibility of assessing NOS in the science classroom.

As illustrated in Chapter 6, the teachers were faced with a number of theoretical and practical challenges that raised questions about the possibility of assessing NOS. These included out-of-rubric answers, the tension between views and understandings, different enactments of activities, and the demands for reading literacy and scientific content knowledge. The teachers in this study were able to, to some extent, circumvent these issues by adopting a generous marking scheme where all but 'absolutely wrong' student answers were equally counted for marks. Specifically, they agreed that it was almost impossible to assess the *quality* of each answer, so they developed a marking system that did not require the teachers to struggle with individual answers' quality. On this occasion, the teachers' approach worked without causing serious problems partly due to the low accountability pressure for the subject. However, such a strategy might not be possible in circumstances where stakes and demand for accountability are high. Imagine that NOS is included in Korean students' school-leaving grades as importantly as usual subjects, or as a more extreme case, included in the high-stakes external examinations[23]; We would be unable to avoid facing the issue of whether and to what extent NOS is assessable. This difficulty relates to the belief that 'there is no right answer about NOS'—How can we assess NOS if any ideas about it are correct? More theoretical and practical considerations are called for to distinguish between allowing multiple answers and allowing any answers.

There were several other concerns related to the possibility of assessing NOS in terms of summative assessment, such as validity and reliability. The teachers' concerns about the risk of NOS tests being about tests of drawing skills, content knowledge and reading skills are exactly on this point, since these were not learning goals but still influenced student answers. To some extent, these issues are inevitable to arise, given the contextualised and interdisciplinary nature of NOS knowledge. In this sense, we may see NOS as consisting of *rich knowledge* (Davis, 1995) that is 'interconnected and many layered'. Halliday (2010) rightly pointed out the conflict between validity and reliability when assessing such rich knowledge. For rich knowledge to be assessed, the criteria for marking should be 'open to interpretation and exemplar resources should be widespread' (p. 371) so that the assessment should not be a test of mere memory of facts. This validity requirement, however, can be met at the cost of reliability since it becomes more difficult to assess consistently as diverse answers are permitted (Halliday, 2010) and develop a rubric with clear criteria. Such a conflict means that a key task in designing NOS assessments is to balance validity and reliability. These issues will need to be addressed by developing and providing NOS assessment resources that can balance the conflicting quality criteria (Hanuscin et al., 2020; Ryder, 2009)

---

[23] The idea of including NOS in high-stakes testing was suggested by Clough and Olson (2008), who asserted that it would give a clear message about what is important and what needs to be emphasised in the instruction.

When it comes to the formative assessment of NOS, Dr Kim's classroom practices showed that an experienced teacher with knowledge of NOS can effectively use formative assessment strategies in the classroom. Her lessons showed how NOS learning can be facilitated by the teacher's use of instructional tasks (i.e., formal) as well as questioning and feedback (i.e., informal), and how these practices of formative assessment are instrumental in 'making connections' across content knowledge, NOS and history of science, which has been stressed as a key element of NOS learning recently (Mulvey et al., 2021; Peters-Burton, Parrish & Mulvey, 2019). In her lesson, there were some positive indicators of students' learning, such as students' peer feedback on others' presentations became more frequent and critical in the second session, which suggests that they took responsibility for assessing their and their peers' learning. This shift in authority regarding assessment in her lesson is important because it may be indicative of students' increased confidence, willingness and ability to participate as active agents in the lesson (Heritage & Heritage, 2003). Nevertheless, there were still some challenges involved in this process, mainly because it requires the teacher an extensive knowledge of all these domains to enact prompt and responsive formative assessment throughout the instruction. Given that Dr Kim was more knowledgeable and experienced in NOS than many teachers, it is likely that these difficulties would be more problematic for other teachers.

Lastly, if NOS is to be assessed in the classroom, an essential condition is that the teacher has knowledge and skills required for NOS assessment, in addition to those in science assessment (see Section 8.10 below). The teachers in this study had different levels of NOS backgrounds, which was also reflected in their views about NOS teaching and NOS assessment. Dr Kim's feedback and questioning practices in the paradigm shift lesson support the possibility of the formative assessment of NOS in the classroom when the teacher has understandings of NOS and its pedagogy. The design of the three NOS assessment tasks used by the teachers in Chapter 5 and Chapter 6 also indicates some possible ways in which classroom assessment can promote NOS learning through engagement in different learning activities. Nevertheless, when the issue of grading and accountability emerged, the teachers struggled with various problems in the summative assessment of NOS (Chapter 6). While there are no simple, clear-cut solutions to these challenges, they could be reduced through the provision of NOS assessment resources and teacher professional learning opportunities.

8.3 Why are teachers not sure about assessing NOS?

Having discussed how NOS can be assessed in the classroom, I now consider some possible reasons that teachers are not convinced about assessing NOS in the classroom, particularly for the summative assessment. Previous studies on teachers' instructional practices in NOS have suggested, albeit anecdotally, that teachers do not usually consider including any formative or summative assessments in their NOS instruction (Abd-El-Khalick et al., 1998; Brock & Taber, 2019; Hanuscin et al., 2001). This thesis's findings offer some further insights into science teachers' views on the necessity of NOS classroom assessment and the backgrounds to their views. Interviews with Ms Lee and Ms You suggested that teachers can hesitate to assess NOS because they do not think assessing NOS is necessary, even when they agree with the idea of including NOS as an instructional goal. In other words, because they do not think of NOS as something to be assessed in the classroom.

How, then, can we decide if something needs to be assessed in the classroom? The most basic answer to the 'what to assess' question is that teachers assess student achievement against the standards set by the curriculum (Klenowski & Wyatt-Smith, 2014). Under usual circumstances where NOS is not included as a learning goal in the curriculum or presented as secondary to content knowledge, teachers' lack of recognition is understandable. However, the curriculum in this study did explicitly present NOS as its learning goal. Given this context, the fact that Ms Nam and Ms You remained largely unconvinced of the value of assessing NOS contradicts the curriculum's explicit focus on NOS as achievement standards. This inconsistency suggests that teachers' views about what to teach about NOS and what to assess about NOS can be different, due to their beliefs about the necessity of assessing it.

One important factor regarding the possibility and worthiness of NOS assessment seems to be the issue of distinguishing 'views' and 'understandings'. The popular VNOS questionnaire starts with an instruction that says, '*There are no "right" or "wrong" answers to the following questions. We are only interested in your opinion on a number of issues about science*' (VNOS-C; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). However, after marking students' VNOS worksheets, Dr Kim remarked that there *were* certainly wrong answers, for example, when a student writes that 'science is absolute' (Informal conversation, July 2020). Teachers aim for students' cognitive understanding of NOS, teach the key aspects of NOS, but if these are still about opinions rather than knowledge, how can we justify marking and grading these opinions? Do 'informed' answers deserve more marks and thus a higher grade, which can have implications for students' future lives? Hanuscin et al. (2020) rightly pointed out that the focus on *views* of NOS 'is inconsistent with tasks with which teachers are charged—the summative evaluation of student learning and assigning of grades' (p. 413). The seeming conflict between interpreting NOS as views (perspectives, opinions) and understandings requires further considerations to support teachers to develop rationales for assessing NOS in the classroom, particularly for summative assessment. The teachers in this study were cognizant of this issue while engaging in the NOS assessment, but their decision to mark generously allowed underplaying this issue at the expense of fine discrimination among students' NOS understandings. The low examination pressure for the subject was an important factor that allowed such a decision. This decision can be understood as a practical rather than logical choice, but it can be criticised in that generous marking and low level of discrimination mean that the assessments provide limited information about student learning in NOS, which is a crucial purpose of summative assessment (Black, 1998).

My view is aligned with Allchin's (2013) that understandings, instead of views, should be the focus of NOS summative assessment. However, I think that VNOS and similar open-ended questionnaires can still be used as a summative assessment to the extent that it elicits students' *understandings* of NOS (which I consider as the object of NOS assessment; see Section 2.5.3) in writing. The assumption here is that the teacher taught the target NOS aspects as an explicit lesson goal before VNOS is administered as a summative assessment. Otherwise, a summative assessment with VNOS would be decontextualised and detached from the instructional context and the teacher's instructional goals. When treated as a matter of an understanding rather than a view, the summative assessment of NOS can be justified more easily. I also suggest that the development of LPs for NOS (Section 2.3.4) can be helpful to offer an alternative approach to grading NOS. Students' NOS understandings can be marked against the levels specified in LP for each NOS aspect, rather than as naïve, mixed or informed. To use Abd-El-Khalick's (2012) example about the tentative NOS, students can start with the simple version of the idea that 'Scientific knowledge is subject to change over time', and then progress into more sophisticated understandings such as the different ways in which such changes occur (e.g., expansion through accretion, replacement with new knowledge) and the main reasons for changes (e.g., new evidence, theoretical advances). Therefore, I acknowledge the possibility of VNOS being used as a summative assessment with adaptations, but at the same time, efforts to develop other methods of summative assessment in addition to questionnaires will need to be made.

8.4 How does assessing NOS compare to assessing science?

In Chapter 5, I discussed that the teachers' views of NOS assessment differed from those of science assessment. While they commonly put emphasis on the process or product of problem-solving for science assessment, for NOS assessment, the teachers had varying foci such as the application of NOS understandings (Dr Kim), participation and attitudes (Ms Nam) and transforming the assessment into collective learning (Ms You). This perceived difference between science assessment and NOS assessment relates to the domain-specificity of classroom assessment that I have discussed earlier (Section 0) and its possible implications for NOS assessment. It means that the aims and methods of classroom assessment are different across school subjects and the specific content being taught (Black et al., 2003). There is evidence on the differences between science content knowledge and NOS in terms of teachers' perceptions and pedagogical approaches (Hanuscin, 2013; Mesci et al., 2020), but whether and how the two differ in the context of classroom assessment have been little discussed. In this section, I consider some aspects of this issue by reflecting on my findings and relevant literature.

The findings in Chapters 5 through 7 illustrated that the differences between scientific content knowledge and NOS played an important role in shaping the teachers' perceptions and practices as well as giving rise to some of the challenges that they faced. The teachers generally felt that learning NOS could complement the content-focused curriculum by promoting critical thinking and science communication skills, suggesting their awareness that NOS can serve different goals than developing scientific content knowledge. However, when it comes to summative assessment, two teachers, Ms Nam and Ms You, were unsure of whether NOS needs to be subjected to summative assessment. This suggests that the teachers considered NOS to be different from science content in terms of assessment, but they did not have a clearly articulated view of the differences. Dr Kim, who had prior knowledge and experience in NOS instruction, viewed that NOS summative assessment should focus on the application of NOS knowledge and suggested an oral test where students talk about NOS ‘in their own words’ as her preferred assessment method. Such a contrast suggests that developing teachers’ awareness of how scientific knowledge and NOS knowledge differ can be key to their coherent ideas about the classroom assessment of NOS.

In Chapter 5, the perception of NOS as a matter of an ‘opinion’ was one central consideration in the team’s decision making about assessment, which shows another contrast with scientific content knowledge. This in turn raised the question of how to distinguish the right answers from the wrong ones, particularly for the VNOS task where the instruction clearly stated that there are no right or wrong answers. The challenges of marking (arguably) ‘no-right-answer’ questions show that science teachers, quite expectedly, are not used to such type of marking. Such challenges recurred for the other tasks, whenever the question was open-ended and not all student answers had been predicted by teachers. Meanwhile, Chapter 6 showed that the formative assessment of NOS during instruction can be challenging and requires the teacher to be knowledgeable about content knowledge, history of science, NOS and have the ability to interconnect these elements throughout the lesson. This clearly illustrates that formative assessment skills for NOS are not naturally derived from the skills needed for science content knowledge and calls for increased attention to the specificities of NOS assessment.

Teaching NOS and teaching science have both common and different aspects, and so do the classroom assessment of science and that of NOS. Broadly speaking, both science and NOS are taught (and assessed) to contribute to students’ scientific literacy, the goal of science education. However, there are differences, too. Not only does NOS teaching require a different set of PCK than science teaching (Abd-El-Khalick & Lederman, 2000; Hanuscin et al., 2011; Mesci et al., 2020), it also poses new challenges to the teachers. Höttecke and Silva (2012) illustrated the difference between teaching science and teaching NOS in the case of physics, noting that teachers struggle to shift from the disciplinary culture of physics, where the focus is on physics as a collection of truths about nature, to that of NOS where physics is portrayed as ‘a matter of empirical investigation, discourse, and negotiation among scientists that result in knowledge that has changed and may change in the future' (p. 299). Despite these differences, assessing NOS cannot be considered in isolation from assessing science because NOS cannot be separated from scientific knowledge and inquiry skills (as suggested in Chapter 6), and importantly, because they are assessed by the same teacher. Further theoretical and empirical investigations into the relationship between science assessment and NOS assessment will be a crucial step in promoting teachers' knowledge and skills in NOS assessment. For example, studies with a larger number of teachers to explore how teachers' science PCK and NOS-PCK for classroom assessment are related will be useful to further develop some of the issues pertaining to the science-NOS contrast raised in this thesis.

8.5 How should curriculum, instruction and assessment align in NOS lessons?

Having discussed different aspects of NOS classroom assessment, let me now turn to how it can relate to other elements of teaching and learning in the classroom, such as curriculum and instruction. The alignment between curriculum, instruction and assessment, or how these three elements of an educational system operate in close cooperation to be mutually supportive, are integral to effective classroom teaching (Tyler, 1949). Ideally, what is set out in the curriculum standards need to cohere with what students are taught and tested on. Such an alignment is particularly important in this age of accountability to provide stakeholders with a clear message about what is expected (Martone & Sireci, 2009). Moreover, achieving the coherence of curriculum and instructional and assessment practices is particularly challenging for teachers in the time of a national curriculum reform where new learning goals such as NOS are introduced (Ryder & Banner, 2013).

Although the SIE curriculum made progress from earlier national science curricula in Korea where NOS had been stated only as a higher-level goal, it still provided no information about how to discern different levels of student learning in the NOS aspects, and how student learning is expected to develop in each NOS domain. The difficulties of marking that the teachers encountered in Chapter 6 partly stemmed from the absence of agreed standards in terms of what level of understanding is desired about individual NOS aspects and how to evaluate whether it is met. Similarly, the three teachers' different implementations of instruction illustrated in Chapter 5 can also be understood in relation to the lack of agreement about the desired NOS learning outcomes. To address these challenges and inconsistencies, the provision of more specific learning goals about NOS and criteria for deciding the achievement in the curriculum will be necessary.

One possible way to enhance the curriculum-instruction-assessment alignment is through the use of LPs (Duncan & Hmelo-Silver, 2009) that are specifically developed for NOS. LPs are aimed at 'successively more sophisticated ways of reasoning within a content domain that follow one another as students learn' (Smith, Wiser, Anderson & Krajcik, 2006, p. 1) and have been suggested as a promising framework for developing meaningful assessments in science and beyond (Alonzo & Steedle, 2008; Duncan, Rogat & Yarden, 2009). LPs for key NOS domains (e.g., aims and values, practices, methods, social dimensions) can provide teachers useful guidance to plan their instruction, map formative assessment opportunities, and assess students' NOS learning summatively. Recently, some potential benefits of NOS LPs have been suggested by NOS researchers (Abd-El-Khalick, 2012; Akerson et al., 2011; Bell & Clair, 2015; Duschl & Grandy, 2013; McComas, Clough & Nouri, 2020). For example, Abd-El-Khalick (2012) called for an approach where NOS could be 'addressed at increasing levels of depth as learners move along the educational ladder from elementary school to college-level science teacher education programmes' (p. 1027).[24] The development and empirical validation of NOS LPs can draw on the rich existing research on students' (mis)perceptions of NOS at different grade levels, as well as research on the effect of different instructional interventions on NOS. Empirically grounded LPs about major aspects of NOS will provide practical guidance to teachers compared to what informed, mixed and naïve categories can offer, and also increase the coherence between instruction and assessment, and between formative and summative assessment. Besides,

[24] Note that Abd-El-Khalick's (2012) argument for learning progressions is still based upon the consensus list. While I concur with his idea of a progressive approach to NOS learning quoted here, my view is that these progressions should not be limited to the NOS aspects in the consensus list (see Section 2.5).

through the use of LPs, teachers can elicit evidence about the gap between students' current NOS understanding and the desired level, provide feedback based on the identified gap, and also possibly involve students in the formative assessment in the form of self-assessment and peer assessment (Heritage, 2008).

8.6 How can content knowledge and NOS learning goals be balanced in classroom assessment?

In addition to the alignment within the curriculum, instruction and assessment, NOS learning goals should be well aligned and balanced with other non-NOS goals such as conceptual learning of scientific content knowledge. Science curricula have traditionally had an emphasis on conceptual learning, and many science curricula and standards worldwide still have a strong emphasis on key concepts of each science discipline (Akerson, Carter, Pongsanon & Nargund-Joshi, 2019; Murakami & Sumida, 2014). While the profound value of conceptual learning should not be dismissed, science education researchers have made consistent efforts to incorporate the epistemic aspects of science as a goal of science education (e.g., Kelly & Duschl, 2002; Schwab, 1962). How can the two goals be balanced in assessment practice? Duschl (2008) suggested the relation between conceptual and epistemic learning to be symbiotic rather than an either-or situation, and they should reinforce and mutually establish each other. In particular, teacher-designed activities and assessment practices should be able to monitor learning and provide feedback for both learning goals (Duschl, 2008).

The summative and formative assessment practices illustrated in this thesis suggest that NOS learning goals, even when it is targeted in the lesson, can be easily obscured or downplayed by conceptual goals unless the teacher makes conscious efforts to emphasise the former. In the case of summative assessment (Chapter 5 and Chapter 6), even though the teachers included worksheet questions to elicit students' reflections about NOS from their activities, these questions turned out to bear various issues relating to validity and reliability. This led the teachers to downplay the discriminating function of the activities by adopting a generous marking scheme. In the case of formative assessment (Chapter 7), some of the student presentations were focused on describing the content of scientific discovery (that is, the conceptual aspect) without interpreting it in terms of the paradigm shift idea (that is, the epistemic aspect). Students' questions after the presentations were also often focused on the specific aspects of the discovery (e.g., Kepler's law as evidence of the heliocentric model, the critical density in the models of the universe) that sometimes exceeded the science curriculum, and the teacher had to shut down the Q&A and refocus it onto the NOS aspects of the discovery, by providing feedback or asking questions specifically targeted at NOS. Together, these episodes point to the importance of teachers in striking a balance between conceptual and NOS learning goals. To tackle this, efforts will be necessary to support teachers to develop a clear idea about the difference between science and NOS as learning goals and acquire instructional skills to teach NOS explicitly through professional development efforts.

8.7 How do the summative assessment and formative assessment of NOS relate?

Given that this thesis addressed both formative and summative assessment of NOS, it is worth briefly discussing the relation between the two purposes of classroom assessment in the context of this study. When formative assessment and summative assessment are understood as the dual purposes of classroom assessment rather than different forms of it (see Section 2.2.1), these two purposes can be complementary in classroom instruction (Looney, 2011). However, the two functions of assessment do not always work in service of each other. For example, when a classroom summative assessment involves high stakes, its 'washback effect' that negatively impacts learning has been recognised (Dolin et al., 2018). Several moments in this study seem to exemplify such an antagonistic relationship between formative and summative assessment. The teachers' decision to use diverse assessment methods that were based on differently contextualised student activities, rather than paper-and-pencil assessments, is reflective of their focus on the formative functions of NOS assessment. These activity-centred assessments that were aimed at the promotion of learning often conflicted with the need for grading. An example is when Ms You regretted that she had not given much feedback to students during activities due to her concerns about fairness, after finding out other teachers had given feedback. As another example, the teachers could prioritise the learning of NOS in the design of assessments due to the reduced accountability pressure for the subject in Chapter 6, and this reduced pressure was also why Dr Kim could allocate enough time for student presentations in her lesson in Chapter 7.

Although tensions arose on several occasions, overall, the teachers felt that they were generally successful in facilitating student learning from the assessment activities as well as achieve a fair level of accountability. This suggests that they managed to strike a balance, to some extent, between the two main aims of classroom assessment—promoting learning and producing evidence of student achievement. Specifically, the NOS activities described in Chapters 5 and 6, and the paradigm shift activity in Chapter 7, illustrates how the summative and formative assessment of NOS can be complementary in instructional practice. For example, Dr Kim's formative assessment practices constantly make connections between different NOS aspects, between different historical examples and scientific content knowledge. Given that the facilitation of connections is increasingly recognised as a core aim of NOS learning (Mulvey et al., 2021; Peters-Burton et al., 2019), her practices show a good example of how it could be done in the classroom. To understand NOS, learners require a 'rich understanding' (Davis, 1995) of it; Simply memorising the tenets is far from 'understanding' NOS (Allchin, 2013; Matthews, 2014; McComas, Clougn & Nouri, 2020). In this sense, recognising the interconnections is something that was well supported in Dr Kim's formative assessment during the lesson, but not as well in the case of summative assessment. Given that the teachers marked each question separately and then aggregated the marks, it was difficult to discriminate student understandings of NOS based on the connections between different elements of NOS learning.

8.8 How should controversies about NOS be addressed?

Having discussed the issues of classroom assessment, let me now turn to an intrinsic issue related to the nature of NOS knowledge. It is well recognised that NOS has its roots in historical, philosophical and social studies of science. The process in which knowledge of NOS is generated is, therefore, more akin to the methods of the humanities and social sciences than those of the natural sciences. For this reason, claims about NOS are often controversial, and there exist several competing views about a single aspect of NOS. An example from Dr Kim's lesson is whether Watson and Crick's discovery of DNA double helix was 'paradigm-shifting' that suggested a new paradigm, the answer to which relies on the concept of paradigms and scientific revolutions that has been intensely debated since Kuhn's proposal (Cohen, 1985; Masterman, 1970; McMullin, 1993; Sankey, 1993; Shapere, 1971; Toulmin, 1972) and also revised by Kuhn himself several times in subsequent works (Kuhn, 1974; Kuhn, 2000). Likewise, NOS researchers have argued that most of the tenets in the NOS consensus list are in fact more controversial than they appear (Allchin, 2013; Hodson, 2014).

Kötter and Hammann (2017) argued that NOS teaching should address the controversies about the scope and limits of science because controversies are at the core of philosophy, history and sociology of science on which NOS is based. However, science educators have been divided about the inclusion of controversial aspects in NOS teaching. The consensus view, as self-evident from the name, regards controversies as 'impractical' and 'irrelevant' at the level of pre-university science education (Abd-El-Khalick, 2012; Lederman, 2006) so have less priority in deciding what aspects of NOS to include in the science curriculum. This claim has attracted criticism from NOS researchers, who argued that science cannot be properly described in a list of simple statements due to its inherent heterogeneity (Hodson, 2014; Irzik and Nola, 2011; van Dijk, 2011). In this sense, Matthews (2014) criticised the attempt to 'list up' the NOS and argued that NOS instead needs to be 'elaborated, discussed and inquired about' in the science classroom (p. 18). He argued that there are questions about NOS that can be critically reflected on and discussed in the classroom without necessarily engaging in higher-level philosophical debates. He suggests, rather than aiming to teach the mere tenet that 'scientists use experiments', teachers should discuss and explore various features of scientific experimentation with historical examples such as Newton's pendulum experiments in mechanics and prism experiments in optics.

I would argue that one promising solution to the question of controversies versus consensus in NOS is to consider them not as an either-or choice but as a matter of progression. In other words, the idea of LPs introduced earlier (Section 2.3.4 and Section 8.5) can be beneficial here, too. At the lower grades, students can be introduced to the idea that 'science is empirically based' and explore how Darwin developed the theory of evolution based on his empirical observation of finches in Galápagos, and how British scientist Luke Howard formed useful knowledge about the types of clouds by observing and classifying numerous clouds according to the altitude level and shape. The similarities of these cases in terms of their empirical element can be highlighted at this level. As a student progresses into higher grade levels, the NOS idea can be elaborated and expanded by changing the focus to variations, specificities and exceptions. The idea of empirical basis in science can be challenged by introducing scientific achievements where conceptual and non-empirical processes played a key role, as in the examples of Galileo's thought experiment about a ball falling down a slope without friction that led to the idea of inertia, and Einstein's imagination of chasing a light beam which was instrumental to his discovery of special relativity. Through these examples, students can learn a more elaborate notion that scientists sometimes make progress by transcending the empirical experience. When carefully organised, this approach can avoid both oversimplifying NOS and burdening teachers and students with cutting-edge philosophical debates.

8.9 How can we teach and assess the domain-specific aspects of NOS?

Part of the controversies about NOS relates to the domain-general and domain-specific aspects of NOS. What, if any, are some characteristics shared among the sciences and what others are only applicable to some domains within science? What are the implications of this question for NOS education? It has been not only the interest of philosophers of science (Cartwright, 1999; Galison & Stump, 1996; Kellert, Longino & Waters, 2006) but also NOS educators (Duschl & Grandy, 2013; Erduran & Dagher, 2014a; Hodson, 2014; Irzik & Nola, 2014). Specifically, NOS researchers have debated whether to address the domain-specific aspects of NOS to students. Here again, the fact that there are NOS aspects specific to some domains itself is not controversial; The debate is rather about how it is taught at the school level. Consider the use of experiments as a method and practice of science. Although experiments were suggested as a key method of knowledge acquisition at least since the seventeenth century by

Francis Bacon and continue to be one, it is not all scientists who conduct experiments.[25] For astronomers and geologists, the massive scale of time and space they deal with makes it impossible to conduct experiments in a controlled environment, although they instead count on observational methods. Meanwhile, there are even subfields of science that develop with little empirical evidence. For example, the work of string theorists that is nearly impossible to test through experiments is still considered a major branch of physics. The overemphasis of experimentation in NOS as *the* scientific method can thus obscure the role of historical, observational and theoretical sciences in the development of knowledge. These domain-specificities of NOS have invoked criticisms of the consensus view that stresses the 'general aspects' applicable to all sciences (Dagher & Erduran, 2014b; Duschl & Grandy, 2013; Irzik & Nola, 2014; Park & Song, 2019; Schizas, Psillos & Stamou, 2016).

Several excerpts in this study point to the practical challenges relating to domain-general and domain-specific NOS. In many countries, secondary science teachers have a background in one of the science disciplines and often are qualified to teach that discipline (Davies & Amos, 2017; Im, Yoon & Cha, 2016; Pederson, Isozaki & Hirano, 2017). It is therefore possible that teachers tend to be more familiar with NOS aspects that are common to all sciences or specific to their own discipline than those of other disciplines (Höttecke & Silva, 2012; Topcu, 2013). This possibility at least partly explains the struggles that the teachers had with the 'finding similar cases in the history of science' question in the worksheet. Alternatively, the teachers could have decided whether the case is similar to another specific case. Likewise, Dr Kim's paradigm shift lesson could have been easier for the teacher to handle if the task were less open-ended in terms of topics, given that part of the difficulty stemmed from the fact that the examples were drawn from multiple domains of science such as astronomy, meteorology, thermodynamics, and molecular biology. Instead, the teacher may instead give selected historical examples and ask students to evaluate whether each case has a

---

[25] Here I subscribe to Hacking's (1983) view that an experiment creates noteworthy and discernible phenomena through intervention, as opposed to an observation where the focus is either on phenomena in nature or the *result* of the intervention. The distinction between experiment and observation is a debated topic among philosophers of science (Daston, 2011; Gooding, 1992; Malik, 2017).

paradigm shift in it and why. Then the teacher can have some expectations about the possible student answers and also a better sense of how to respond to their answers (i.e., formative assessment) and how to mark and differentiate among the answers (i.e., summative assessment).

8.10 How could we support teachers to teach and assess NOS effectively and meaningfully?

In this thesis, I have argued that teachers' practices in NOS assessment are a missing piece in research in effective and meaningful NOS instruction, drawing on the review of existing literature on NOS. This study's findings indicate that the classroom assessment of NOS requires teachers to have additional knowledge and skills to implement this assessment effectively, and efforts are needed to support teachers in this regard. What kind of support can be provided to prepare teachers to assess NOS? The sociocultural approach to classroom assessment (Section 3.1) views it as 'a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledge with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students' (Willis et al., 2013, p. 242). To facilitate such a social practice, a teacher will need a good level of knowledge and experience in both NOS itself and how to assess it in the science classroom, let alone have general skills for classroom assessment. Here I will discuss some possible ways forward, based on the specific challenges and difficulties identified from this study and research literature in teacher knowledge in NOS and assessment.

Let us begin with the teacher knowledge for NOS classroom assessment. Hanuscin et al.'s (2011) and Mesci et al.'s (2020) conceptualisations of NOS-PCK both included knowledge of assessment as a component, but what constitutes teachers' knowledge of NOS assessment was not defined or specified in these studies. The framework for teacher assessment literacy by Abell and Siegel (2011) can be a useful starting point here. It has four main areas of teacher PCK for science assessment: knowledge of assessment purposes, knowledge of what to assess, knowledge of assessment strategies, and knowledge of assessment interpretation and action taking. At the core of these knowledge domains are teachers' ideas about assessment values and principles, grounded in their views of student learning (Abell & Siegel, 2011)

Nearly all of these components, from the views of learning to taking actions upon the assessment, need some adaptations to be applied to teachers' NOS assessment in the classroom, given that PCK is dependent upon the subject matter and that scientific content knowledge and NOS knowledge have different characteristics as subject matters (Mesci et al., 2020). An explicitly targeted NOS learning goal must influence the teacher's knowledge of assessment purposes and knowledge of what to assess. Specific strategies for assessing controversial, open-ended aspects of NOS should be developed and provided to teachers. Not to mention the question of what constitutes 'learning NOS' underlies all these considerations (Hanuscin et al., 2020).

The fact someone has good knowledge in science does not necessarily mean that they are good at teaching or assessing science (Gess-Newsome & Lederman, 1999; Shulman, 1987). The same idea should hold true for NOS. In this sense, although it is hopeful that an increasing number of science teacher education programmes and standards across the world are including NOS (Dibattista & Morgese, 2014; McComas, 2014; Song & Joung, 2014), these programmes must further include training on the instruction and assessment of NOS in addition to developing teacher candidates' ideas about NOS itself. Systematic supports to increase teachers' NOS-PCK and NOS assessment literacy should be provided besides discussing the aspects of NOS. McComas (2020) suggested four potential ways to incorporate NOS into teacher education: as part of the science methods course, by addressing NOS in science content courses, through engagement in authentic science research, and finally, through a special science education course or professional development solely focused on NOS. These four approaches can be pursued concurrently, starting from the first three that are relatively modest goals, and in the long run, reforming the science teacher education system to include NOS as a major goal by creating specialised courses and programmes.

What may also be beneficial is to engage science teachers in cross-subject learning opportunities with teachers of other subjects or those of other subject specialisms within science. Cross-subject professional development settings have proven to be effective in the context of SSIs for developing teachers' understanding of the content as well as pedagogical approaches (Christenson, Gericke & Rundgren, 2017). Given that science teachers have relatively low familiarity and competence in discussing controversial issues in the classroom (Erduran, Guilfoyle & Park, in press; Hand & Levinson, 2012), collaborative classroom assessment projects with teachers of social studies, ethics, religious studies and history can provide opportunities for developing assessment skills in the historical, philosophical and sociological aspects of NOS and the controversies around it. Dr Kim's focus on assessing what a student can talk about NOS 'in their own words' and Ms Nam's focus on assessing the reasoning process in the student answers about NOS indicate the possibility of fruitful exchange with literature and literacy teachers, who can offer insights on assessment of students' verbal and written expressions.

8.11 Future research on classroom assessment of NOS

8.11.1 Assessing diverse aspects of NOS

Due to the nature of this study as an observational rather than interventional one, the analysis has been limited to the NOS aspects that the teachers selected to teach. In Chapter 7, the focal NOS aspect was the idea of the paradigm shift, although Dr Kim attempted to build connections to NOS aspects that had not been brought up by each student group. For example, when one group presented plate tectonics as an example of a paradigm shift, she tried to link the historical case also to the role of technology in the advancement of science (see Section 7.3.4). Still, the overarching focus in her lesson was the paradigm shift idea and how it leads to scientific development which, in the FRA framework, belongs to the cognitive-epistemic categories. Likewise, the summative assessment activities used in the second year (Chapter 5 and Chapter 6) also featured mostly epistemic issues, such as observation and classification as scientific practices (the clouds activity), and diverse methods for making inferences (the scientific methods activity). The VNOS task included one question pertaining to the social and cultural aspects of science *('If you believe that science reflects social and cultural values [or is universal], explain why and how. Defend your answer with examples'*), but generally speaking, most assessment activities were focused on the cognitive-epistemic NOS aspects. Accordingly, my analysis has been focused on teacher practices in assessing these NOS aspects and the challenges in doing so. More research that considers the classroom assessment of social-institutional aspects will provide further insights as to how different categories of NOS knowledge can influence teacher practices in classroom assessment. This is an important avenue for further research, particularly in line with the increasing attention to the economic, political, social, ethical dimensions of science in NOS research (Erduran & Mugaloglu, 2013; Kaya, Erduran, Birdthistle & McCormack, 2018; OECD, 2020).

Just as some aspects of NOS can be taught more effectively with a particular instructional method (e.g., teaching the development of science with historical examples, and teaching the social aspect of science with SSIs), it can be postulated that suitable assessment methods may also differ for different NOS ideas. In future research, it would be helpful to consider the influence of different assessment methods such as written tests, oral tests, performance assessment and portfolio assessment on the quality of NOS assessment, as well as how can relate to the nature of NOS idea being taught and assessed.

8.11.2 Teacher knowledge, views and practices of NOS assessment

Three teachers' views of classroom assessment in relation to NOS illustrated in Chapter 5 offer some useful insights for future investigations in this area. The fact that teachers can take different approaches to science assessment and NOS assessment, for example, calls for research on the factors influencing such a disparity as well as how the two can be aligned coherently. Building on earlier works such as Mesci et al.'s (2020) PCK for NOS framework, and Abell and Siegel's (2011) science assessment literacy model, future research can consider developing a framework of teacher knowledge of NOS classroom assessment. A synthesis of these frameworks can lead to a useful theoretical account of teachers' NOS classroom assessment knowledge and views, particularly their components and the interplay between such components. This can in turn develop into research with statistical generalisability, which I did not aim for in this thesis. Given that most existing studies on teachers' NOS views are focused on their views of NOS ideas themselves rather than those of *teaching* and *assessing* NOS, large-scale studies on teachers' NOS instruction and assessment will provide critical evidence that can inform preparing teachers for NOS. Moreover, the mismatch between teacher views and practices reported in Chapter 5 also deserves continued attention. Although this thesis focused on the process and challenges of summative NOS assessment through a case study, future studies with larger samples can reveal the trends in teachers' assessment practices in diverse settings.

In considering teacher knowledge and practices of NOS assessment, it would be crucial to further investigate the extent to which domain-general knowledge and domain-specific knowledge shape teacher practice (see Section 0). Further research of teacher knowledge, views and practices of NOS assessment guided by topic-specificity of PCK (Veal & MaKinster, 1999) will enable a systematic understanding of the exploratory findings reported in this thesis. Given the interactions between NOS and scientific content knowledge reported in this thesis, how teachers make sense of the two for instruction and assessment will be an interesting area to investigate.

8.11.3 Collaborative summative assessment of NOS

One of the key motivations for Chapter 6 was to investigate how a *collaborative* setting can impact the summative assessment of NOS. As I discussed, the collaboration among the three teachers made their concerns relating to NOS assessment more explicit by bringing them to a public space, which otherwise might have been kept to each teacher. Still, there are some aspects of teacher practices that are not addressed in this thesis and thus can be explored in further research efforts.

First, there is a need for research addressing questions such as to what extent and how such collaboration can facilitate teachers' professional development, particularly in terms of their NOS assessment literacy. This thesis employed a naturalistic research design that did not involve researcher interventions, but in future research, interventions could be useful to promote teacher collaborations to assess NOS. Outside NOS research, Black and colleagues' studies (Black, 2010, 2021) provide an exemplar where researchers worked with teachers to improve the validity of summative assessment. Their model of collaboration engaged teachers in moderation exercises within and between schools, which could be particularly useful for assessing NOS that many science teachers reportedly struggle with (Höttecke & Silva, 2011; Wong, Kwan, Hodson & Yung, 2009). Lund's (2008) study also showed the possibility of collaborative assessment to facilitate teachers' professional development, particularly in an online environment. In this study, 14 preservice teachers of English as a foreign language were asked to grade two exam papers individually and give reasons for the grade, and then exchanged feedback in an online discussion forum. From the analysis of the online discussion, Lund argued that the threaded discussion elicited the negotiation and coordination of the participants' ideas about the grades and became a learning opportunity for the participants (p. 43).

Second, a more systematic analysis of the effect of social moderation processes on the quality of NOS assessment, and eventually its effect on student learning of NOS, will be necessary. Several studies, although not specifically in the context of NOS, have suggested the benefits of forming a community of assessors to the validity of assessment (Black et al., 2011; Lund, 2008). The findings in Chapter 6 suggest that the moderation process brought several important issues relating to NOS to the teachers' attention, which otherwise might have been ignored and not come to light. However, it remains unanswered if such collaboration can have a positive impact on the assessment quality such as validity and reliability. Further investigations of teacher markings and student data will provide insights into these unresolved issues.

8.11.4 Formative assessment of NOS

For nearly 25 years, formative assessment has drawn considerable attention from researchers (Black & Wiliam, 2009; Ruiz-Primo & Furtak, 2007; Shavelson et al., 2008) as well as policymakers (Assessment Reform Group, 1999; Lucas et al., 2013; National Research Council, 1996; OECD, 2013) in an effort to shift away from the dominance of summative assessment in many countries. At the same time, however, evidence suggests that it is not easy for teachers to change their practices to pay attention to students throughout instruction and use appropriate strategies to create learning opportunities based on collected information about student learning (Atkin, Coffey, Moorthy, Sato & Thibeault, 2005). NOS instruction can benefit from the theories of formative assessment, particularly because NOS researchers have emphasised using classroom interactive activities to teach NOS (Allchin, Andersen & Nielsen, 2014; McComas, 2020), where feedback and conversation can occur more frequently than in teacher-centred direct instructions.

There is a fair amount of research on NOS instruction in the science classroom, but few of them have looked at NOS instruction through the lens of formative assessment, or with the focus on teachers' process of collecting evidence of student learning and acting on it. Ryder and Leach's (2008) study is a rare example that investigated teachers' use of classroom talk in the context of NOS instruction and found the variations in the degree to which seven science teachers noticed students' ideas, recasted them to the whole class, and sometimes challenged their ideas to facilitate NOS learning. Although this thesis's analysis of Dr Kim's classroom practice through the lens of formative assessment illuminates some important connections between NOS and formative assessment, more theoretical and empirical discussion on this connection will be necessary.

First, given that teachers accumulate and aggregate information on students' ideas as they teach more classes (Cowie & Harrison, 2020), considering the temporality of formative assessment practice will be useful. It would be interesting to explore how a teacher's formative assessment practice such as noticing, questioning and feedback evolves over time as she accumulates more experience with NOS. Dr Kim's case suggested the possibility of improvement through experience when she told me about how she became more confident as she taught the paradigm shift lesson to more classes. The students' engagement in the assessment conversations also increased over time, showing the potential power of peer assessment in developing sophisticated and connected ideas about NOS. A further look into the use of peer assessment in NOS instruction will provide insights for making NOS learning more interactive and learner-centred. Second, how NOS formative assessment practices are influenced by teacher characteristics will need investigation. Since Chapter 7 is a study of one teacher, possible variations in practices and challenges of NOS formative assessment can be explored by including more teachers from different backgrounds in terms of subject specialism, NOS knowledge and pedagogical orientations. This will enable understanding the potential variation across teachers' NOS formative assessment practices, providing more nuanced and focused support to teachers in different circumstances. For example, a study with more teachers that investigates the relationship between their NOS knowledge, assessment views, NOS-PCK and actual assessment practices will give further insights useful to provide individualised support for teachers.

8.11.5 Students and teacher assessment practices

Although this thesis's focus was the views and practices of teachers, that is not the end of the story. Like instruction, classroom assessment is a mutual process between the teacher and her students. I have discussed several ways in which students can shape teacher assessment practices. When it comes to the formative assessment, which is inherently interactive and responsive, I showed that Dr Kim was responding spontaneously to student presentations about paradigms and that she became more confident and skilful after teaching the lesson to several classes. For summative assessment, it was mainly the marking and moderation process that was influenced by student performances in the assessment tasks. Evidence of student learning came from the teachers' observations while students were on task and the worksheets that students had completed. During the moderation process, the teachers used such information to establish marking criteria to guide their marking.

In the future, more detailed research work focusing on students' role in NOS classroom assessment. There are several remaining questions in this regard. For example, to what extent do students interpret the purposes of NOS assessment tasks as the teacher intended? After taking NOS lessons, what do students feel would be a fair and appropriate method of assessing NOS? Are their perceptions coherent with their teacher's? How would the alignment between the teacher's and student's perceptions of NOS impact the assessment practice? There are questions around formative assessment, too—Do students actually build connections between components of NOS learning when the teacher makes conscious efforts to do so, as in the case of Dr Kim? What was the impact of the peer assessment in Dr Kim's lesson on student learning of NOS? Research focusing on these questions will complement what we know about NOS classroom assessment from the teacher's point of view.

8.11.6 Initial teacher education and continuing professional development

Providing opportunities for professional learning is integral to challenging teachers' conceptions of classroom assessment as well as their assessment practices (Correia & Harrison, 2019). This points to the need for research-informed professional development, both at the initial and in-service teacher education levels, although they will take time to have an effect in classrooms. Given that the current study was of exploratory and naturalistic nature, an intervention study that investigates the design and evaluation of teacher education programmes focused on NOS assessment will be necessary in the future. Previous large-scale intervention studies on classroom assessment such as the King's-Medway-Oxfordshire Formative Assessment Project (Black et al., 2003), the King's Oxfordshire Summative Assessment Project (Black et al., 2010) and the more recent Assess Inquiry in Science, Technology and Mathematics Education Project (ASSIST-ME) (Dolin & Evans, 2018) have demonstrated that the provision of professional learning by researchers can motivate teachers to explore and experiment with classroom assessment and make changes in their classroom practice (Cowie & Harrison, 2020).

Meanwhile, these earlier projects suggested the importance of creating mutual learning opportunities between researchers and teachers (Cowie & Harrison, 2020), rather than adopting a one-way diffusion model of knowledge transfer where teachers passively participate in the programme provided by 'experts'. Given the lack of evidence on the topic, I suggest that the two parties can work best by co-creating knowledge about the classroom assessment of NOS—not only its purpose and possible methods but also what works and what does not, why teachers make certain assessment decisions, and how individual context and culture of the classroom can be addressed, particularly in the context of high-stakes examinations. This way, teachers can be producers of knowledge as they engage in research, learn new things, implement them in their classrooms, and reflect on the process (Cochran-Smith & Lytle, 1993). Taking a collaborative, action research approach will maximise what we can learn from the professional development interventions about improving NOS assessment in the classroom.

The opportunities and challenges identified in this study provide some information that can be useful to establish design principles for NOS assessments. For example, NOS assessments should be targeted at the connections between various elements of

NOS rather than discrete facts. If an assessment can be answered only by rote memorisation of several statements, it will fail to yield useful information for teachers to identify students' needs and facilitate learning. This seems to discourage the use of multiple-choice questions in favour of document-based and activity-based assessments. However, the findings of Chapter 6 suggest that excessive use of texts and complex tasks can also be problematic since it can reduce a NOS assessment to an assessment of something else, such as relevant content knowledge and reading and writing skills. Given these challenges, it is instrumental for researchers to collaborate with teachers to design and validate NOS assessments that can be used and adapted for classroom use.

8.12 Final thoughts

The idea of teaching students what science is and how it works has a longer tradition than it initially sounds. Matthews (2012) finds its roots in William Whewell's Royal Institution of Great Britain lecture in 1854, and Jenkins (2013) in the Duke of Argyll's presidential address at the British Association for the Advancement of Science in 1855. These nascent precursors to NOS education have since been echoed and elaborated by influential figures such as Henry Armstrong (1898), John Dewey (1916), Frederick Westaway (1929) and Joseph Schwab (1962). Nevertheless, NOS continues to struggle to find its right place in school science. In Korea, the national curriculum from as early as the 1980s stated some ideas about NOS as a key learning goal. However, as with many other elements of the national curriculum, NOS was presented as a higher-level goal no more than something that might be good to learn (MOE, 1988). In the past curricula, NOS, at best, was expected to be automatically learnt as students study scientific content knowledge, without having to teach or assess consciously. In this respect, the new curriculum that foregrounded NOS as an explicit learning goal is an encouraging shift. This thesis capitalised on this opportunity to reveal various aspects of NOS classroom assessment as teachers in the wake of a NOS-oriented curriculum reform.

In this chapter, I argued that my case study showed that NOS classroom assessment is possible, but supports are needed as to providing resources and supporting teachers to develop clear rationales for assessing NOS in the classroom. I also discussed more specific issues and tasks such as the comparison of science assessment and NOS assessment, the alignment of curriculum, instruction and assessment, the balance between content knowledge and NOS, and controversial and domain-specific aspects of NOS. In so doing, I often referred to the recent developments in science education research such as LPs to consider possible solutions to the challenges and obstacles identified from the study. Drawing on the teachers' practices in formative and summative assessment in the study, I also discussed the potential tension arising from the two functions of classroom assessment in the context of NOS instruction. Although the study focused on only three teachers, the findings offer useful insights and raise important questions about NOS that are not limited to the specific participants and context of the study, as outlined in this chapter.

Along with answering the research questions, this thesis also raised new questions about the teaching and assessment of NOS. Science educators agree that NOS should be taught—but what kind of cognitive state do we expect of learners? Do we aim to develop certain 'views' of NOS, or is it more akin to an 'understanding' of it? To what extent can NOS learning outcomes be assessed as right or wrong, better or worse, or informed or naïve? What kinds of information can teachers accept as evidence of NOS learning? Can we assess NOS learning separately from students' scientific content knowledge and scientific reasoning skills? From the analysis of teachers' views and practices, I have discussed that there exists little consensus on these fundamental questions about NOS instruction and assessment. A coherent model of NOS classroom assessment will need to consider these questions and establish clear connections between the learning goals and how they can be measured.

In this age of accountability, what students learn largely depends on what they are tested on, both in the classroom and high-stake examinations. Given such centrality of assessment in today's education system, the 'backgrounding' of NOS in the science curriculum is problematic. This thesis started from the idea that NOS should be brought to the fore, targeted clearly as a learning goal, and assessed formatively and summatively. Given that the science curriculum is already overcrowded, arguments for the inclusion, and even assessment, of NOS may seem to imply the relative demotion of content knowledge, which some people regard as anti-scientific.[26] This thesis has presented concrete evidence that 'scientific' knowledge (i.e., science content knowledge) and NOS knowledge are inseparable and complementary in the context of instruction and assessment (Sections 6.4.3 and 7.3.3). Take, for an everyday example, the science of the human immune system against viruses. In the biology class, we are told that humans develop immunity after an infection, so we do not catch the same virus again. This 'hard fact' fails to explain the cases of COVID-19 reinfection and therefore questions our trust in the scientific knowledge about viruses. To make sense of this seeming failure of science, one also needs to understand that science can change based on available evidence, and when the evidence is insufficient, our knowledge can also be limited in its scope and explanatory power. An awareness of NOS as such would not only influence one's course of action in response to the outbreak but also allow for a more elaborate understanding of fast-changing science.

Going back to the current state of the world, the COVID-19 pandemic has vividly revealed the nature of scientific knowledge and practices in a way that would not have been possible without it. Not only does it show the tentative NOS (e.g., the changes in the scientific understanding of how the virus is spread) and the variety of evidence collection in science (e.g., epidemiological modelling and randomised controlled trials to assess the effects of clinical interventions), but it also points to various aspects of science as an institution. It has shown us the impact of funding and governmental support on the speed of research on the disease, and how social and political interests can drive the conduct of scientific research. The fact that many prominent COVID-19 vaccines were developed jointly by universities and biotechnology companies tells us the prevalence of academic-industry collaboration in the application of science today. Looking into the process of policy decision making, one can realise that the line between science and politics is not as solid as often thought. All these examples indicate

[26] When the *Korean Science Education Standards* (KSES) that foregrounded the 'action and participation' through science learning first came out, a major science media in Korea published an article criticising KSES, titled 'Bizarre science education standards with "science" disappeared and "society" emphasised" (Jo, 2019). This episode shows that the emphasis on aspects of science other than conceptual understanding can provoke backlashes, pointing to the need for more efforts to rationalise the curricular inclusion of NOS to relevant stakeholders.

that an awareness of NOS is not a luxury but a necessity for responsible citizens.

Due to its nature as an educational research project, this thesis has called for a number of changes across curriculum and assessment policy, teacher education and classroom practice. The shift towards NOS classroom assessment, however, needs not to be revolutionary but should be rather incremental through small, practical changes in the science curriculum and teacher education. The constraints that impede NOS *teaching*, such as the lack of available resources, insufficient pedagogical skills, time limits, and the difficulty of justifying NOS against the curriculum and colleagues (Henke & Höttecke, 2015), are problematic for teachers' NOS classroom *assessment* as importantly. Despite these known challenges, the assessments that the teachers in this study used were a good start. Except for the VNOS task, the assessment tasks included both student-led activities that link scientific knowledge and real-life applications and reflection questions to elicit NOS understandings. The challenges associated with marking were evident, but these do not necessarily mean that the assessments were ill-designed, but rather indicate the need for theoretically and empirically grounded support for teachers. It would be essential for research and educational practice to inform each other to turn the suggestions into reality.

Given that NOS is a conventionally unfamiliar aspect of science learning, school- and university-based professional development efforts will be crucial in maximising NOS learning through effective classroom assessment. Although many efforts have been made to develop preservice and in-service teachers' understandings of NOS and its instruction (Akerson, Cullen & Hanson, 2010; Donnelly & Argyle, 2011; Erduran & Kaya, 2019; Kaya, Erduran, Aksoz & Akgun, 2019; Mesci et al., 2020), assessment has rarely been the focus of these professional development programmes. Teacher educators should provide preservice and in-service teachers with training and support to improve their knowledge and skills for NOS assessment. As more teachers sympathise with the value of assessing NOS in the classroom through these opportunities, and the curriculum policy allows spaces to do so, more concrete efforts could be made to support teachers at all career stages to use the classroom assessment to promote effective and meaningful learning of NOS. In doing so, the cultural and historical contexts of assessment must be carefully considered, given that all assessments are socially situated and should be coherent with the rest of the education system.

While the increasing attention to the value of NOS in the wake of pressing global challenges such as COVID-19, misinformation, climate change denial and vaccine hesitancy is encouraging, to tackle these issues, more is needed than presenting NOS as an abstract, higher-level curriculum goal that never comes down to the classroom instruction. To empower students to have an epistemic and functional understanding of NOS, we need to assess their NOS learning in the classroom and support teachers to be competent at NOS assessment. In this sense, the mismatch between teachers' NOS teaching intentions and NOS assessment intentions (Chapter 5), the various challenges of summative and formative assessment of NOS (Chapter 6 and Chapter 7) can be a useful baseline for developing professional development programmes that are more attentive to the assessment of NOS. As with any other empirical research project in education, this study was situated within specific regional, curricular and school contexts, but findings from the three teachers' views and experiences shed light on how various issues of classroom assessment can unfold in the context of NOS and what opportunities and challenges are there, which offer insights that will be useful beyond this particular case. Ultimately, supporting science teachers for effective and meaningful NOS assessment will enable students to develop skills crucial for responsible citizenship through learning about what science is and how it operates.

References

Abd-El-Khalick, F. (1998). The influence of history of science courses on students' conceptions of nature of science. Unpublished doctoral dissertation. Oregon State University, Corvallis.

Abd-El-Khalick, F. (2005). Developing deeper understandings of nature of science: The impact of a philosophy of science course on preservice science teachers' views and instructional planning. *International Journal of Science Education*, *27*(1), 15–42.

Abd-El-Khalick, F. (2012). Nature of science in science education: Toward a coherent framework for synergistic research and development. In B. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 1041-1060). Springer, Dordrecht.

Abd-El-Khalick, F. (2013). Teaching with and about nature of science, and science teacher knowledge domains. *Science & Education*, *22*(9), 2087–2107.

Abd-El-Khalick, F., & Akerson, V.L. (2009). The influence of metacognitive training on preservice elementary teachers' conceptions of nature of science. *International Journal of Science Education*, *31*(16), 2161–2184.

Abd-El-khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education*, *22*(7), 665–701.

Abd-El-Khalick, F., Belarmino, J., Brunner, J., Le, A.-P., Myers, J. Y., Summers, R. G., … Zeineddin, A. A. (2017). A longitudinal analysis of the extent and manner of representations of nature of science in U.S. high school chemistry, biology, and physics textbooks. In C. V. McDonald & F. Abd-El-Khalick (Eds.), *Representations of nature of science in school science textbooks: A global perspective* (pp. 20–60). London: Routledge.

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, *82*(4), 417–436.

Abell, S. K., & Lederman, N. G. (2007). *Handbook of research on science education*. Routledge.

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Dordrecht: Springer.

Abell, S. K., & Smith, D. C. (1994). What is science?: Preservice elementary teachers' conceptions of the nature of science. *International Journal of Science Education*, *16*(4), 475–487.

Adams, P. E., & Krockover, G. H. (1997). Concerns and perceptions of beginning secondary science and mathematics teachers. *Science Education*, *81*(1), 29–50.

Adie, L., Addison, B., & Lingard, B. (2021). Assessment and learning: An in-depth analysis of change in one school's assessment culture. *Oxford Review of Education*. Advance online publication. https://doi.org/10.1080/03054985.2020.1850436

Ahrenkiel, L., & Worm-Leonhard, M. (2014). Offering a forensic science camp to introduce and engage high school students in interdisciplinary science topics. *Journal of chemical education*, *91*(3), 340-344.

Ahrenkiel, L., & Worm-Leonhard, M. (2014). Offering a forensic science camp to introduce and engage high school students in interdisciplinary science topics. *Journal of Chemical Education*, *91*(3), 340–344.

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: Views on Science-Technology-Society (VOSTS). *Science Education*, *76*(5), 477–491.

Akerson, V. L., & Donnelly, L. A. (2010). Teaching nature of science to K-2 students: What understandings can they attain? *International Journal of Science Education*, *32*(1), 97–124.

Akerson, V. L., Abd-El-Khalick, F., & Lederman, N. G. (2000). Influence of a reflective explicit activity-based approach on elementary teachers' conceptions of nature of science. *Journal of Research in Science Teaching*, *37*(4), 295–317.

Akerson, V. L., Buck, G. A., Donnelly, L. A., Nargund-Joshi, V., & Weiland, I. S. (2011). The importance of teaching and learning nature of science in the early childhood years. *Journal of Science Education and Technology*, *20*(5), 537–549.

Akerson, V. L., Carter, I. S., & Elcan, N. (2016). On the nature of professional identity for nature of science. In L. Avraamidou (Ed.), *Studying science teacher identity: Theoretical, methodological and empirical explorations* (pp. 89–110). Rotterdam: Sense.

Akerson, V. L., Carter, I., Pongsanon, K., & Nargund-Joshi, V. (2019). Teaching and learning nature of science in elementary classrooms. *Science & Education*, *28*(3–5), 391–411.

Akerson, V. L., Cullen, T. A., & Hanson, D. L. (2009). Fostering a community of practice through a professional development program to improve elementary teachers' views of nature of science and teaching practice. *Journal of Research in Science Teaching*, *46*(10), 1090–1113.

Akerson, V. L., Cullen, T. A., & Hanson, D. L. (2010). Experienced teachers' strategies for assessing nature of science conceptions in the elementary classroom. *Journal of Science Teacher Education*, *21*(6), 723–745.

Akerson, V. L., Donnelly, L. A., Riggs, M., & Eastwood, J. L. (2012). Developing a community of practice to support preservice elementary teachers' nature of science instruction. *International Journal of Science Education*, *34*(9), 1371–1392.

Akerson, V. L., Hanson, D. L., & Cullen, T. A. (2007). The influence of guided inquiry and explicit instruction on K-6 teachers' views of nature of science. *Journal of Science Teacher Education*, *18*(5), 751–772.

Akerson, V. L., Pongsanon, K., Rogers, M. A. P., Carter, I., & Galindo, E. (2017). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science. *International Journal of Science and Mathematics Education*, *15*(2), 293–312.

Akerson, V.L., Abd-El-Khalick, F., & Lederman, N.G. (2000). Influence of a reflective explicit activity-based approach on elementary teachers' conceptions of nature of science. *Journal of Research in Science Teaching*, *37*(4), 295–317.

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20-34.

Allal, L., & Mottier Lopez, L. (2014). Teachers' professional judgment in the context of collaborative assessment practice. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 151–165). Dordrecht: Springer.

Allchin, D. (2013). *Teaching the nature of science: Perspective & resources*. Saint Paul, MN: SHiPs Education Press.

Allchin, D. (2017). Beyond the consensus view: Whole science. *Canadian Journal of Science, Mathematics and Technology Education*, *17*(1), 18–26.

Allchin, D. (2020). From nature of science to social justice: The political power of epistemic lessons. In H. A. Yacoubian, & L. Hansson (Eds). *Nature of science for social justice* (pp. 23–39). Dordrecht: Springer.

Allchin, D., Andersen, H. M., & Nielsen, K. (2014). Complementary approaches to teaching nature of science: Integrating student inquiry, historical cases, and contemporary cases in classroom practice. *Science Education*, *98*(3), 461–486.

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, *93*(3), 389–421.

American Association for the Advancement of Science. (AAAS). (1989). *Science for all Americans*. Washington, DC: AAAS.

American Association for the Advancement of Science. (AAAS). (1993). *Benchmarks for science literacy*. Oxford: Oxford University Press.

American Federation of Teachers. (AFT), National Council on Measurement in Education. (NCME), & National Education Association. (NEA) (1990). Standards for teacher competence in educational assessment. Retrieved 28 September 2020 from https://buros.org/standards-teacher-competence-educational-assessment-students

Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. London: Routledge.

Andrade, H., & Cizek, G. J. (Eds.) (2010) *Handbook of formative assessment*. New York, NY: Routledge.

Antink-Meyer, A., & Brown, R. A. (2019). Nature of Engineering Knowledge. *Science & Education*, *28*(3–5), 539–559.

Aragón, M., José, M., & Acevedo, A. (2018). Prospective biology teachers' understanding of the nature of science through an analysis of the historical case of Semmelweis and childbed fever. *Cultural Studies of Science Education*. Advance online publication. https://doi.org/10.1007/s11422-018-9868-y

Armstrong, H.E. (1898/1903) *The teaching of scientific method and other papers on education*. London: Macmillan.

Atkin, J. M., Coffey, J. E., Moorthy, S., Sato, M., & Thibeault, M. (2005). *Designing everyday assessment in the science classroom*. New York: Teachers College Press.

Ban, J.-C., Kim, S., Park, C., & Kim, H. (2018). Teachers' perceptions of teacher process fortified assessment policy. *The Journal of Curriculum and Evaluation, 21*(3), 105–130.

Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Mahwah, NJ: Lawrence Erlbaum.

Bell, B. (2007). Classroom assessment of science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (1st ed., pp. 965–1005). Mahwah, NJ: Lawrence Erlbaum Associates.

Bell, R. L. (2008). *Teaching the nature of science through process skills: Activities for grades 3-8*. New York, NY: Pearson.

Bell, R. L., & Clair, T. L. S. (2015). Too little, too late: Addressing nature of science in early childhood education. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 125–141). Dordrecht: Springer.

Bell, R. L., Blair, L., Crawford, B., & Lederman, N. (2003). 'Just do it': The impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching, 40*(5), 487–509.

Bell, R. L., Lederman, N. G., & Abd-El-Khalick, F. (2000). Developing and acting upon one's conception of the nature of science: A follow-up study. *Journal of Research in Science Teaching*, *37*(6), 563–581.

Bell, R. L., Mulvey, B. K., & Maeng, J. L. (2016). Outcomes of nature of science instruction along a context continuum: Preservice secondary science teachers' conceptions and instructional intentions. *International Journal of Science Education*, *38*(3), 493–520.

Benn, M. (2020, August 11). England's exam results fiasco has exposed its flawed education system. The Guardian. https://www.theguardian.com/commentisfree/2020/aug/11/england-results-exams-fiasco-ofqual-inequality

Benn, M. (2020, August 11). England's exam results fiasco has exposed its flawed education system. *The Guardian. https://www.theguardian.com/commentisfree/2020/aug/11/england-results-exams-fiasco-ofqual-inequality*

Bentley, M. L. (2000). Improvisational drama and the nature of science. *Journal of Science Teacher Education, 11*(1), 63–75.

Bianchini, J. A., Cavazos, L. M., & Helms, J. V. (2000). From professional lives to inclusive practice: Science teachers and scientists' views of gender and ethnicity in science education. *Journal of Research in Science Teaching*, *37*(6), 511–547.

Bilican, K., Akerson, V., & Nargund, V. (2020). Learning by teaching: A case study of co-teaching to enhance nature of science pedagogy, successes, and challenges. *International Journal of Science and Mathematics Education*. Advance online publication. https://doi.org/10.1007/s10763-020-10094-6

Birks, M., Chapman, Y., & Francis, K. (2008). Memoing in qualitative research. *Journal of Research in Nursing, 13*(1), 68–75.

Black, P. J. (1998). *Testing, friend or foe?: the theory and practice of assessment and testing*. London: Falmer.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: GL Assessment.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215–232.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, *18*(4), 451–469.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.

Bordwell, D., & Thompson, K. (1990). *Film art: An introduction* (5th ed.) New York, NY: McGraw-Hill.

Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties and principles. In J. H. McMillan (Ed.), *SAGE Handbook on research on classroom assessment* (pp. 87–106). Thousand Oaks, CA: Sage.

BouJaoude, S., Dagher, Z. R., & Refai, S. (2017). The portrayal of nature of science in Lebanese ninth grade science textbooks. In C. V. McDonald & F. Abd-El-Khalick (Eds.), *Representations of nature of science in school science textbooks: A global perspective* (pp. 79–97). London: Routledge.

Boujaoude, S., Sowwan, S., & Abd-El-Khalick, F. (2005). The effect of using drama in science teaching on students' conceptions of the nature of science. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 259–267). Dordrecht: Springer.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative*

Research in Psychology, *3*(2), 77–101.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (pp. 57–71). Washington, DC: APA Books.

Brickhouse, N. W. (1990). Teachers' beliefs about the nature of science and their relationship to classroom practice. *Journal of Teacher Education*, *41*(3), 53–62.

Brock, R., & Taber, K. S. (2019). 'I'm sad that it's gone': A case study of teachers' views on teaching the nature of science at Key Stage 4. *School Science Review*, *100*(373), 29–34.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5–12.

Brown, G. T. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological reports*, *99*(1), 166–170.

Brown, G. T. L. (2002). Teachers' conceptions of assessment. Unpublished doctoral dissertation. University of Auckland, Auckland.

Brunner, J. L., & Abd-El-Khalick, F. (2017). Representations of nature of science in US elementary science trade books. In C. McDonald & F. Abd-El-Khalick (Eds), *Representations of nature of science in school science textbooks* (pp. 135–151). London: Routledge.

Brunner, J. L., & Abd-El-Khalick, F. (2020). Improving nature of science instruction in elementary classes with modified science trade books and educative curriculum materials. *Journal of Research in Science Teaching*, *57*(2), 154–183.

Burgin, S. R., & Sadler, T. D. (2015). Learning nature of science concepts through a research apprenticeship program: A comparative study of three approaches. *Journal of Research in Science Teaching*, *53*(1), 31–59.

Byun, S.-Y. (2014). Shadow education and academic success in Republic of Korea. In H. Park & K.-K. Kim (Eds)., *Korean education in changing economic and demographic contexts* (pp. 39–58). Singpore: Springer.

Byun, T., Baek, J., Shim, H.-P., & Lee, D. (2019). An investigation on the implementation of the 'scientific inquiry experiment' of the 2015 revised curriculum. *Journal of the Korean Association for Science Education*, *39*(5), 669–679.

Cakici, Y., & Bayir, E. (2012). Developing children's views of the nature of science through role play. *International Journal of Science Education*, *34*(7), 1075–1091.

Champagne, A. B. (2014). Assessment: An overview. In R. Gunstone (Ed.), *Encyclopedia of science education* (pp. 87–90). Dordrecht: Springer.

Chang, H. (2011). How historical experiments can improve scientific knowledge and science education: The cases of boiling water and electrochemistry. *Science & Education*, *20*(3), 317–341.

Chen, H. Y., & Boore, J. R. (2010). Translation and back-translation in qualitative nursing research: Methodological review. *Journal of clinical nursing*, *19*(1-2), 234–239.

Chesky, N. Z., & Wolfmeyer, M. R. (2015). *Philosophy of STEM education: A critical investigation*. New York, NY: Palgrave Macmillan.

Cheung, K. K. C. (2020). Exploring the inclusion of nature of science in biology curriculum and high-stakes assessments in Hong Kong. *Science & Education*, *29*(3), 491–512.

Christenson, N., Gericke, N., & Rundgren, S. N. C. (2017). Science and language teachers' assessment of upper secondary students' socioscientific argumentation. *International Journal of Science and Mathematics Education*, *15*(8), 1403–1422.

Christenson, N., Gericke, N., & Rundgren, S. N. C. (2017). Science and language teachers' assessment of upper secondary students' socioscientific argumentation. *International Journal of Science and Mathematics Education*, *15*(8), 1403–1422.

Clough, M. P. (2006). Learners' responses to the demands of conceptual change: Considerations for effective nature of science instruction. *Science & Education, 15*(5), 463–494.

Clough, M. P., Herman, B. C., & Olson, J. K. (2020). Preparing science teachers to overcome common obstacles and teach nature of science. In W. F. McComas

(Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 239–252). Dordrecht: Springer.

Cochran-Smith, M. & Lytle, S. L. (1993). *Inside/outside: Teacher research and knowledge*. New York: Teachers College Press.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, *48*(10), 1109–1136.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational evaluation and policy analysis*, *25*(2), 119-142.

Cohen, I. B. (1985). *Revolution in science*. Cambridge MA: Harvard University Press.

Committee for Public Deliberation on the Revision of College Entrance System. (CPDRCES). (2018). *Public deliberation on the revision of college entrance system white paper*. Seoul: CPDRCES. Retrieved 3 November 2020 from http://19edutalk.pa.go.kr/reference/whiteBook

Conant, J. B. (1950). *General education in a free society: Report of the Harvard Committee*. Cambridge, MA: Harvard University Press.

Cowie, B., & Harrison, C. (2020). The what, when & how factors: reflections on classroom assessment in the service of inquiry. *International Journal of Science Education*. Advance online publication. https://doi.org/10.1080/09500693.2020.1824088

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2016). *30 essential skills for the qualitative researcher*. Thousand Oaks, CA: Sage.

Crimmins, G., Nash, G., Oprescu, F., Liebergreen, M., Turley, J., Bond, R., & Dayton, J. (2016). A written, reflective and dialogic strategy for assessment feedback that can enhance student/teacher relationships. *Assessment & Evaluation in Higher Education*, *41*(1), 141–153.

Crisp, V. (2008). A review of literature regarding the validity of coursework and the rationale for its inclusion in the GCSE. *Research Matters*, *5*, 20–24. Retrieved 28 September 2020 from https://www.cambridgeassessment.org.uk/Images/499640-a-review-of-literature-regarding-the-validity-of-coursework-and-the-rationale-for-its-inclusion-in-the-gcse.pdf

Cronin-Jones, L. L. (1991). Science teacher beliefs and their influence on curriculum implementation: Two case studies. *Journal of Research in Science Teaching*, *28*(3), 235–250.

Cumming, J. J., Van der Kleij, F. M., & Adie, L. (2019). Contesting educational assessment policies in Australia. *Journal of Education Policy*, *34*(6), 836–857.

Dagher, Z. R., & Boujaoude, S. (2005). Students' perceptions of the nature of evolutionary theory. *Science Education*, *89*(3), 378–391.

Darling-hammond, L., Adamson, F., Kahl, S., Lane, S., Montague, W., Olson, J., Owens, M., & Pecheone, R. (2010). *Beyond basic skills: Achieving 21st century standards of learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Daston, L. (2011). The empire of observation. In L. Daston and E. Lunbeck (Eds.), *Histories of scientific observation* (pp. 81–113). Chicago, IL: The University of Chicago Press.

Davies, P &, Amos, R. (2017). Science teacher education: A multi-country comparison - United Kingdom. In J. E. Pederson, T. Isozaki, & T. Hirano (eds.), *Model science teacher preparation programes: An international comparison of what works* (pp. 1–28). Charlotte, NC: Information Age Publishing.

DeCoito, I. (2020). The use of digital technologies to enhance learners' conceptions of nature of science. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 343–358). Dordrecht: Springer.

del Mar Aragón-Méndez, M., Acevedo-Díaz, J. A., & García-Carmona, A. (2019). Prospective biology teachers' understanding of the nature of science through an analysis of the historical case of Semmelweis and childbed fever. *Cultural Studies of Science Education*, *14*(3), 525–555.

DeLuca, C., & Braund, H. (2019). Preparing Assessment Literate Teachers. In *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-487

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, *21*(4), 248–266.

Deng, F., Chen, D.-T., Tsai, C.-C., & Tsai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, *95*(6), 961–999.

Dewey, J. (1916). *Democracy and education*. New York: Macmillan.

Dhingra, K. (2003). Thinking about television science: How students understand the nature of science from different program genres. *Journal of Research in Science Teaching*, *40*(2), 234–256.

Dibattista, L., & Morgese, F. (2014). Incorporation of HPS/NOS content in school and teacher education programmes in Europe. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 2083–2118). Dordrecht: Springer.

Dolin, J., & Evans, R. (2018). *Transforming assessment: Through an interplay between practice, research and policy*. Dordrecht: Springer.

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring Relations Between Formative and Summative Assessment. In J. Dolin & R. Evans (Eds.), *Transforming Assessment: Contributions from Science Education Research* (pp. 53–80). Dordrecht: Springer.

Dolphin, G., & Dodick, J. (2014). Teaching controversies in earth science: The role of history and philosophy of science. In *International handbook of research in history, philosophy and science teaching* (pp. 553–599). Springer, Dordrecht.

Donmoyer, R. (1990). Generalizability and the single-case study. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 175–200). New York, NY: Teachers College Press.

Donnelly, L. A., & Argyle, S. (2011). Teachers’ willingness to adopt nature of science activities following a physical science professional development. *Journal of Science Teacher Education*, *22*(6), 475–490.

Driver, R., Leach, J., Millar, R., & Scott, P. (1997). A study of progression in learning about ‘the nature of science’: Issues of conceptualisation and methodology. *International Journal of Science Education*, *19*(2), 147–166.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, *46*(6), 606–609.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. *Journal of Research in Science Teaching*, *46*(6), 655–674.

Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*(1), 268–291.

Duschl, R. A., & Bismack, A. S. (Eds.). (2016). *Reconceptualizing STEM education: The central role of practices*. London: Routledge.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, *4*(1), 37–41.

Duschl, R. A., & Grandy, R. (2013). Two views about explicitly teaching nature of science. *Science & Education*, *22*(9), 2109–2139.

Dwyer, S. C., & Buckle, J. L. (2009). The space between: On being an insider-outsider in qualitative research. *International Journal of Qualitative Methods*, *8*(1), 54–63.

Eastwood, J. L., Sadler, T. D., Zeidler, D. L., Lewis, A., Amiri, L., & Applebaum, S. (2012). Contextualizing nature of science instruction in socioscientific issues. *International Journal of Science Education*, *34*(15), 2289–2315.

Egan, K. (1989). Memory, imagination, and learning: Connected by the story. *Phi Delta Kappan*, *70*(6), 455–459.

Elwood, J. (2006) Formative assessment: possibilities, boundaries and limitations, *Assessment in Education*, 13(2), 215–232

Enderle, P., Cohen, S., & Scott, J. (2020). Communicating about science and engineering practices and the nature of science: An exploration of American Sign Language resources. *Journal of Research in Science Teaching*, *57*(6), 968–995.

Engelsen, K., & Smith, K. (2010). Is "Excellent" good enough?. *Education inquiry*, *1*(4), 415–431.

Erduan, S., Kaya, E., Cullinane, A., Imren, O., & Kaya, S. (2020). Practical learning resources and teacher education strategies for understanding nature of science. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 377–398). Dordrecht: Springer.

Erduran, S. (2015). Introduction to the focus on ... scientific practices. *Science Education*, *99*(6), 1023–1025. https://doi.org/10.1002/sce.21192

Erduran, S. (2020). Bringing nuance to 'the science' in public policy and science understanding. *Science & Education*, *29*(3), 487–489.

Erduran, S., & Dagher, Z. R. (2014a). *Reconceptualizing the nature of science for science education*. Dordrecht: Springer.

Erduran, S., & Dagher, Z. R. (2014b). 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., & Kaya, E. (2018). Drawing nature of science in pre-service science teacher education: Epistemic insight through visual representations. *Research in Science Education*,*48*(6), 1133–1149.

Erduran, S., & Kaya, E. (2019). *Transforming teacher education through the epistemic core of chemistry*. Dordrecht: Springer.

Erduran, S., & Mugaloglu, E. Z. (2013). Interactions of economics of science and science education: Investigating the implications for science teaching and learning. *Science & Education*, *22*(10), 2405–2425.

Erduran, S., Guilfoyle, L., & Park, W. (in press). Science and religious education teachers' perceptions of argumentation and its teaching. *Research in Science Education*.

Erduran, S., Kaya, E., Cilekrenkli, A., Akgun, S., & Aksoz, B. (2020). Perceptions of nature of science emerging in group discussions: A comparative account of pre-service teachers from Turkey and England. *International Journal of Science and Mathematics Education*. Advance online publication. https://doi.org/10.1007/s10763-020-10110-9

Fay, B. (1996). *Contemporary philosophy of social science: A multicultural approach*. Cambridge, UK: Blackwell.

Fernandes, G. W. R., Rodrigues, A. M., & Ferreira, C. A. (2018). Conceptions of the nature of science and technology: a study with children and youths in a non-formal science and technology education setting. *Research in Science Education*, *48*(5), 1071–1106.

Fraser, B. J., Tobin, K., & McRobbie, C. J. (2012). *Second International Handbook of Science Education*. Dordrecht: Springer.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, *25*(7), 739–755.

Galili, I. (2012). Promotion of cultural content knowledge through the use of the history and philosophy of science. *Science & Education*, *21*(9), 1283–1316.

Galili, I. (2013). On the power of fine arts pictorial imagery in science education. *Science & Education*, *22*(8), 1911–1938.

Galili, I. (2019). Towards a refined depiction of nature of science: Applications to physics education. *Science & Education*, *28*(3–5), 503–537.

García-Carmona, A., & Díaz, J. A. A. (2016). Learning about the nature of science using newspaper articles with scientific content. *Science & Education*, *25*(5–6), 523–546.

Gardner, J. (2010). Developing teacher assessment: An introduction. In J. Gardner, W. Harlen, L. Hayward (Eds.), *Developing teacher assessment* (pp. 1–13). London: McGraw-Hill Education.

Gess-Newsome J, & Lederman, N. G. (Eds.). (1999). *Examining pedagogical content* *knowledge: The construct and its implications for science education*. Dordrecht: Kluwer Academic.

Giere, R. N. (1991) *Understanding scientific reasoning* (3rd ed.). Fort Worth, TX: Holt, Rinehart and Winston.

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

Gioka, O. (2009). Teacher or examiner? The tensions between formative and summative assessment in the case of science coursework. *Research in Science Education*, *39*(4), 411–428.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. New York, NY: Routledge.

Gooding, D. (1992). Putting agency back into experiment. In A. Pickering (Ed.), *Science as practice and culture* (pp. 65–112). Chicago, IL: University of Chicago Press.

Grinyer, A., & Thomas, C. (2012). The value of interviewing on multiple occasions or longitudinally. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti & K. D. McKinney (Eds.), *The Sage handbook of interview research: The complexity of the craft* (pp. 219-230). Thousand Oaks, CA: Sage.

Guilfoyle, L., Erduran, S., & Park, W. (2020). Science and religious education teachers' views of the comparison of argumentation in science and religion. *Journal of Beliefs and Values*. Advance online publication. https://doi.org/10.1080/13617672.2020.1805925

Güney, B. G., & Şeker, H. (2017). Discovering socio-cultural aspects of science through artworks. *Science & Education*, 26(7–9), 867–887.

Hadzigeorgiou, Y., Klassen, S., & Klassen, C. F. (2012). Encouraging a "romantic understanding" of science: The effect of the Nikola Tesla story. *Science & Education*, *21*(8), 1111-1138.

Hall, W. A., & Callery, P. (2001). Enhancing the rigor of grounded theory: Incorporating reflexivity and relationality. *Qualitative Health Research*, *11*(2), 257–272.

Halliday, J. (2010). Educational assessment. In R. Bailey, R. Barrow, D. Carr & C. McCarthy (Eds.), *The SAGE handbook of philosophy of education* (pp. 369–383). Thousand Oaks, CA: Sage.

Hand, M., & Levinson, R. (2012). Discussing controversial issues in the classroom. *Educational Philosophy and Theory, 44*(6), 614–629.

Hanuscin, D. L. (2013). Critical incidents in the development of pedagogical content knowledge for teaching the nature of science: A prospective elementary teacher's journey. *Journal of Science Teacher Education*, *24*(6), 933–956.

Hanuscin, D. L., Lee, M. H., & Akerson, V. L. (2011). Elementary teachers' pedagogical content knowledge for teaching the nature of science. *Science Education*, *95*(1), 145–167.

Hanuscin, D., Khajeloo, M., & Herman, B. C. (2020). Considering the clssroom assessment of nature of science. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 409–423). Dordrecht: Springer.

Haraway, D. (1988). Situated knowledge: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575–599

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research papers in education*, *20*(3), 245-270.

Harlen, W. (2011). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 95–110). Thousand Oaks, CA: Sage.

Hattiem, J, & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Heering, P. (2000). Getting shocks: Teaching secondary school physics through history. *Science & Education*, *9*(4), 363-373.

Henke, A., & Höttecke, D. (2015). Physics teachers' challenges in using history and philosophy of science in teaching. *Science & Education*, *24*(4), 349–385.

Heritage, M. (2008). Learning progressions: Supporting instruction and formative assessment. Washington, DC: Chief Council of State School Officers.

Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Cambridge, MA: Harvard Education Press.

Herman, B. C., Clough, M. P., & Olson, J. K. (2017). Pedagogical reflections by secondary science teachers at different NOS implementation levels. *Research in Science Education*, *47*(1), 161–184.

Herman, B. C., Owens, D. C., Oertli, R. T., Zangori, L. A., & Newton, M. H. (2019). Exploring the complexity of students' scientific explanations and associated nature of science views within a place-based socioscientific issue context. *Science & Education*, *28*(3), 329–366.

HM Government. (2021). Press release: Teacher assessed grades for students: Fairness and flexibility at the heart of plans to ensure young people can progress to next stage of education or career. Retrieved 1 March 2021 from https://www.gov.uk/government/news/teacher-assessed-grades-for-students

Hodson, D. (2014). Nature of science in the science curriculum: Origin, development, implications and shifting emphases. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 911–969). Dordrecht: Springer.

Hodson, D., & Wong, S. L. (2014). From the horse's mouth: Why scientists' views are crucial to nature of science understanding. *International Journal of Science Education*, *36*(16), 2639–2665.

Hodson, D., & Wong, S. L. (2017). Going beyond the consensus view: Broadening and enriching the scope of NOS-oriented curricula. *Canadian Journal of Science, Mathematics and Technology Education*, *17*(1), 3–17.

Holstein, J. A. & Gubrium, J. F. (2003). Inside interviewing: New lenses, new concerns. In Holstein, J. A., & Gubrium, J. F. (Eds.) *Inside interviewing* (pp. 2–30). Thousand Oaks, CA: Sage.

Höttecke, D., & Allchin, D. (2020). Reconceptualizing nature-of-science education in the age of social media. *Science Education*, *104*(4), 641–666.

Höttecke, D., & Silva, C. C. (2011). Why implementing history and philosophy in school science education is a challenge: An analysis of obstacles. *Science & Education*, *20*(3), 293–316.

Im, E.-B. (2019, January 10). 'Sky Castle' discusses realities of parenthood behind education craze. The Korea Herald. http://m.koreaherald.com/view.php?ud=20190110000747#cb

Im, S., Yoon, H. G., & Cha, J. (2016). Pre-service science teacher education system in South Korea: Prospects and challenges. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(7), 1863–1880.

Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education*, *20*(7–8), 591–607.

Irzik, G., & Nola, R. (2014). New directions for nature of science research. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 999–1021). Dordrecht: Springer.

Jesson, R. (2018). Stanines. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement and evaluation* (pp. 1610–1611). Thousand Oaks, CA: Sage.

Jo, S. (2019, August 7). Bizarre science education standards with "science" disappeared and "society" emphasised. Donga Science. http://dongascience.donga.com/news.php?idx=30432

Kampourakis, K. (2016). The 'general aspects' conceptualization as a pragmatic and effective means to introducing students to nature of science. *Journal of Research in Science Teaching*, *53*(5), 667–682.

Kaya, E., & Erduran, S. (2016). From FRA to RFN, or how the Family Resemblance Approach can be transformed for science curriculum analysis on nature of science. *Science & Education*, *25*(9–10), 1115–1133.

Kaya, E., Erduran, S., Akgün, S., & Aksöz, B. (2017). Nature of science in teacher education: A holistic approach. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *11*(2), 464–501.

Kaya, E., Erduran, S., Aksoz, B., & Akgun, S. (2019). Reconceptualised family resemblance approach to nature of science in pre-service science teacher education. *International Journal of Science Education*, *41*(1), 21–47.

Kelly, G. J., Chen, C., & Crawford, T. (1998). Methodological considerations for studying science-in-the-making in educational settings. *Research in Science Education*, *28*(1), 23–49.

Khishfe, R. (2012a). Nature of science and decision-making. *International Journal of Science Education*, *34*(1), 67–100.

Khishfe, R. (2012b). 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., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, *39*(7), 551–578.

Kiang K. M., & Colanero, K. (2020). A classics reading approach to nurture epistemic insight in a multidisciplinary and higher education context. In T. Teo, A. L. Tan, & Y. S. Ong (Eds.), *Science education in the 21st century* (pp. 51–65). Singapore: Springer.

Kim, I.-H., Cha, J., Kim, C., & Kim, H. (2010). Difficulties experienced by preservice science teachers in studying the theory f science education for teacher selection test. *Journal of Korean Association for Science Education*, *30*(4), 429–436.

Kim, S. Y., Yi, S. W., & Cho, E. H. (2014). Production of a science documentary and its usefulness in teaching the nature of science: Indirect experience of how science works. Science & Education, 23(5), 1197–1216.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. Thousand Oaks, CA: Sage.

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society*, *33*(2), 211–221.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science*, 328(5977), 456–459.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Kuhn, T. S. (1974). Second thoughts on paradigms. In F. Suppe (Ed), *The structure of scientific theories* (pp. 459–482). Urbana, IL: University of Illinois Press.

Kuhn, T. S. (2000). *The road since Structure*. Chicago: University of Chicago Press.

Kwak, Y. (2020). Trend analysis of curriculum application status of the 2015 revised integrated science and scientific laboratory experiment curriculum. *Journal of Korean Society of Earth Science Education*, *13*(1), 53–63.

Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, *24*(1), 60–77.

Laplane, L., Mantovani, P., Adolphs, R., Chang, H., Mantovani, A., McFall-Ngai, M., et al. (2019). Why science needs philosophy. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 3948–3952.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ Princeton University Press.

Lawson, A. E. (1982). The nature of advanced reasoning and science instruction. *Journal of Research in Science Teaching*, *19*(9), 743–760.

Leblebicioglu, G., Abik, N. M., Capkinoglu, E., Metin, D., Dogan, E. E., Cetin, P. S., & Schwartz, R. (2019). Science camps for introducing nature of scientific inquiry through student inquiries in nature: Two applications with retention study. *Research in Science Education*, *49*(5), 1231–1255.

Leblebicioglu, G., Metin, D., Capkinoglu, E., Cetin, P. S., Dogan, E. E., & Schwartz, R. (2017). Changes in students' views about nature of scientific inquiry at a science camp. *Science & Education*, *26*(7–9), 889–917.

Lederman, N. G. (1995). Translation and transformation of teachers' understanding of NOS into classroom practice. Paper presented at the meeting of the National Association of Research in Science Teaching, San Francisco, CA.

Lederman, J. S., & Khishfe, R. (2002). *Views of nature of science, form D2*. Illinois Institute of Technology.

Lederman, J. S., Lederman, N. G., Bartels, S., Jimenez, J., Akubo, M., Aly, S., … Zhou, Q. (2019). An international collaborative investigation of beginning seventh grade students' understandings of scientific inquiry: Establishing a baseline. *Journal of Research in Science Teaching*, *56*(4), 486–515

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching*, *51*(1), 65–83.

Lederman, N. G. (1999). Teachers' understanding of the nature of science and classroom practice: Factors that facilitate or impede the relationship. *Journal of Research in Science Teaching*, *36*(8), 916–929.

Lederman, N. G. (2006). Syntax of nature of science within inquiry and science instruction. In W. McComas (Ed.) *Scientific inquiry and nature of science* (pp. 301–317). Dordrecht: Springer.

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–880). Mahwah, NJ: Erlbaum.

Lederman, N. G., & Zeidler, D. L. (1987). Science teachers' conceptions of NOS: Do they really influence teacher behavior? *Science Education*, *71*(5), 721–734.

Lederman, N. G., Abd-El-Khalick, F., & Lederman, J. S. (2020). Avoiding de-natured science: Integrating nature of science into science instruction. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 295–326). Dordrecht: Springer.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521.

Lederman, N. G., Schwartz, R. S., Abd-El-Khalick, F., & Bell, R. L. (2001). Pre-service teachers' understanding and teaching of nature of science: An intervention study. *Canadian Journal of Science, Mathematics and Technology Education*, *1*(2), 135–160.

Lee, B., Shim, K.-C., Shin, M.-K., Kim, J., Park, E., Yoon, J., Kwon, Y.-J., & Kim, Y.-J. (2013). Analysis of science education theories in the question items of the examination for appointing secondary school science teachers. *Journal of Korean Association for Science Education*, *33*(4), 794–806.

Lee, H., & Witz, K. G. (2009). Science teachers' inspiration for teaching socio-scientific issues: Disconnection with reform efforts. *International Journal of Science Education*, *31*(7), 931–960.

Leite, L. (2002). History of science in science education: Development and validation of a checklist for analysing the historical content of science textbooks. *Science & Education*, *11*(4), 333–359.

Li, X., Tan, Z., Shen, J., Hu, W., Chen, Y., & Wang, J. (2018). Analysis of five junior high school physics textbooks used in China for representations of nature of science. *Research in Science Education*. Advance online publication. https://doi.org/10.1007/s11165-018-9713-z

Lipset, S. M., Trow, M. A., & Coleman, J. S. (1956). Union democracy: the internal politics of the International Typographical Union. Garden City, NY: Anchor.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442–467.

Looney, J. W. (2011). *Integrating Formative and Summative Assessment: Progress toward a Seamless System?* OECD Education Working Papers (No. 58). Paris: OECD Publishing. Retrieved 1 March 2021 from https://www.oecd-ilibrary.org/education/integrating-formative-and-summative-assessment_5kghx3kbl734-en

Lund, A. (2008). Assessment made visible: Individual and collective practices. *Mind, Culture, and Activity*, *15*(1), 32–51.

Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education*, *35*(7), 1208–1229.

Ma, H. (2012). *The images of science through cultural lenses: A Chinese study*. Rotterdam: Sense Publishers.

Magnusson, S., Krajcik, J., & Borko, H. (2006). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *PCK and science education* (pp. 95–132). Dordrecht: Kluwer Academic Publishers.

Malik, S. (2017). Observation versus experiment: an adequate framework for analysing scientific experimentation?. *Journal of General Philosophy of Science*, *48*(1), 71–95.

Mamlok-Naaman, R., Ben-Zvi, R., Hofstein, A., Menis, J., & Erduran, S. (2005). Learning science through a historical approach: Does it affect the attitudes of non-science-oriented students towards science? *International Journal of Science and Mathematics Education*, *3*(3), 485–507.

Mansour, N., Wegerif, R., Skinner, N., Postlethwaite, K., & Hetherington, L. (2016). Investigating and promoting trainee science teachers' conceptual change of the nature of science with digital dialogue games 'InterLoc'. *Research in Science Education*, *46*(5), 667–684.

Marso, R. N., & Pigge, F. L. (1993). Teachers' testing knowledge, skills, and practice. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 129–185). Lincoln , NE : Buros Institute of Mental Measurements.

Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, *14*(2), 78–102.

Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge: Cambridge University Press.

Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. London: Routledge.

Matthews, M. R. (2012). Changing the focus: From nature of science (NOS) to features of science (FOS). In M. S. Khine (Ed.), *Advances in nature of science research: Concepts and methodologies* (pp. 3–26). Dordrecht: Springer.

Matthews, M. R. (2014). *Science teaching: The contribution of history and philosophy of science* (2nd ed.). London: Routledge.

Mauthner, N. S., & Doucet, A. (2003). Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology*, *37*(3), 413–431.

McComas, W. F. (2014). Nature of science in the science curriculum and in teacher education programs in the United States. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1993–2023). Dordrecht: Springer.

McComas, W. F. (2020). Nature of science and classroom practice: A review of the literature with implication for effective NOS instruction. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 67–111). Dordrecht: Springer.

McComas, W. F., & Olson, J. K. (2002). The nature of science in international science education standards documents. In *The nature of science in science education* (pp. 41–52). Dordrecht: Springer.

McComas, W. F., Clougn, M. P. & Nouri, N. (2020). Nature of science and classroom practice: A review of literature with implications for effective NOS instruction. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 67–111). Dordrecht: Springer.

McDermott, R. P. (1999) On becoming labelled–The story of Adam. In P. Murphy (Ed.), *Learners, learning and assessment* (pp. 1–21). London: Paul Chapman.

McDonald, C. V. (2017). Exploring representations of nature of science in Australian junior secondary school science textbooks: A case study of Genetics. In C. V. McDonald (Ed.), *Representations of nature of science in school science textbooks:*

A global perspective (pp. 98–117). London: Routledge.

McDonald, C. V., & Abd-El-Khalick, F. (Eds.). (2017). *Representations of nature of science in school science textbooks: a global perspective*. London: Routledge.

McGregor, D., Baskerville, D., Anderson, D., & Duggan, A. (2019). Examining the use of drama to develop epistemological understanding about the nature of science: a collective case from experience in New Zealand and England. *International Journal of Science Education Part B, 9*(2), 171–194.

McKinney, S. E., Chappell, S., Berry, R. Q., & Hickman, B. T. (2009). An examination of the instructional practices of mathematics teachers in urban schools. *Preventing School Failure: Alternative Education for Children and Youth*, *53*(4), 278–284.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32.

McMillan, J. H. (2018). *Classroom assessment: Principles and practice that enhance student learning and motivation*. New York, NY: Pearson.

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40-49.

McMullin, E. (1993). Rationality and paradigm change in science. In P. Horwich (Ed.), World changes: Thomas Kuhn and the nature of science (pp. 55–78). Cambridge, MA: MIT Press.

Mehan, H. (1982). The structure of classroom events and their consequences for student performance. In P. Gilmore & A. A. Glatthorn (Eds.), *Children in and out of school*. Washington, DC: Center for Applied Linguistics.

Merriam, S. B., Johnson-Bailey, J., Lee, M. Y., Kee, Y., Ntseane, G., & Muhamad, M. (2001). Power and positionality: Negotiating insider/outsider status within and across cultures. *International Journal of Lifelong Education*, *20*(5), 405–416.

Mesci, G., Schwartz, R. S., & Pleasants, B. A. S. (2020). Enabling factors of preservice science teachers’ pedagogical content knowledge for nature of science and nature of scientific inquiry. *Science & Education*, *29*(2), 263–297.

Metz, D. (2014). The history and philosophy of science in science curricula and teacher education in Canada. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 2025–2044). Dordrecht: Springer.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook and the coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.

Millar, R., & Osborne, J. (1998). *Beyond 2000: Science education for the future: A report with ten recommendations*. London: King's College London School of Education.

Ministry of Education (MOE). (1988). *National curriculum: Science*. Seoul: MOE.

Ministry of Education (MOE). (2015). *National curriculum: Science*. Sejong: MOE.

Ministry of Education, Science and Technology (MEST). (2009). *National curriculum: Science*. Seoul: MEST.

Morse, J. M. (1997). 'Perfectly healthy, but dead': The myth of inter-rater reliability. *Qualitative Health Research*, *7*(4), 445–447.

Moss, C. M. (2013). *Research on classroom summative assessment*. In J. H. McMillan (Ed.), SAGE Handbook of research on classroom assessment (p. 235–256). Thousands Oaks, CA: Sage.

Mpemba, E. B., & Osborne, D. G. (1969). Cool?. *Physics Education*, *4*(3), 172.

Mulvey, B. K., Parrish, J. C., Reid, J. W., Papa, J., & Peters-Burton, E. E. (2021). Making Connections: Using individual epistemic network analysis to extend the value of nature of science assessment. *Science & Education*, *30*(3), 527–555.

Murakami, Y. & Sumida, M. (2014). History and philosophy of science in Japanese Education: A historical overview. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 2217–2246). Dordrecht: Springer.

National Council on Measurement in Education. (NCME). (2018). Position statement on K-12 classroom assessment. Retrieved from https://higherlogicdownload.s3.amazonaws.com/NCME/c53581e4-9882-4137-987b-

4475f6cb502a/UploadedImages/Classroom_assessment_position_statement_-_Board_approved.pdf

National Research Council. (NRC) (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (NRC) (2001). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (NRC) (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academy Press.

Ng, W. (2011). Why digital literacy is important for science teaching and learning. *Teaching Science*, *57*(4), 26–32.

NGSS Lead States. (2013). *Next generation science standards*. Washington DC: National Academy Press.

Niaz, M. (2012). *From 'science in the making' to understanding the nature of science: An overview for science educators.* London: Routledge.

Niaz, M. (2016). History and philosophy of science as a guide to understanding nature of science. *Revista científica*, *1*(24), 7–16.

Nickles, T. (2017). Scientific Revolutions. In N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 ed.). Retrieved 2 March 2021 from https://plato.stanford.edu/archives/win2017/entries/scientific-revolutions/

Nieminen, P., Correia, C. F., Hähkiöniemi, M., Serret, N., Viiri, J., & Harrison, C. (2016). Formative assessment in inquiry-based science education using interactions on-the-fly. Paper presented in NARST Annual International conference, Baltimore, MD.

Nieminen, P., Hähkiöniemi, M., & Viiri, J. (2020). Forms and functions of on-the-fly formative assessment conversations in physics inquiry lessons. *International Journal of Science Education*. Advance online publication. https://doi.org/10.1080/09500693.2020.1713417

Nikou, S. A., & Economides, A. A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior*, *55*(B), 1241–1248.

Noh, T., Lee, J., Kang, S., & Kang, H. (2015). Secondary school science teachers' actual and preferred types of assessment. *Journal of Korean Association for Science Education*, *35*(4), 725–733.

Nussbaum, M. (1997). *Cultivating humanity: A classical defense of reform in liberal education*. Cambridge, MA: Harvard University Press.

O'Kane, P., Smith, A., & Lerman, M. P. (2019). Building Transparency and Trustworthiness in Inductive Research Through Computer-Aided Qualitative Data Analysis Software. *Organizational Research Methods*. Online first publication. https://doi.org/10.1177/1094428119865016

OECD. (2019). *TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners*. Paris: OECD Publishing. https://doi.org/10.1787/1d0bc92a-en

OECD. (2020). *PISA 2024 Strategic Direction and Vision for Science*. Paris: OECD Publishing. Retrieved 1 April 2020 from https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf

Oldham, V., & Brouwer, W. (1984). Mendelian genetics: Paradigm, conjecture, or research program. *Journal of Research in Science Teaching*, *21*(6), 623–637.

Olson, J. K. (2018). The inclusion of the nature of science in nine recent international science education standards documents. *Science & Education, 27*(7–8), 637–660.

Osborne, J., Sedlacek, Q. C., Friend, M., & Lemmi, C. (2016). Learning to read science. *Science Scope*, *40*(3), 36–42.

Osborne, J., Sedlacek, Q. C., Friend, M., & Lemmi, C. (2016). Learning to Read Science. *Science Scope*, *40*(3), 36–42. https://doi.org/10.5949/liverpool/9781781382615.003.0002

Oxford English Dictionary. (n.d.). Stakeholder. Retrieved 24 February 2021 from https://www.oed.com/view/Entry/246856?redirectedFrom=stakeholder#eid

Park, H. (2018, March 28). More money spent on private education as early decisions increase. Seoul Shinmun. http://www.seoul.co.kr/news/newsView.php?id=20180326500142

Park, M. (n.d). School District 8. The encyclopedia of Gangnam-gu. Retrieved 28 September 2020 from http://e-gangnam.grandculture.net/Contents?local=e-gangnam&dataType=01&contents_id=EC04801558

Park, W. (2020). Beyond the 'two cultures' in the teaching of disaster: or how disaster education and science education could benefit each other. *Educational Philosophy and Theory, 52*(13), 1434–1448.

Park, W., & Song, J. (2019). Between realism and constructivism: A sketch of pluralism for science education. In E. Herring, K. Jones, K. Kiprijanov, & L. Sellers (Eds.), *The past, present and future of integrated history and philosophy of science* (pp. 228–247). London: Routledge.

Park, W., & Song, J. (2019). Between realism and constructivism: A sketch of pluralism for science education. In E. Herring, K. Jones, K. Kiprijanov, & L. Sellers (Eds.), *The past, present, and future of integrated history and philosophy of science* (pp. 228–247). Routledge. https://doi.org/10.4324/9781351214827-14

Park, W., Park, J., & Song, J. (2016). *Goethe's Theory of Colours and its implications for science education*. Manuscript in preparation.

Park, W., Wu, J. Y., & Erduran, S. (2020). The nature of STEM disciplines in the science education standards documents from the USA, Korea and Taiwan: Focusing on disciplinary aims, values and practices. *Science & Education, 29*(4), 899–927.

Park, W., Yang, S., & Song, J. (2019). When Modern Physics Meets Nature of Science: The Representation of Nature of Science in General Relativity in New Korean Physics Textbooks. *Science & Education*, *28*(9–10), 1055–1083.

Park, W., Yang, S., & Song, J. (2020). Eliciting students' understanding of nature of science with text-based tasks: insights from new Korean high school textbooks. *International Journal of Science Education*, *42*(3), 426–450. Park, W., & Song, J. (2018). Goethe's Conception of "Experiment as Mediator" and Implications for Practical Work in School Science. *Science & Education*, *27*(1–2), 39–61.

Parkes, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed.), *SAGE Handbook on research on classroom assessment* (pp. 107–124). Thousand Oaks, CA: Sage.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5), 1189–1208.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage.

Pederson, J. E., Isozaki, T., & Hirano, T. (Eds.). (2017). *Model science teacher preparation programes: An international comparison of what works*. Charlotte, NC: Information Age Publishing.

Peräkylä, A. (2004). Reliability and validity in research based on naturally occurring social interaction. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (2nd ed.) (pp. 283–304). Thousand Oaks, CA: Sage.

Peters-Burton, E. E., Parrish, J. C., & Mulvey, B. K. (2019). Extending the Utility of the Views of Nature of Science Assessment through Epistemic Network Analysis. *Science & Education*, *28*(9–10), 1027–1053.

Pickering, A. (1992). *Science as practice and culture*. Chicago, IL: University of Chicago Press.

Pleasants, J., & Olson, J. K. (2019). What is engineering? Elaborating the nature of engineering for K-12 education. *Science Education*, *103*(1), 145–166.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, *47*(11), 1451–1458.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*(1), 4–11.

Preece, P. F., & Skinner, N. C. (1999). The national assessment in science at Key Stage 3 in England and Wales and its impact on teaching and learning. *Assessment in Education: Principles, Policy & Practice*, *6*(1), 11–25.

Price, R., & Perez, K. (2018). Many paths toward discovery: A module for teaching how science works. *Journal of College Science Teaching*, *47*(3), 78–87.

Pring, R. (2000). The 'false dualism'of educational research. *Journal of Philosophy of Education*, *34*(2), 247–260.

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20.

Punch, M. (1986). *The politics and ethics of fieldwork*. Beverly Hills, CA: Sage

Quigley, C., Pongsanon, K., & Akerson, V. L. (2011). If we teach them, they can learn: Young students views of nature of science during an informal science education program. *Journal of Science Teacher Education*, *22*(2), 129–149.

Rached, E., & Grangeat, M. (2020). French teachers' informal formative assessment in the context of inquiry-based learning. *International Journal of Science Education*. Online first publication. https://doi.org/10.1080/09500693.2020.1740818

Reiss, M. J. (2020). Science education in the light of COVID-19. *Science & Education*, *29*(4), 1079–1092.

Reiss, M. J., & McComas, W. F. (2020). Informal learning sites and their role in communicating the nature of science. In W. F. McComas (Ed.), *Nature of science in science instruction: Rationales and strategies* (pp. 711–730). Dordrecht: Springer.

Rieg, S. A. (2007). Classroom assessment strategies: What do students at-risk and teachers perceive as effective and useful? *Journal of Instructional Psychology*, *34*(4), 214–225.

Roehrig, G. H., & Garrow, S. (2007). The impact of teacher classroom practices on student achievement during the implementation of a reform-based chemistry curriculum. *International Journal of Science Education*, *29*(14), 1789–1811.

Roehrig, G. H., Kruse, R. A., & Kern, A. (2007). Teacher and school characteristics and their influence on curriculum implementation. *Journal of Research in Science Teaching*, *44*(7), 883–907.

Romero-Maltrana, D., & Duarte, S. (2020). A new way to explore the nature of science: Meta-categories rather than lists. *Research in Science Education*. Online first publication. https://doi.org/10.1007/s11165-020-09940-y

Rosen, L. H. & Underwood, M. K. (2012). Observations. In Salkind, N. J. (Ed.), *Encyclopedia of research design*. Retrieved 28 September from https://sk.sagepub.com/reference/researchdesign

Rowe, M.B. (1974). A humanistic intent: The program of preservice elementary education at the University of Florida. *Science Education*, *58*(3), 369–376.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation, 37*(1), 15–24.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, *11*(3–4), 237–263.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, *44*(1), 57–84.

Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education*, *36*(1), 1–44.

Ryder, J. (2009). Enhancing engagement with science/technology-related issues. In A. T. Jones & M. J. de Vries (Eds.), *International handbook for research and development in technology education* (pp. 287–296). Rotterdam: Sense Publishers.

Ryder, J., & Banner, I. (2013) School teachers' experiences of science curriculum reform. *International Journal of Science Education*, *35*(3), 490–514.

Ryder, J., & Leach, J. (2008). Teaching about the epistemology of science in upper secondary schools: An analysis of teachers' classroom talk. *Science & Education*, *17*(2), 289–315.

Sadler, T. D. (2014). Assessment of socio-scientific reasoning: Linking progressive aims of science education to the realities of modern education. In M. P. Mueller, D. J. Tippins, & A. J. Stewart (Eds.), *Assessing schools for Generation R (responsibility)* (pp. 101–113). Dordrecht: Springer.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.

Sandoval, W. A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. *Journal of Research in Science Teaching*, *40*(4), 369–392.

Sankey, H. (1993). Kuhn's changing concept of incommensurability. *The British Journal for the Philosophy of Science*, *44*(4), 759–774.

Savin-Baden, M., Major, C. H. (2013). *Qualitative research: The essential guide to theory and practice*. Abingdon: Routledge.

Schellinger, J., Mendenhall, A., Alemanne, N., Southerland, S. A., Sampson, V., & Marty, P. (2019). Using technology-enhanced inquiry-based instruction to foster the development of elementary students' views on the nature of science. *Journal of Science Education and Technology*, *28*(4), 341–352.

Schizas, D., Psillos, D., & Stamou, G. (2016). Nature of Science or Nature of the Sciences? *Science Education*, *100*(4), 706–733.

Schuman, D. (1982). *Policy analysis, education, and everyday life*. Lexington, MA: Heath.

Schwab, J. J. (1962). *The teaching of science as enquiry*. Cambridge, MA: Harvard University Press.

Schwartz, R. S., & Crawford, B. A. (2006). Authentic scientific inquiry as context for teaching nature of science: Identifying critical element. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science* (pp. 331–355). Dordrecht: Springer.

Schwartz, R. S., & Lederman, N. G. (2002). "It's the nature of the beast": The influence of knowledge and intentions on learning and teaching nature of science. *Journal of Research in Science Teaching*, *39*(3), 205–236.

Schwartz, R. S., & Lederman, N. G. (2008). What scientists say: Scientists' views of nature of science and relation to science context. *International Journal of Science Education*, *30*(6), 727–771.

Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, *88*(4), 610–645.

Schwartz, R. S., Lederman, N. G., Khishfe, R., Lederman, J. S., Matthews, L., & Liu, S.-Y. (2002). Explicit/reflective instructional attention to nature of science and scientific inquiry: Impact on student learning. Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science. Retrieved from https://eric.ed.gov/?id=ED465622

Seedhouse, P., & Richards, K. (Eds.). (2005). *Applying conversation analysis*. London: Palgrave Macmillan.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York, NY: Teachers College Press.

Shapere, D. (1971). The paradigm concept. *Science*, *172*(3984), 706–709.

Shapiro, B. (1994) What children bring to light: A constructivist perspective on children's learning in science. New York, NY: Teachers College Press.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ... & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied measurement in education*, *21*(4), 295–314.

Shin, K. (2019, January 23). China can’t hide its love for ‘SKY Castle’: Without a distribution rights holder, pirating is only way to watch. Korea JoongAng Daily. https://koreajoongangdaily.joins.com/news/article/article.aspx?aid=3058571&cloc=joongangdaily

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Education Review*, *57*(1), 1–22.

Silverman, D. (2014). *Interpreting qualitative* data (5th ed.). Thousand Oaks, CA: Sage.

Simons, H. (2014). Case study research: In-depth understanding in context. In P. Leavy

(Ed.), *The Oxford handbook of qualitative research* (pp. 455–470). Oxford: Oxford University Press.

Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse*. Oxford: Oxford University Press.

Smith, C., Wiser, M., Anderson, C.W., & Krajcik, J. (2006). Implications for children's learning for assessment: A proposed learning progression for matter and the atomic molecular theory. *Measurement*, *14*(1–2), 1–98.

Solbes, J., & Traver, M. (2003). Against a negative image of science: History of science and the teaching of physics and chemistry. *Science & Education*, *12*(7), 703–717.

Son, J. (2016). Perception of pre-service science teachers for the possibility of common science majors' preparation of the 2015 science curriculum 'integrated science' and 'scientific inquiry experiments'. *Teacher Education Research*, *55*(4), 472–484.

Song, J. (2013). The disparity between achievement and engagement in students' science learning: A case of East-Asian regions. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 285–306). Springer Netherlands.

Song, J., & Joung, Y. J. (2014). Trends in HPS/NOS Research in Korean Science Education. In *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 2177–2215). Springer.

Songer, N. B. and Linn, M. C. (1991) How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28* (9), 71–84.

Sorensen, C. W. (1994). Success and education in South Korea. *Comparative Education Review*, *38*(1), 10–35.

Sorenson, C. W. (1994). Success and education in South Korea. *Comparative Education Review*, *38*(1), 10–35.

Stanley, M. (2016). Why should physicists study history? *Physics Today*, *69*(7), 36–44.

Stiggins, R. J., Frisbie, R. J., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and*

Practice, *8*(2), 5–14.

Stobart, G. (2008) *Testing times: The uses and abuses of assessment*. London: Routledge.

Summers, R., Alameh, S., Brunner, J., Maddux, J. M., Wallon, R. C., & Abd-El-Khalick, F. (2019). Representations of nature of science in U.S. science standards: A historical account with contemporary implications. *Journal of Research in Science Teaching*. Advance online publication. https://doi.org/10.1002/tea.21551

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 399-417.

The Royal Society Science Policy Centre. (2014). *Vision for science and mathematics education*. London: The Royal Society.

Thomas, G. (2013). (Ed.). *Case study methods in education*. Thousand Oaks, CA: Sage.

Thornberg, R., & Charmaz, K. (2014). Grounded theory and theoretical coding. In Flick, U. (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 153–170). Thousand Oaks, CA: Sage.

Tilley, L., & Woodthorpe, K. (2011). Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research*, *11*(2), 197–212.

Tobin, K., & McRobbie, C. J. (1996). Cultural myths as constraints to the enacted science curriculum. *Science Education*, *80*(2), 223–241.

Topcu, M. S. (2013). Preservice teachers' epistemological beliefs in physics, chemistry, and biology: A mixed study. *International Journal of Science and Mathematics Education*, *11*(2), 433–458.

Topping, K. J. (2013) Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. 395–412). London: Sage.

Toulmin, S. (1972). *Human understanding*. Oxford: Clarendon Press.

Twinn, S. (1997). An exploratory study examining the influence of translation on the validity and reliability of qualitative data in nursing research. *Journal of Advanced Nursing*, *26*(2), 418–423.

Tyler, R. W. (1949) *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.

van Dijk, E. M. (2011). Portraying real science in science communication. *Science Education*, *95*(6), 1086–1100.

Vanblaere, B., & Devos, G. (2016). Relating school leadership to perceived professional learning community characteristics: A multilevel analysis. *Teaching and Teacher Education, 57*(1), 26–38.

Veal, W. R., & MaKinster, J. G. (1999). Pedagogical content knowledge taxonomies. *The Electronic Journal for Research in Science & Mathematics Education*, *3*(4). https://doi.org/10.1016/B978-0-08-044894-7.00642-4VERBI Software. (2019). MAXQDA 2020 [computer software]. Berlin, Germany: VERBI Software. Available from maxqda.com.

Wahbeh, N., & Abd-El-Khalick, F. (2014). Revisiting the translation of nature of science understandings into instructional practice: Teachers' nature of science pedagogical content knowledge. *International Journal of Science Education*, *36*(3), 425–466.

Waight, N., & Abd-El-Khalick, F. (2012). Nature of technology: Implications for design, development, and enactment of technological tools in school science classrooms. *International Journal of Science Education*, *34*(18), 2875–2905.

Walford, G. (2005). Research ethical guidelines and anonymity. *International Journal of Research and Method in Education*, *28*(1), 83–93.

Walker, K. A., & Zeidler, D. L. (2003). Students' understanding of the nature of science and their reasoning on socioscientific issues: A web-based learning inquiry. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (Philadelphia, PA, March 23–26, 2003).

Waters-Adams, S. (2006). The relationship between understanding of the nature of science and practice: The influence of teachers' beliefs about education, teaching and learning. *International Journal of Science Education*, *28*(8), 919–944.

Weiss, K. M. (1996). Is there a paradigm shift in genetics? Lessons from the study of human diseases. *Molecular Phylogenetics and Evolution*, *5*(1), 259–265.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Wenger-Trayner, E., & Wenger-Trayner, B. (2015). Introduction to communities of practice: A brief overview of the concept and its uses. https://wenger-trayner.com/introduction-to-communities-of-practice/

Westaway, F. W. (1929). *Science teaching*. London: Blackie & Son.

Whittaker, F., & Booth, S. (2021, January 6). Teacher assessment grades to replace GCSE and A-level exams in 2021. Schools Week. https://schoolsweek.co.uk/teacher-assessment-grades-to-replace-gcse-and-a-level-exams-in-2021/

Wiliam, D. (2010). Standardized Testing and School Accountability. *Educational Psychologist*, *45*(2), 107–122.

Wilkinson, J. (1999). A quantitative analysis of physics textbooks for scientific literacy themes. *Research in Science Education*, *29*(3), 385–399.

Willis, J. (2010). Assessment for learning as a participative pedagogy. *Assessment Matters*, *2*(1), 65–84.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241–256.

Wolf, D., Bixby, J., Glenn, H., & Gardner, H. (1991) To use their minds well: Investigation of new forms of students' assessment. *Review of Research in Education*, *17*(1), 31–74.

Wong, S. L., Hodson, D., Kwan, J., & Yung, B. H. W. (2008). Turning crisis into opportunity: Enhancing student-teachers' understanding of nature of science and scientific inquiry through a case study of the scientific research in severe acute respiratory syndrome. *International Journal of Science Education, 30*(11), 1417–1439.

Wong, S. L., Kwan, J., Hodson, D., & Yung, B. H. W. (2009). Turning crisis into opportunity: Nature of science and scientific inquiry as illustrated in the scientific research on severe acute respiratory syndrome. *Science & Education*, *18*(1), 95–118.

Wong, S. L., Wan, Z. H., & Cheng, K. L. (2014). One country, two systems: Nature of science education in Mainland China and Hong Kong. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 2149–2175). Springer.

Woolcock, N. (2021, March 26). It's time to abolish GCSEs, says biggest teaching union. The Times. https://www.thetimes.co.uk/article/abolish-gcses-says-biggest-teaching-union-njmmxw2d7

Wyatt-Smith, C. M., & Cumming, J. J. (2003). Curriculum literacies: Expanding domains of assessment. *Assessment in Education: Principles, Policy & Practice,* 10(1), 47–59.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162.

Yacoubian, H. A. (2015). A framework for guiding future citizens to think critically about nature of science and socioscientific issues. *Canadian Journal of Science, Mathematics and Technology Education, 15*(3), 248–260.

Yacoubian, H. A., & Boujaoude, S. (2010). The effect of reflective discussions following inquiry-based laboratory activities on students' views of nature of science. *Journal of Research in Science Teaching*, *47*(10), 1229–1252.

Yacoubian, H. A., & Hansson, L. (2020) (Eds.). *Nature of science for social justice*. Dordrecht: Springer.

Yang, S., Park, W., & Song, J. (2019). Representations of nature of science in new Korean science textbooks: The case of 'scientific inquiry and experimentation.' In A. L. Tan, T. W. Teo, & Y. S. Ong (Eds.), *Science education in the 21st century:*

Re-searching issues that matters from different lenses (pp. 19–35). Singapore: Springer.

Yang, S., Park, W., & Song, J. (2020). Representations of nature of science in new Korean science textbooks: The case of 'scientific inquiry and experimentation.' In A. L. Tan, T. W. Teo, & Y. S. Ong (Eds.), *Science education in the 21st century: Re-searching issues that matters from different lenses* (pp. 19–35). Springer.

Yin, R. K. (2010). Analytic generalisation. In A. J. Mills, G. Durepos & E. Wiebe (Eds.), Encyclopedia of case study research. Retrieved 28 September 2020 from https://sk.sagepub.com/reference/casestudy/n8.xml

Yin, R. K. (2015). *Qualitative research from start to finish*. New York, NY: Guilford Publications.

Yin, R. K. (2017). *Case study research and applications* (6th ed.). Thousand Oaks: Sage.

Yoon, J., & Kang, S.-J. (2016). The analysis of high school science teachers' expectations and concerns on the Integrated Science and Science Inquiry Experiment subjects in 2015 revised national curriculum. *The Journal of Learner-Centered Curriculum and Instruction*, *16*(5), 515–546.

Yu, H. (Writer), Jo, H., & Kim, D. (Directors) (2019). SKY Castle. [TV Series]. JTBC.

Yung, B. H. W. (2006). *Assessment reform in science: Fairness and fear*. Dordrecht: Springer.

Zeidler, D. L., Applebaum, S. M., Sadler, T. D. (2011). Enacting a socioscientific issues classroom: transformative transformations. In Sadler, T. D. (Ed.). *Socio-scientific issues in science classrooms: Teaching, learning and research* (pp. 277–306). Dordrecht: Springer.

Appendix A Ethical approval and supporting documents

From: Lyudmila Nurse <**lyudmila.nurse@education.ox.ac.uk**>
Sent: Tuesday, March 19, 2019 3:49 PM
To: Sibel Erduran <**sibel.erduran@education.ox.ac.uk**>
Cc: Wonyong Park <**wonyong.park@education.ox.ac.uk**>; Education Research Office <**research.office@education.ox.ac.uk**>
Subject: CUREC Approval

Dear Professor Erduran,

Title and reference number: Physics Teachers' Experiences of Implementing a Science Curriculum Reform Explicitly Focused on Nature of Science - ED-CIA-19-217

The above application has now been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse **http://www.nspcc.org.uk/Inform/research/questions/reporting_child_abuse_wda74908.html**

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to **research.office@education.ox.ac.uk** for consideration.

Good luck with your research study.

Yours sincerely,

Lyudmila

Name Lyudmila Nurse
Title Dr

Dr Lyudmila Nurse
Research Fellow
Department of Education

Note: This is an initial approval for the first-year data collection. An extension to this approval was granted later on 25th September 2019 to cover the additional data collection in the second year. This initial approval had been granted for collecting data from physics teachers, but with the approval of the thesis supervisors and DREC, the ultimate data sources also included chemistry and earth science teachers.

Appendix A. Letter to Headteacher

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

Principal investigator: Professor Sibel Erduran
sibel.erduran@education.ox.ac.uk
Primary researcher: Wonyong Park (DPhil student)
wonyong.park@education.ox.ac.uk

[Head teacher name]
[School name and address]
[Date]

Dear [Head teacher name],

I am writing to enquire about conducting some research in your school in the next academic year. I am a doctoral research student at Oxford University, supervised by Professor Sibel Erduran (Professor of Science Education) and Dr Judith Hillier (Associate Professor in Science Education [Physics]). In my research project, *Physics Teachers' Experiences of Teaching a Science Curriculum Reform Explicitly Focused on Nature of Science*, I will explore the experiences of physics teachers while they teach scientific inquiry and experimentation (SIE) and the perceived meanings of nature of science by teachers.

The research will involve a physics teacher and one of their SIE classes in Spring 2019 semester. I am not aiming to change what or how the teacher chooses to teach and will not be making any judgements about teaching. My research focus is on teachers' experiences of curriculum implementation.

By participating in the research, your school would be contributing to a project that will deepen our understanding of science curricula that includes the nature of science, thereby providing initial information about the opportunities and challenges of teaching the nature of science to high school students as a separate school subject.

The commitment from the school would be to allow me to observe three SIE lessons over the course of five weeks, from the second week of March to the second week of April (around 60 minutes for each visit). I would audio-record students engaged in whole class discussion, video the teacher, observe and take notes, and photocopy some of the students' written work. The video camera will focus on the teacher and/or the board throughout and any incidental images of students will not be seen by anyone else or used for research without further explicit permission.

University of Oxford has strict ethical procedures on conducting ethical research with teachers and young people, consistent with current British Educational Research Association guidelines and Bioethics and Safety Act of Korea. Before beginning the research, I would inform parents and guardians about the research and offer the students, parents and guardians the opportunity to refuse to participate. Throughout the research, students, parents and guardians will be able to withdraw from participation at any time.

All participants, including students, teacher and the school, would be made anonymous in all research reports. The data collected would be kept strictly confidential, available only to my supervisors and myself, and not used other than specified without the further consent of all involved being obtained. All tapes and records would be destroyed at the end of the research period, and kept in locked conditions until then. I have enclosed copies of the leaflets for parents and students with this letter.

If you feel you would like to take part in the study, or need more information about what is involved, please contact me. Whether or not you feel it would be appropriate for your school to participate, I would be grateful if you would complete the pro-forma below, and return it to me in the stamped addressed envelope enclosed in this letter.

Thank you for your time and attention. I look forward to hearing from you.

Yours sincerely,

Professor Sibel Erduran
Tel: +44 (0)1865 274019 | Email: sibel.erduran@education.ox.ac.uk

Wonyong Park (DPhil Student)
Tel: +44 (0)7706 069166 | Email: wonyong.park@education.ox.ac.uk

Project Name: Physics Teachers' Experiences of Teaching a Science Curriculum Reform Explicitly Focused on Nature of Science

Ethics Approval Reference: [Insert]

Researcher: Wonyong Park
Department of Education, University of Oxford

[School name]
[School address]
[Head teacher name]

☐ We do not wish to participate in this project.

☐ We would like to find out more about this project.

☐ We would like to take part in this project.

Please return this form in the stamped addressed envelope enclosed with this letter.

Thank you for your help.

Appendix B1. Teacher Information Sheet

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

Principal investigator: Professor Sibel Erduran
sibel.erduran@education.ox.ac.uk
Primary researcher: Wonyong Park (DPhil student)
wonyong.park@education.ox.ac.uk

Physics Teachers' Experiences of Teaching a Science Curriculum Reform Explicitly Focused on Nature of Science

PARTICIPANT INFORMATION SHEET (TEACHER)

Ethics Approval Reference: [Insert]

1. ***What is the purpose of this research?***

   This study investigates the case of Korean science teachers' experiences in the wake of the recent national curriculum reform that highlights the nature of science (NOS) as an explicit learning expectation. The main interests of the study are to understand the obstacles that teachers experience while teaching scientific inquiry and experiment (SIE), the perceived meanings of NOS by teachers and the conditions that lead to successful teaching of SIE. Data will be collected from interviews of 6 physics teachers and classroom observations of three selected interviewees. The findings from this thesis will provide initial information about the conditions of and barriers to teaching NOS as a separate school subject. In doing so, implications will be drawn for curriculum developers and teacher educators to support the successful implementation of NOS-focused curriculum reforms.

2. ***Why have I been invited to take part?***

   You have been invited because you are a healthy and non-vulnerable teacher with a qualification in secondary physics teaching. Only physics teachers who teach SIE to Grade 10 students at the time of research (Spring 2019 semester) are invited.

3. ***Do I have to take part?***

   No. You can ask questions about the research before deciding whether or not to participate. If you do agree to participate, you may withdraw from the study at any time, without giving a reason, by advising the researchers of this decision.

4. ***What will happen to me if I take part in the research?***

   If you are happy to take part in the research, you will be asked to an interview of 45 minutes with the researcher. The interview will include several questions around your experiences and feelings about the new subject SIE. I will ask to arrange a good time and place (most likely your school) to conduct the interview.

   You may be also asked for classroom observations after the interview. In this case, you will be asked to allow me into three SIE lessons over the course of five weeks, from the second week of March to the second week of April (around 60 minutes each). I would audio-record students engaged in whole class discussion, video your teaching, observe and take notes, and photocopy some of the students' written work. The video camera will focus on the teacher and/or the board throughout and any incidental images of students will not be seen by anyone else or used for research without further explicit permission. If you agree to participate, I will talk to you to arrange my visits. The observation phase will include two additional interviews with you, each of which will take up to 30 minutes, asking questions about your teaching practices and views about SIE.

   If you are still happy to take part, you will then be asked to sign a consent form.

5. ***Are there any potential risks in taking part?***

In general, only minimum risks are expected for your participation. If you express any physical, emotional or psychological difficulties in the course of interviews or observations, the data collection will stop immediately until you are willing to resume. All collected information will be used only for research purposes and pseudonymised before published in any forms.

6. ***Are there any benefits in taking part?***

There will be no direct benefit to you from taking part in this research. However, the interviews can bring several benefits to you by, for example, (a) reflecting upon your own teaching practice and (b) sharing your curricular concerns to the researcher. Your participation will make significant contribution to science curriculum-makers and science education researchers by providing information about your implementation experience.

7. ***Expenses and payments***

There will be no payment for taking part in this study.

8. ***What happens to the data provided?***

All interviews, classroom observation videotapes and pupils' work data will be transcribed and immediately pseudonymised, and kept securely in the researcher's external hard drive. All identifiable personal data (audiotaped interviews, videotaped lessons and student works with names) will be destroyed immediately after transcription.

All collected data will be stored in (a) the researchers' personal external hard drive and (b) the researchers' business account in Oxford University Nexus365 cloud storage. Both will be secured with passwords and anti-virus software and encrypted. Personally identifiable data (audiotaped interviews, videotaped lessons) will be stored only until the data collection and transcription are finished, and then permanently destroyed.

The researcher and two supervisors will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

All research data will be stored for at least three years after publication or public release of the work of the research. We may retain and store your personal data for an additional period of time as necessary for the purposes of the study, and for further research.

9. ***Will the research be published?***

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be openly accessible to researchers. Part of the thesis could also be published in academic journals for dissemination of the research.

10. Who is organising and funding the research?

This study is conducted as part of the researcher's doctoral research. The research is not funded by external organisations or companies.

11. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: xxx).

12. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please speak to the researcher Wonyong Park (+44 7706 069166, +82 10 2828 1905, wonyong.park@education.ox.ac.uk) or his supervisor Professor Sibel Erduran (+44 1865 274182, sibel.erduran@education.ox.ac.uk), who will do their best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how they intend to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

Dr Liam Gearon, Chair, Departmental Research Ethics Committee
Oxford University Department of Education
Email: research.office@education.ox.ac.uk
Address: Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY

13. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study.

The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest.

Further information about your rights with respect to your personal data is available from http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/.

14. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Professor Sibel Erduran
Tel: +44 (0)1865 274019 | Email: sibel.erduran@education.ox.ac.uk

Wonyong Park (DPhil Student)
Tel: +44 (0)7706 069166 | Email: wonyong.park@education.ox.ac.uk

Appendix B2. Consent Form (Teacher)

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

Principal investigator: Professor Sibel Erduran
sibel.erduran@education.ox.ac.uk
Primary researcher: Wonyong Park (DPhil student)
wonyong.park@education.ox.ac.uk

15 Norham Gardens, Oxford OX2 6PY
Director: Professor Jo-Anne Baird
Tel: +44 (0)1865 274024 Fax: +44 (0)1865 274027

PARTICIPANT CONSENT FORM (TEACHERS)

Ethics Approval Reference: [Insert]

Physics Teachers' Experiences of Teaching a Science Curriculum Reform Explicitly Focused on Nature of Science

Purpose of Study: Understanding the obstacles that teacher experience while teaching SIE, the perceived meanings of nature of science by teachers and the conditions that lead to successful implementation of SIE.

		Please initial each box
1	I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without any adverse consequences or academic penalty.	
3	I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this study. I give permission for these individuals to access my data.	
4	I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.	
5	I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	
6	I understand how this research will be written up and published.	
7	I understand how to raise a concern or make a complaint.	
8	I consent to being audio recorded	
9	I consent to being video recorded	
10	I understand how audio recordings / videos will be used in research outputs	
11	I agree to take part in the study	

______________	dd / mm / yyyy	______________
Name of Participant	Date	Signature
______________	dd / mm / yyyy	______________
Name of person taking consent	Date	Signature

Appendix B Interview protocols

1. Year 1 Interview (60 mins)

Background questions

- Could you tell me about your career in teaching briefly?
- Have you taught SIE before?
- To how many classes are you teaching SIE this semester?

Views of SIE

- What did you know about NOS and SIE before teaching it? Have you received any training about them?
- What was your first impression about the NOS part in SIE?
- Did it change as you teach it this semester?
- Do you think SIE would be helpful to students? If so, how could it be helpful?
- How do you feel about assessing SIE in terms of three levels?
- How would you compare teaching SIE and teaching chemistry [physics, earth science]?
- Do you think your subject specialism in chemistry [physics, earth science] influences how you teach SIE? If so, how?

Questions about observed lesson (examples)

(After the ‘make your own periodic table’ lesson)

- How did you come to use this activity in your lesson?
- Is there a difference between teaching about periodic table in a chemistry lesson and an SIE lesson?

(After the ‘paradigm shift’ lesson)

- Why did you choose to do student presentations for this lesson?
- You told students that the Q&A after presentations were important. Was there any particular reason that you said this?

2. Year 2 Interview (initial, 30 mins)

Usual assessment practices

- How do you usually check for student understanding?
- Do you have any preferred assessment methods?
- As a teacher, what kind of things do you consider when planning assessment?

- How would you compare assessing SIE with assessing chemistry [physics, earth science]?
- It's said that assessment should be fair. How do you usually ensure your assessment is fair?
- How do you deal with the fairness of assessment among other teachers who teach the same year group?

NOS

- How would you define NOS?
- Do you think it is important to teach NOS to students? Why do you think so?

3. Year 2 Interview (final, 30 mins)

Views on assessing NOS

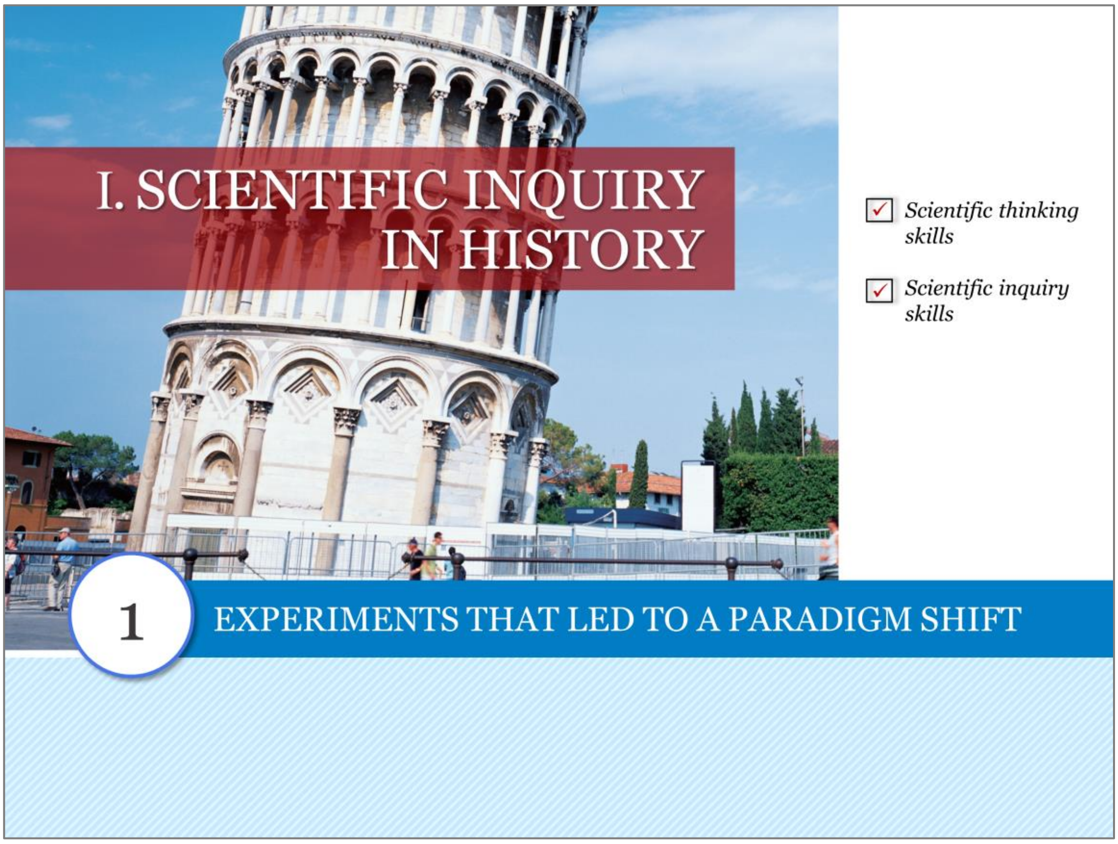
- In your opinion, what should be assessed about NOS?
- In your opinion, what kind of assessment methods would be suitable for NOS?

Cloud activity

- How did you feel about today's activity?
- What could the students have learnt from today's activity?
- How would you mark each student's worksheet from today?
- I noticed that one student wrote ____ for Question number # on the worksheet. How would you mark this answer?

Appendix C Example lesson and assessment materials

1. Paradigm shift lesson slides (selected, translated from Korean)



1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

How did the evolution of science that led to a paradigm shift come about?

Science progress may occur gradually,
but It may also be dramatic enough to transform the 'paradigm'.

Paradigm

A system of thoughts that fundamentally defines the views or thoughts of people living in a certain time

1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

Understanding Aristotle's ideas

1. Which one of heavy or light stones will fall first?

My thoughts on Aristotle's opinion

- ✓ The falling of a heavy object more quickly is consistent with what we know from experience. Actually, I tried to drop the eraser and a piece of tissue at the same time, and the eraser fell faster.
- ✓ I thought it was because of the different size, so I put several pieces of tissue together and made them into a size similar to that of an eraser.

1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

Understanding Aristotle's ideas

2. Can the object continue the linear motion at constant speed?

My thoughts on Aristotle's opinion

- ✓ When I think about playing soccer, I have noticed that the ball doesn't roll forever, and the speed of the rolling ball is getting slower.
- ✓ However, isn't it because the playground has frictional forces due to sand and stones? I wonder what would happen if there was no friction.

1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

Galileo's thought experiment

1. Fill in the blanks following Galileo's thought experiment process, which resulted in a new paradigm for the free fall movement.

Argument 1: A heavy stone falls faster than a light stone. → When two stones are joined, it becomes heavier than the heavy stone. → Therefore, the joined stones will fall _____ than the _____ stone.

Argument 2: A heavy stone falls faster than a light stone. → The light stone falls slower and interrupts the motion of the heavy stone. → Therefore, the joined stones will fall _____ than the _____ stone.

1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

Summing up the results

2. Explain what changes in thought have occurred in explaining natural phenomena through Aristotle, Galilei, and Newton.

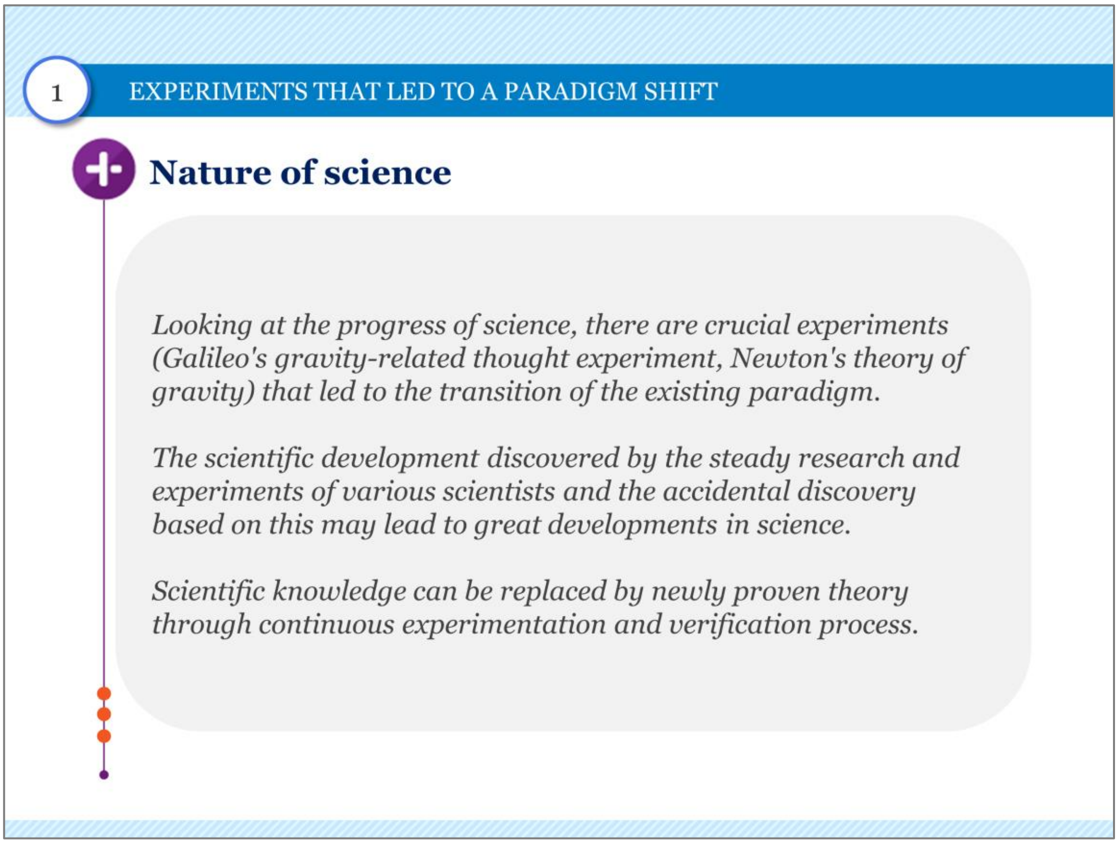
- ✓ Aristotle explained natural phenomena as his own logic, but used the hypothesis that could not be tested for the cause of the phenomena.
- ✓ Galileo's thought experiment revealed Aristotle's disagreement and tried to explain it through experimentation. Newton completed the paradigm shift caused by Galileo through the laws of gravity and motion to the theory of mechanics.
- ✓ Newton's theory of mechanics is significant in that it is characterized by rigorous narrative through mathematics, and through this, it is possible to systematically explain the motion of all objects.

1 EXPERIMENTS THAT LED TO A PARADIGM SHIFT

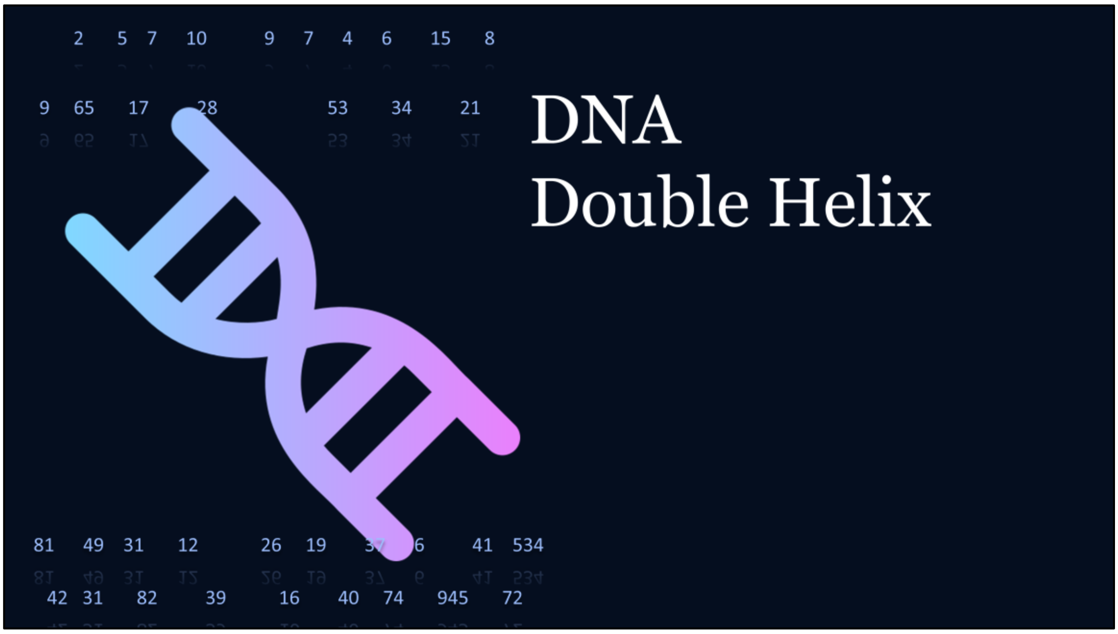
Summing up the results

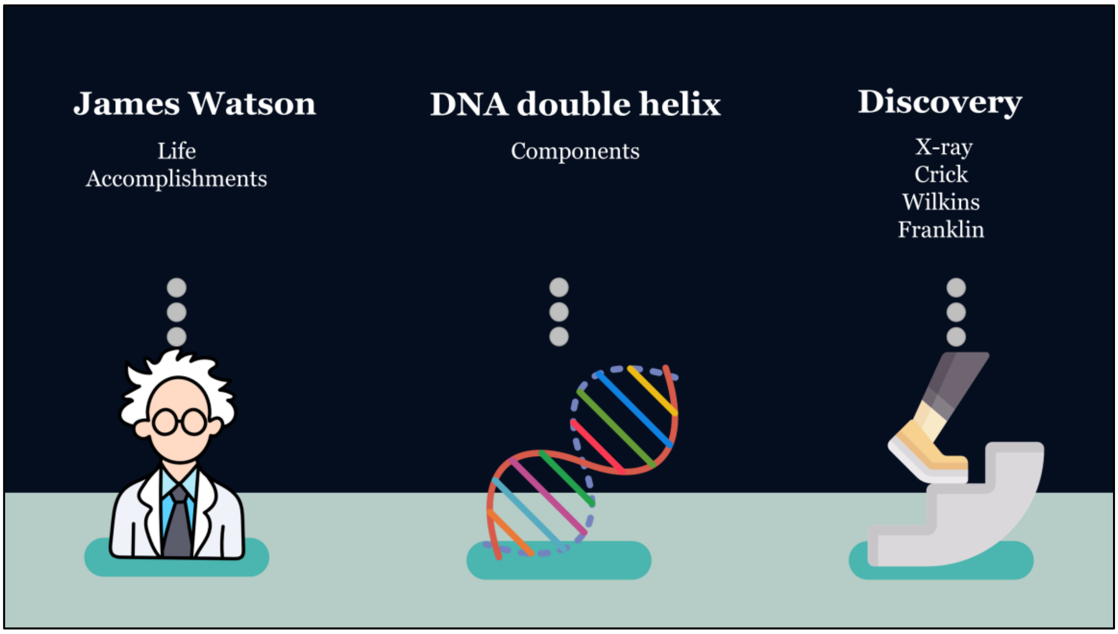
3. Let's investigate what are the paradigm shifts in other areas of science.

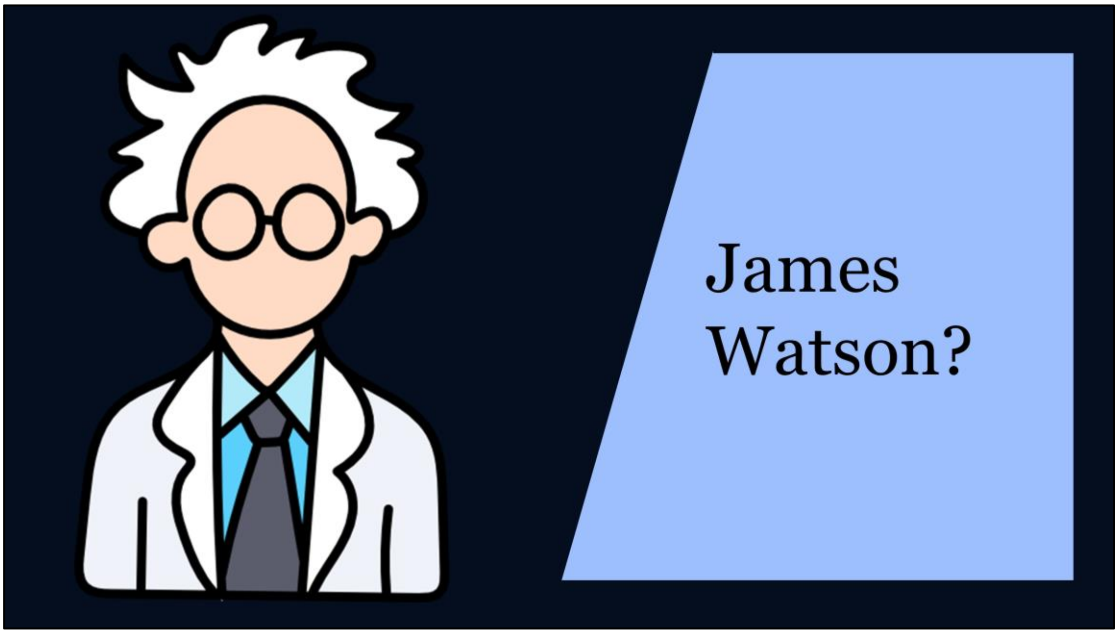
- ✓ The paradigm shift is the process by which the theory explaining the motions of planets and celestial bodies has shifted from the geocentric theory to the heliocentric theory.

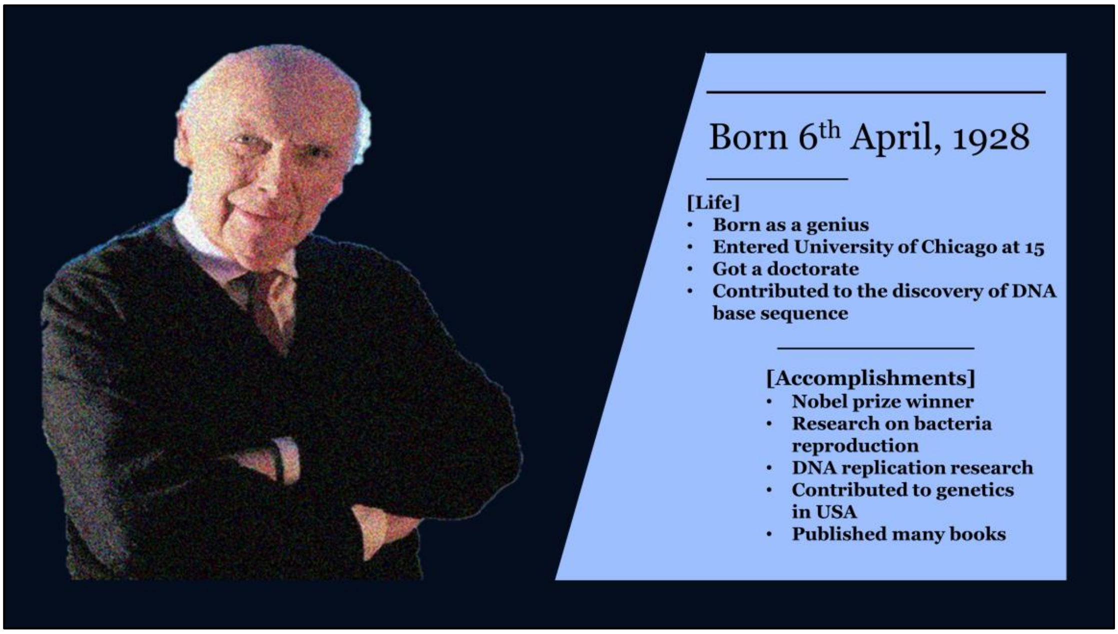

1
EXPERIMENTS THAT LED TO A PARADIGM SHIFT
Nature of science
Looking at the progress of science, there are crucial experiments (Galileo's gravity-related thought experiment, Newton's theory of gravity) that led to the transition of the existing paradigm.
The scientific development discovered by the steady research and experiments of various scientists and the accidental discovery based on this may lead to great developments in science.
Scientific knowledge can be replaced by newly proven theory through continuous experimentation and verification process.

2. Student presentation on the discovery of DNA double helix (translated from Korean)

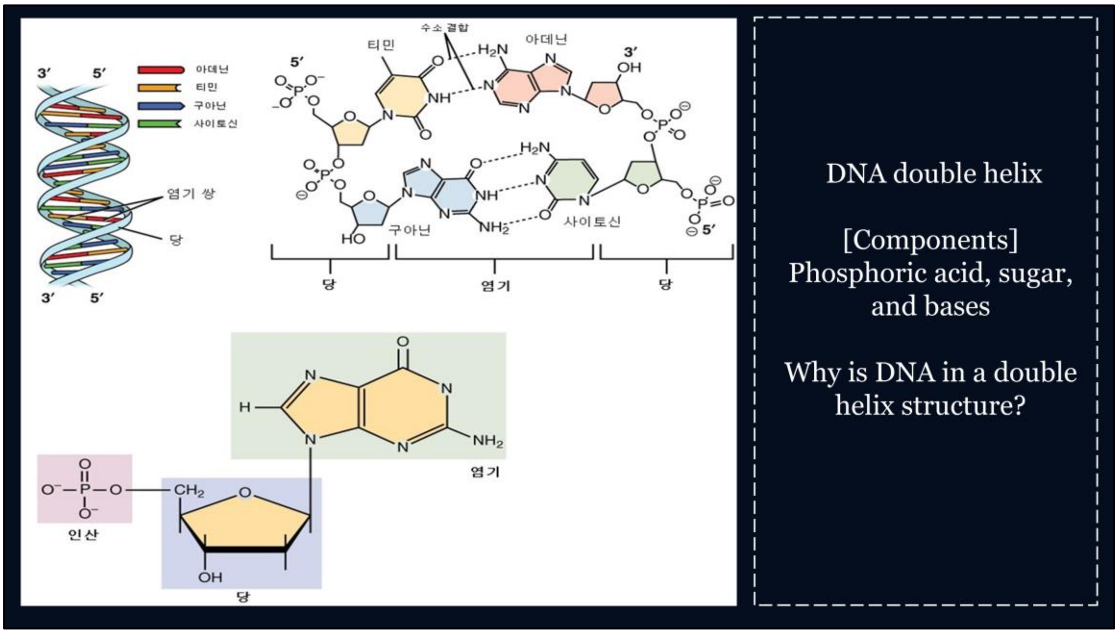




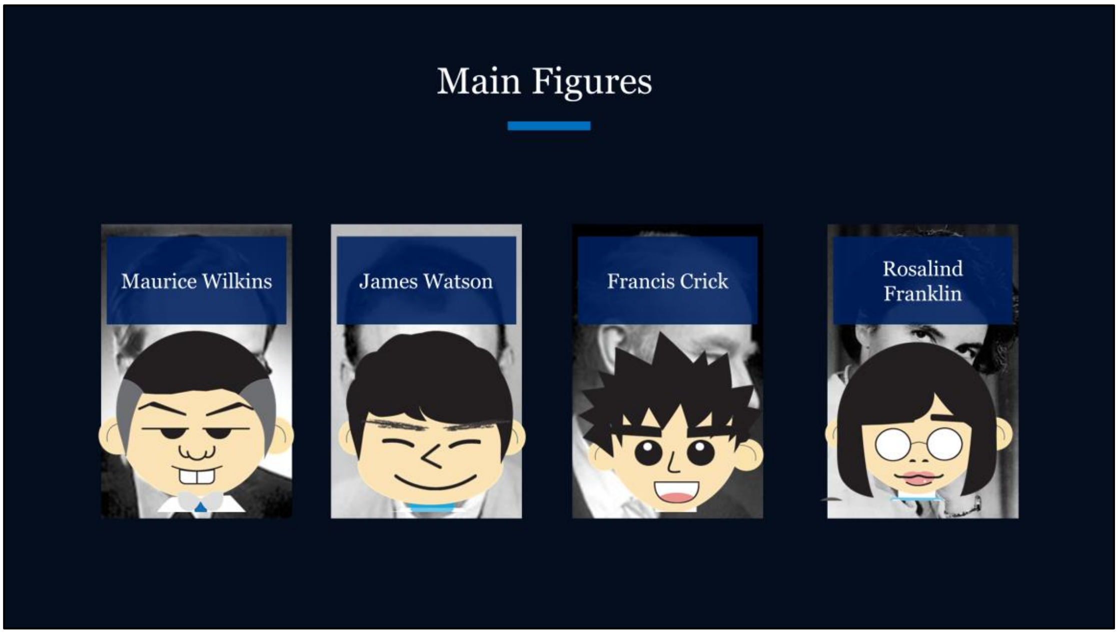

James
Watson?

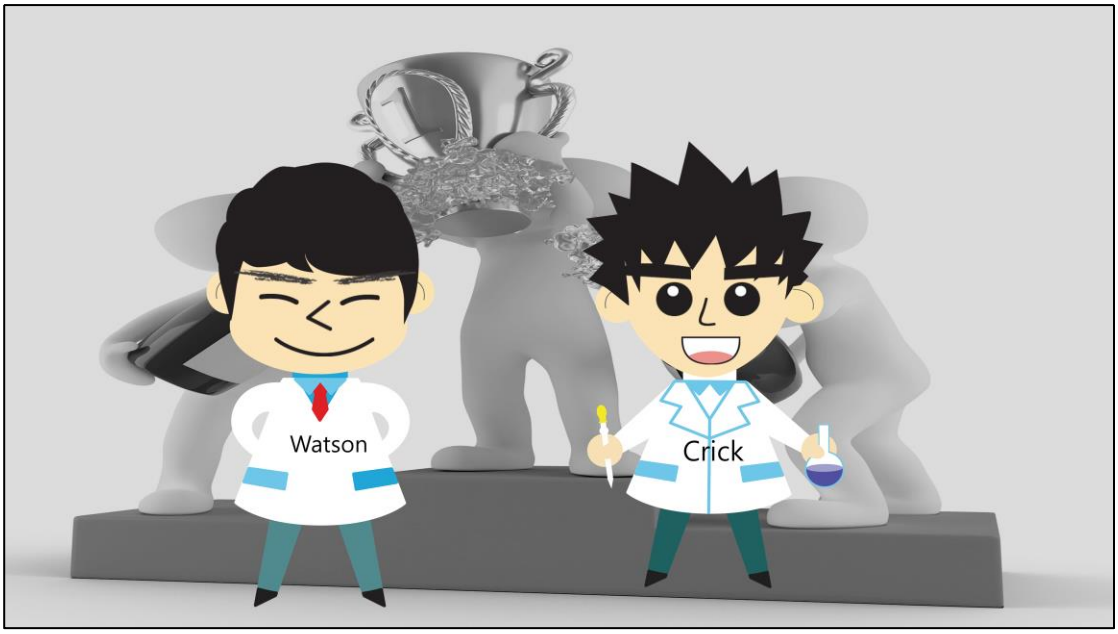

Born 6th April, 1928
[Life]
• Born as a genius
• Entered University of Chicago at 15
• Got a doctorate
• Contributed to the discovery of DNA base sequence
[Accomplishments]
• Nobel prize winner
• Research on bacteria reproduction
• DNA replication research
• Contributed to genetics in USA
• Published many books

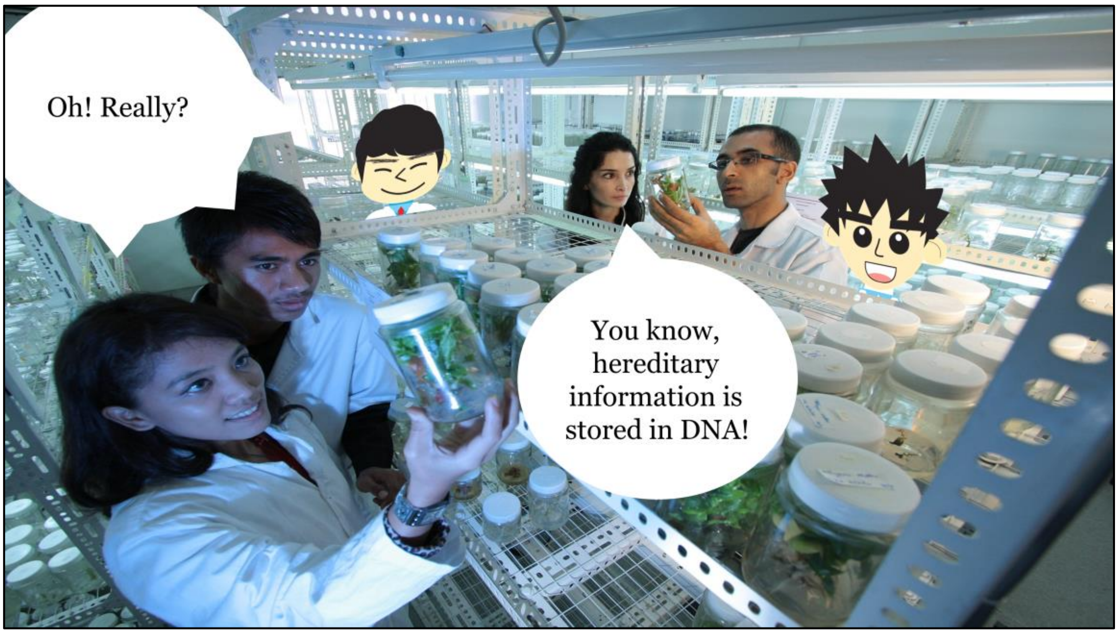

What is
DNA
double
helix?

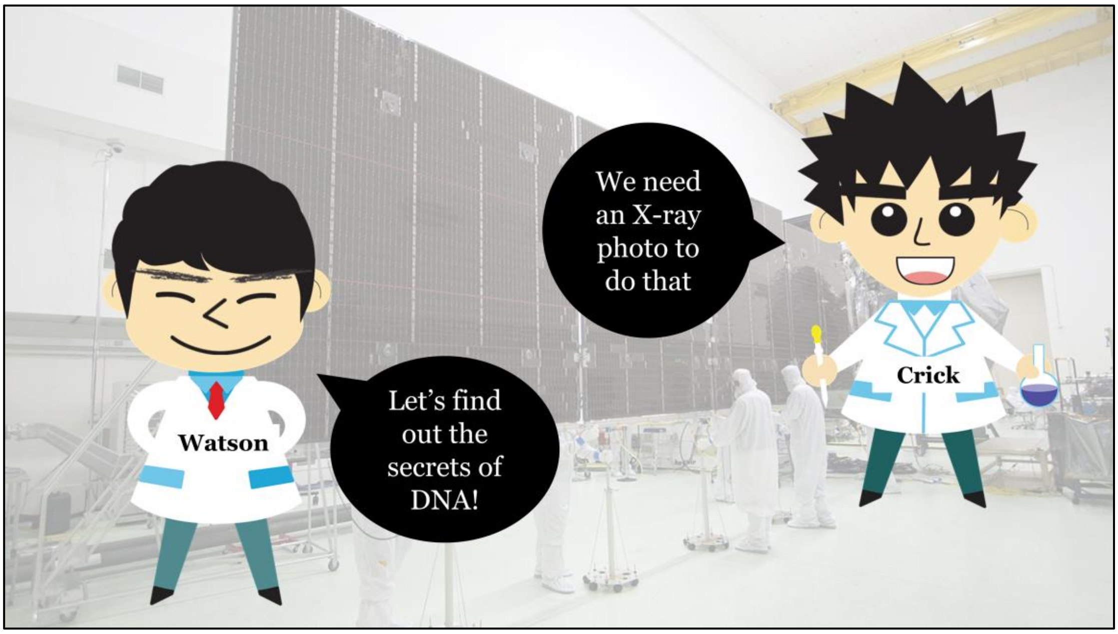

3′
5′
아데닌
티민
구아닌
사이토신
염기 쌍
당
수소 결합
인산
염기
DNA double helix
[Components]
Phosphoric acid, sugar,
and bases
Why is DNA in a double
helix structure?

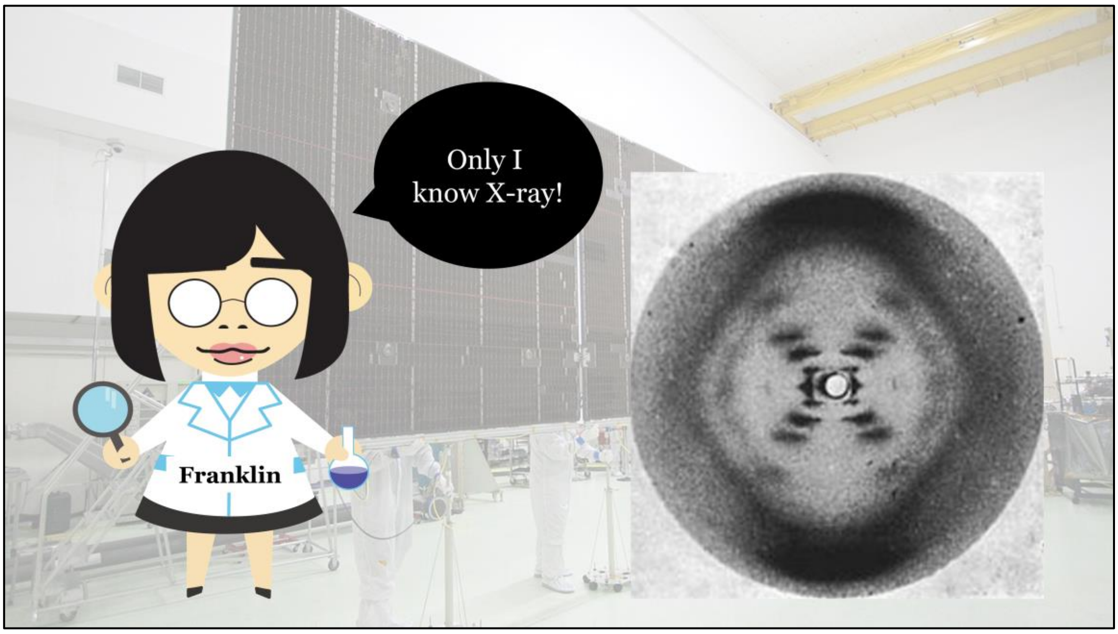

Discovery?

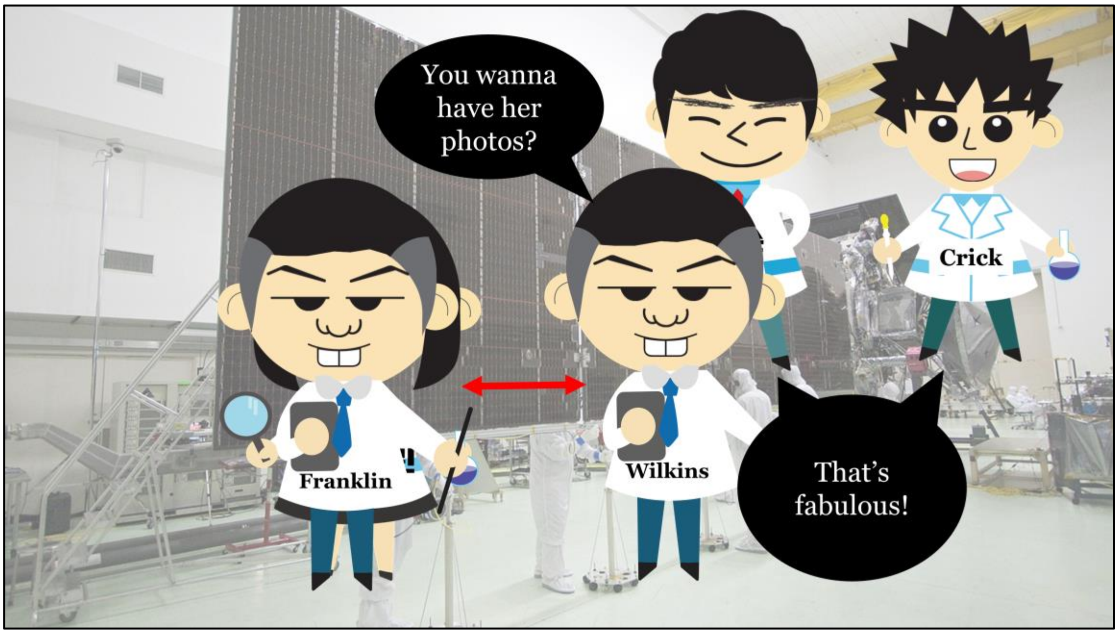

Main Figures
Maurice Wilkins
James Watson
Francis Crick
Rosalind Franklin

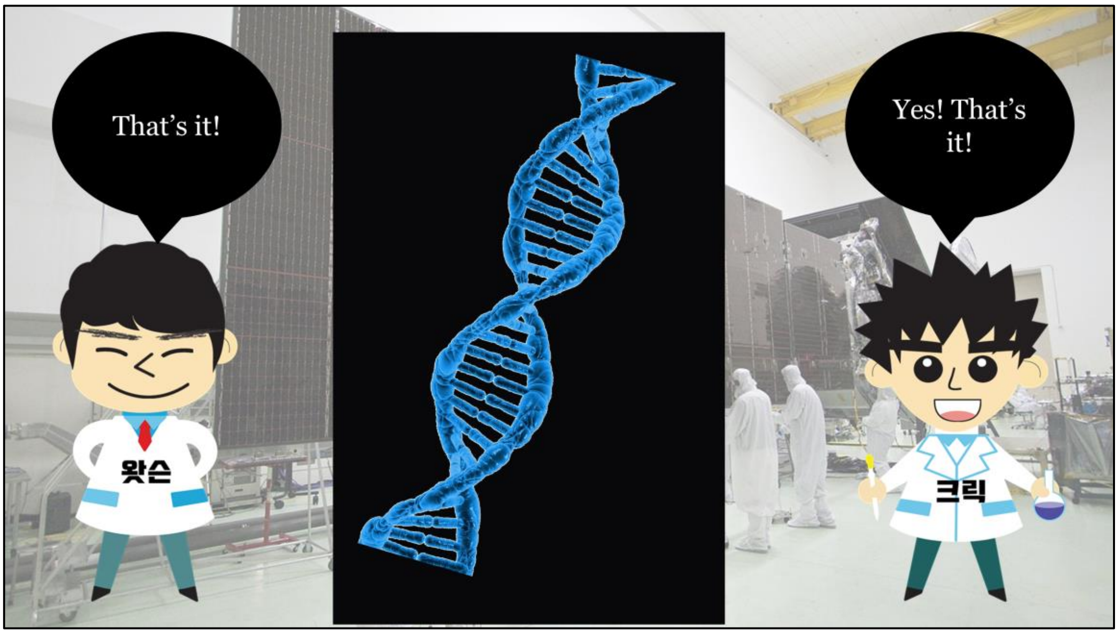

Watson
Crick

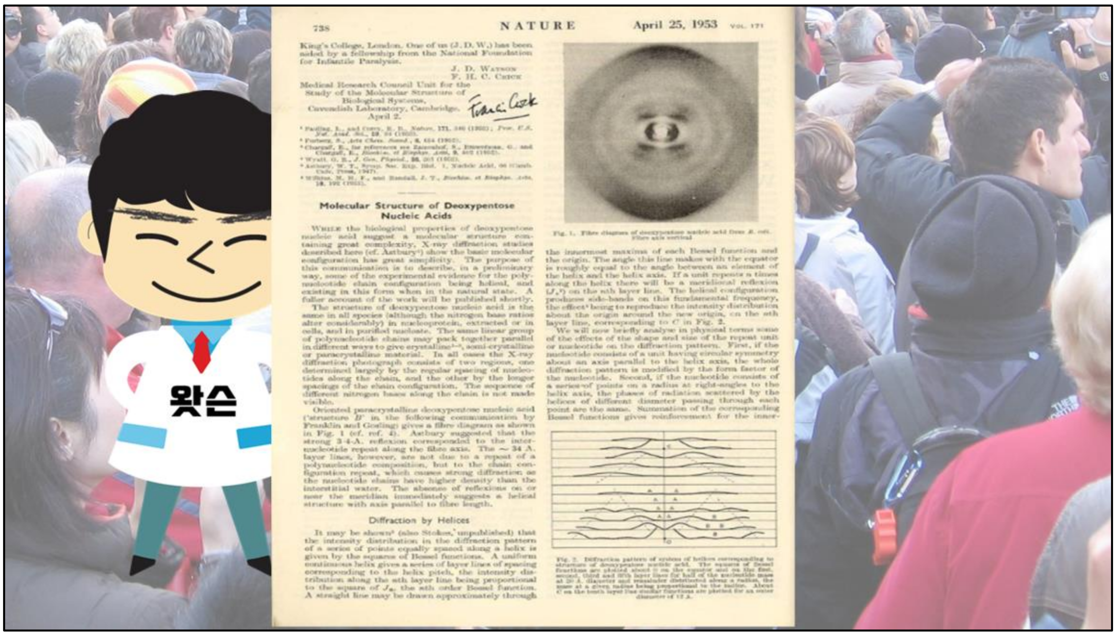

Oh! Really?
You know, hereditary information is stored in DNA!

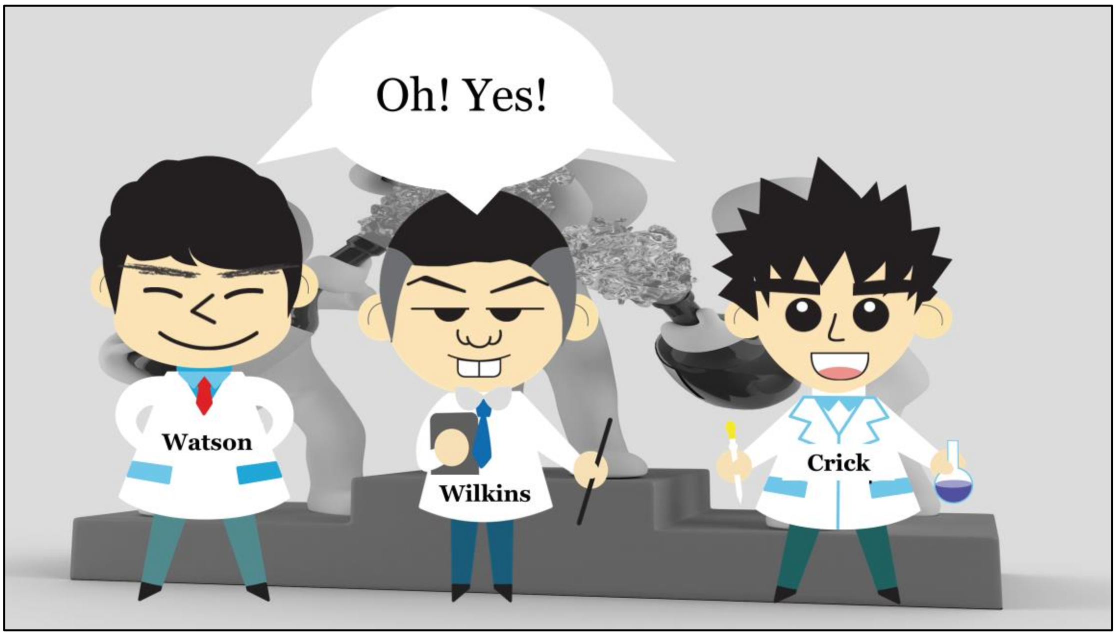

Watson
Let's find out the secrets of DNA!
Crick
We need an X-ray photo to do that

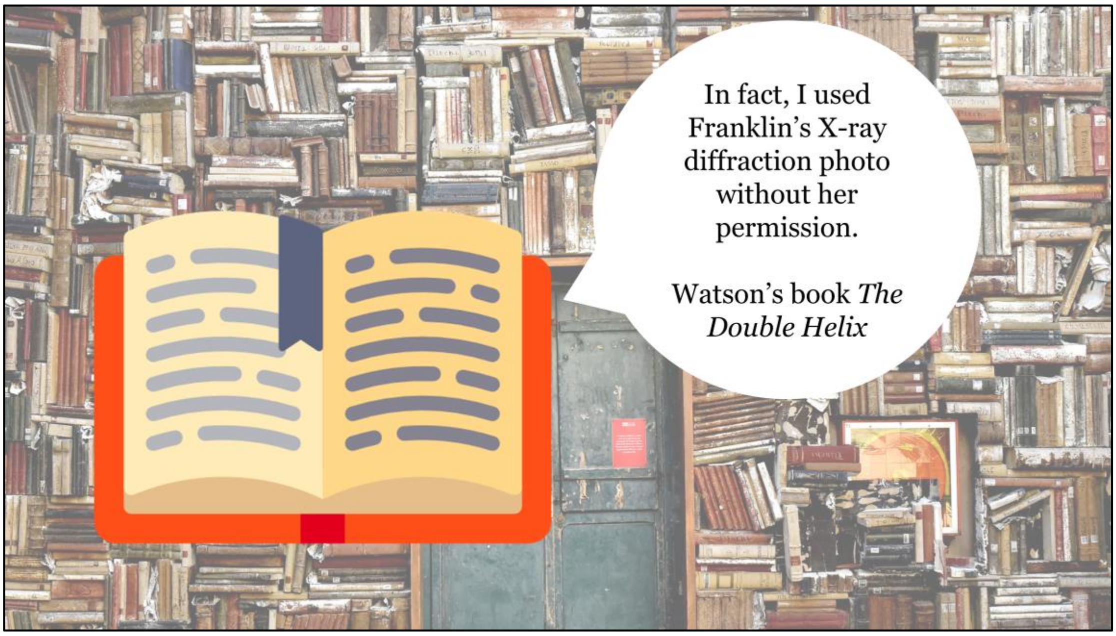

Franklin
Only I know X-ray!


You wanna have her photos?
Crick
Franklin
Wilkins
That's fabulous!

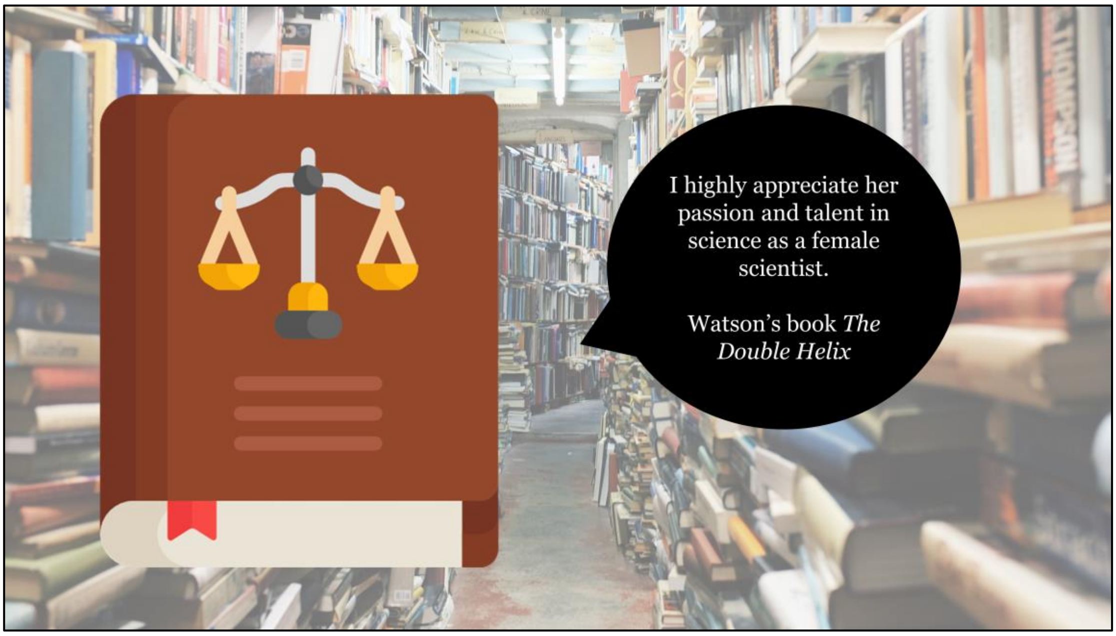

That's it!
Yes! That's it!
왓슨
크릭

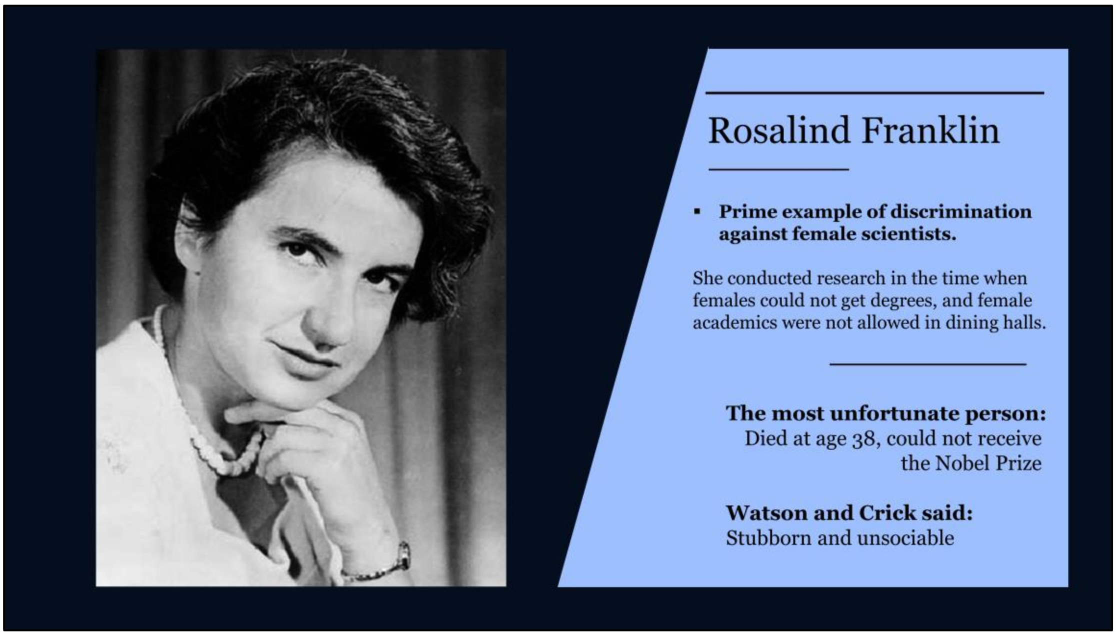

738
NATURE
April 25, 1953
Molecular Structure of Deoxypentose Nucleic Acids
Diffraction by Helices
왓슨


Oh! Yes!
Watson
Wilkins
Crick

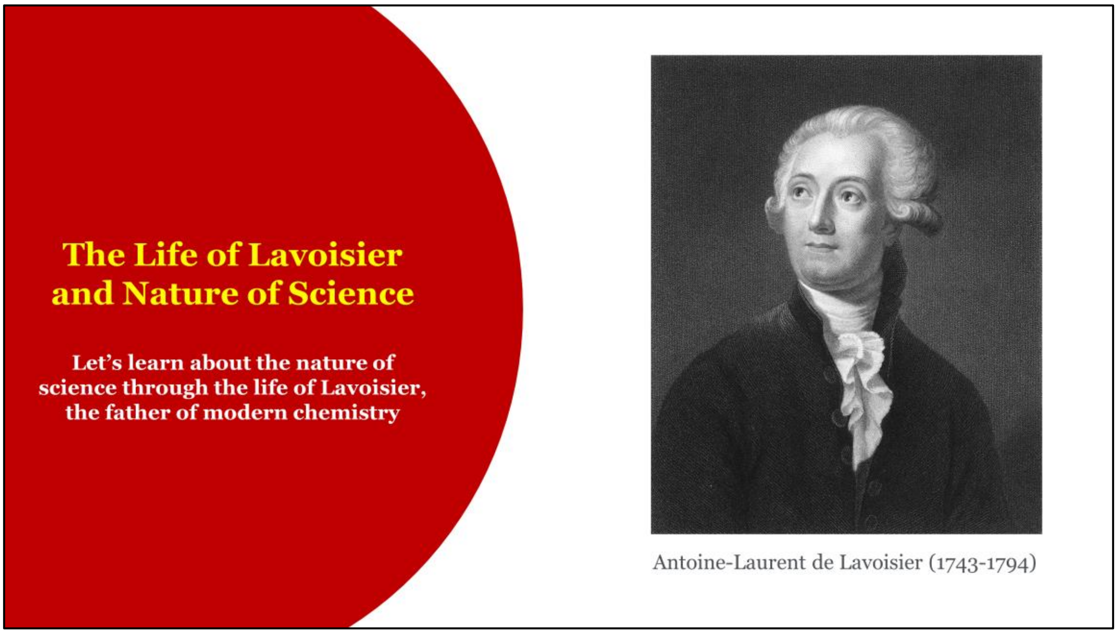

In fact, I used Franklin's X-ray diffraction photo without her permission.
Watson's book *The Double Helix*

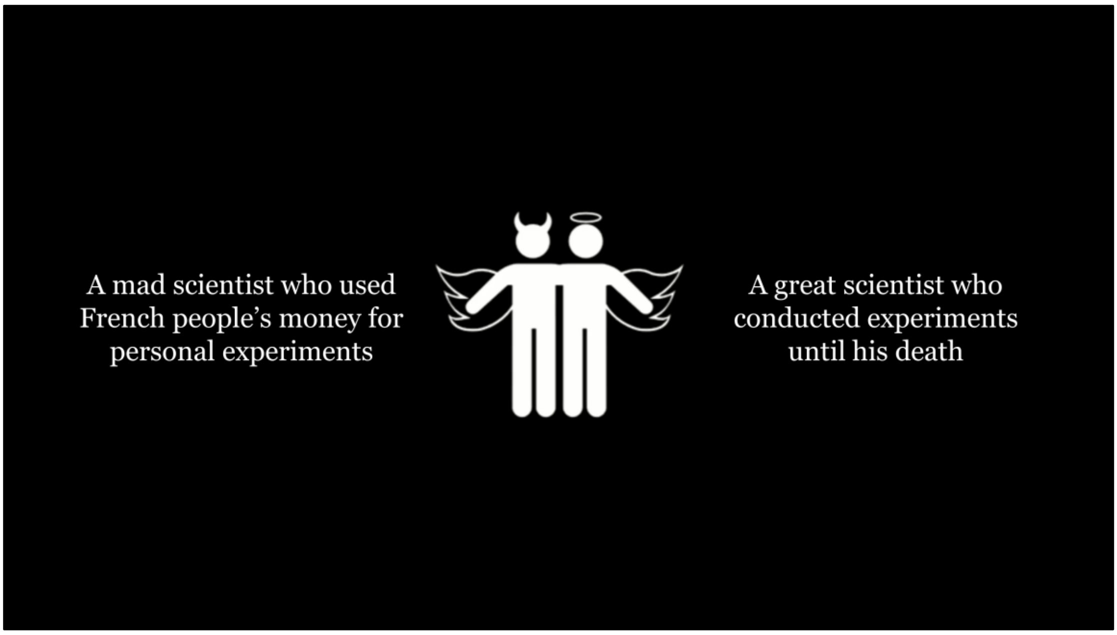

Oh what?
What a @#%!
How can they steal the research!
They stole Franklin's achievement!

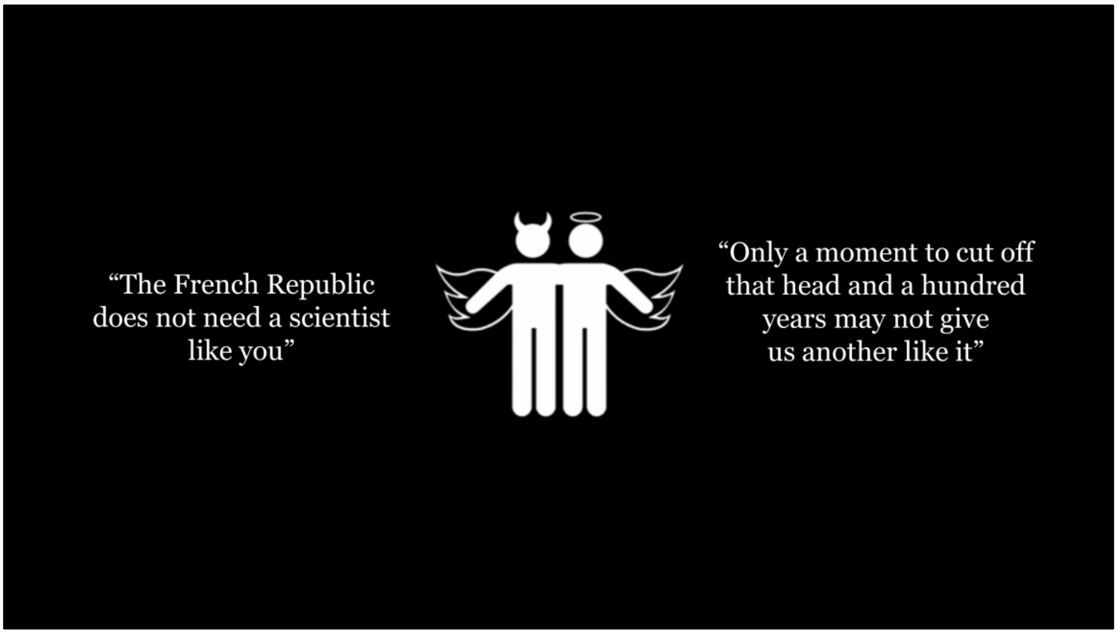

I highly appreciate her passion and talent in science as a female scientist.
Watson's book *The Double Helix*


Rosalind Franklin
Prime example of discrimination against female scientists.
She conducted research in the time when females could not get degrees, and female academics were not allowed in dining halls.
The most unfortunate person: Died at age 38, could not receive the Nobel Prize
Watson and Crick said: Stubborn and unsociable

3. Lavoisier's life and nature of science lesson slides (selected, translated from Korean)







Wealthy family background

Innate brilliance

Inherited a large amount of money from grandmother at age 23

Elected a member of the French Academy of Sciences at age 25

Known in the field of science at age 27

Appointed as a tax collector

Lavoisier's life was really promising.

Lavoisier gained wealth and fame, and then conducted luxurious experiments such as burning a diamond with magnifying glass.

His experiments were done with the money he collected from the French people.

Nature of Science in the Video

(Lavoisier) Proved that Stahl's phlogiston theory was wrong

He heated water to divide it into hydrogen and oxygen
and discovered that water consists of hydrogen and oxygen

When he put a guinea pig in a isolated place,
he found that the oxygen disappeared and carbon dioxide appeared,
which proved that respiration requires oxygen.

…

➢ **Science explains and predicts natural phenomena**

Nature of Science in the Video

In April 1770,

Lavoisier proposed the conservation of mass, which was against the popular idea in his time that water can be turned into soil by boiling.

“When a chemical reaction occurs, the total mass of reacting matters equals to the total mass after the reaction.

It is impossible that matter is created or destroyed through chemical reactions.”

- **Scientific knowledge can change due to the discovery of a new theory**

Nature of Science in the Video

(Lavoisier) Proved that Stahl’s phlogiston theory was wrong

He heated water to divide it into hydrogen and oxygen
and discovered that water consists of hydrogen and oxygen

When he put a guinea pig in a isolated place,
he found that the oxygen disappeared and carbon dioxide appeared,
which proved that respiration requires oxygen.

…

- **Scientific knowledge is based on both observation and inference**

Nature of Science in the Video

In 1794, the great scientist was sent to the guillotine.

- **Science is influenced by society and culture**

Lavoisier's life and nature of science

- **Establishing a new direction for chemistry**

Lavoisier established a new direction for chemistry. People like Galileo in mechanics, Newton in physics, Darwin in biology gave a new order in science. Lavoisier was like them in the field of chemistry.

4. Cloud observation and classification activity sheet (translated from Korean)

Observing and Classifying Clouds

Class (   )   Student No. (   )   Name (   )

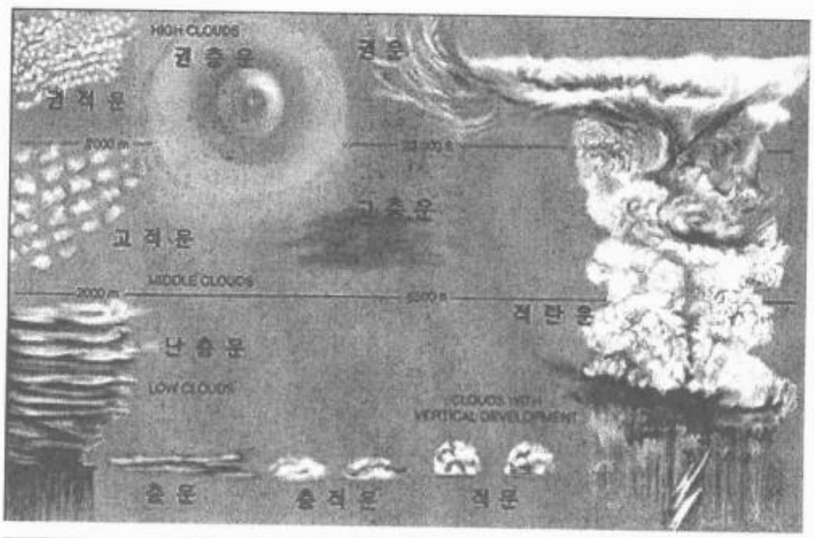
INTRODUCTION

There are various forms of clouds in the sky, but they can be classified in terms of their position and shape. In this activity, we will observe clouds and classify them.

BACKGROUND KNOWLEDGE

The scientific classification of clouds was first suggested by L. Howards in 1803, who suggested that clouds are classified into three types—*cirrus*, *cumulus* and *stratus*. Based on this, many detailed observations were made and in 1956 at the International Meteorological Conference, the first *International Cloud Atlas* was published and have been revised several times.

According to their altitudes, clouds are divided into four kinds: high clouds, middle clouds, low clouds and clouds with vertical development. They in turn are further classified based on their shapes into the basic 10 cloud types.



층별	구름 높이(km)	학명	명칭
상층운	6-13	cirrus(Ci)	권운
		cirrocumulus(Cc)	권적운
		cirrostratus(Cs)	권층운
중층운	2-6	altocumulus(Ac)	고적운
		altostratus(As)	고층운
		nimbostratus(Ns)	난층운
하층운	지면-2	stratocumulus(Sc)	층적운
		stratus(St)	층운
수직운	지면-13	cumulus(Cu)	적운
		cumulonimbus(Cb)	적란운

s

METHODS

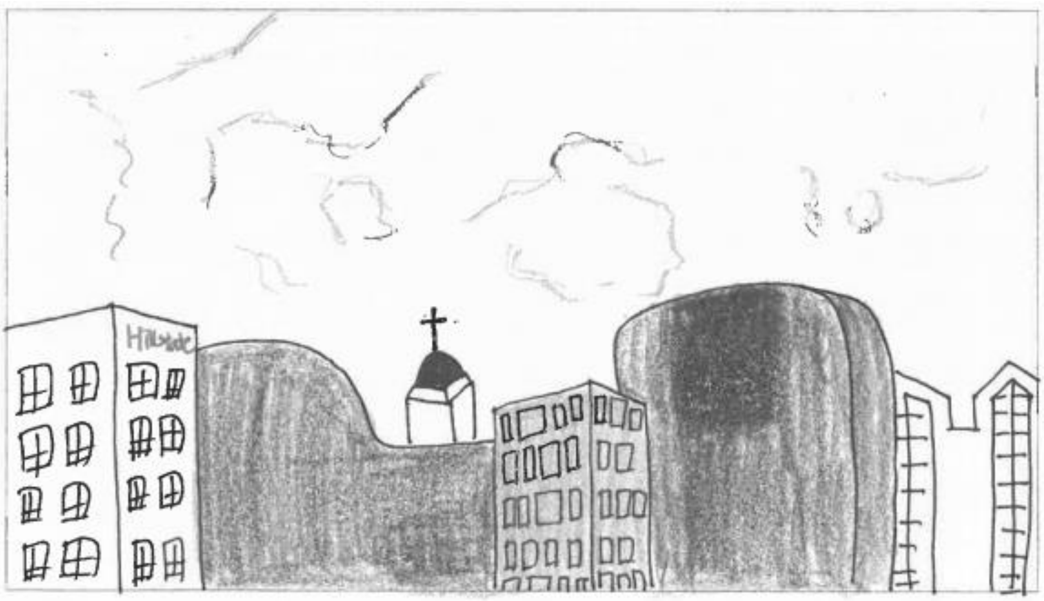
1) Sketching clouds

a. With naked eyes, observe and sketch the shape of clouds. (Note: Draw nearby tall buildings, mountains etc. so you know the altitude)
b. With reference to the nearby building or mountain, determine if the clouds are high, middle or low.
c. Check the cardinal direction and write down where the clouds are moving towards.
d. Write down the date, place and weather of your observation.
e. Draw the rough appearance of clouds in every direction.

2) Classifying clouds

a. Classify the clouds as high, middle or low.
b. Based on the ten basic cloud types, determine which cloud they are.

RESULTS (example student answer)



Altitude: *Low*
Type: *Stratus*

Why did you think so?
The weather was cloudy, there was so many clouds that no sunlight is seen, and clouds are low. I can hardly the sky and it's filled with thick clouds.

Date: ______________________
Place: ______________________
Weather: *Cloudy, Drizzly*
Cardinal direction: *East*
Cloud movement: *South to North*

DISCUSSION

Can you think of an example where the result of observation led to the discovery of classification? Explain your answer.

5. Views of Nature of Science Questionnaire Form C (adapted from Abd-El-Khalick, 1998)

Please answer each of the following questions. Include relevant examples whenever possible. There are no "right" or "wrong" answers to the following questions.

(1) What, in your view, is science? What makes science (or a scientific discipline such as physics, biology, etc.) different from other disciplines of inquiry (e.g., religion, philosophy)?

(2) What is an experiment?

(3) Does the development of scientific knowledge require experiments? • If yes, explain why. Give an example to defend your position. • If no, explain why. Give an example to defend your position.

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

(5) Is there a difference between a scientific theory and a scientific law? Illustrate your answer with an example.

(6) After scientists have developed a scientific theory (e.g., atomic theory, evolution theory), does the theory ever change?

> If you believe that scientific theories do not change, explain why. Defend your answer with examples.
> If you believe that scientific theories do change: (a) Explain why theories change? (b) Explain why we bother to learn scientific theories. Defend your answer with examples.

(7) Scientists perform experiments/investigations when trying to find answers to the questions they put forth. Do scientists use their creativity and imagination during their investigations?

(8) It is believed that about 65 million years ago the dinosaurs became extinct. Of the hypotheses formulated by scientists to explain the extinction, two enjoy wide support. The first, formulated by one group of scientists, suggests that a huge meteorite hit the earth 65 million years ago and led to a series of events that caused the extinction. The second hypothesis, formulated by another group of scientists, suggests that massive and violent volcanic eruptions were

responsible for the extinction. How are these different conclusions possible if scientists in both groups have access to and use the same set of data to derive their conclusions?

(9) Some claim that science is infused with social and cultural values. That is, science reflects the social and political values, philosophical assumptions, and intellectual norms of the culture in which it is practiced. Others claim that science is universal. That is, science transcends national and cultural boundaries and is not affected by social, political, and philosophical values, and intellectual norms of the culture in which it is practiced.

> If you believe that science reflects social and cultural values, explain why and how. Defend your answer with examples.
> If you believe that science is universal, explain why and how. Defend your answer with examples