Assessing Scientific Literacy as Participation in Civic Practices

Affordances and constraints for developing a practice for authentic classroom assessment of argumentation, source critique and decision-making

Jens Anker-Hansen
Assessing Scientific Literacy as Participation in Civic Practices

Affordances and constraints for developing a practice for authentic classroom assessment of argumentation, source critique and decision-making

Jens Anker-Hansen
Abstract

This thesis takes a departure from a view of scientific literacy as situated in participation in civic practices. From such a view, it becomes problematic to assess scientific literacy through decontextualised test items only dealing with single aspects of participation in contexts concerned with science. Due to the complexity of transferring knowledge, it is problematic to assume that people who can explain scientific theories will automatically apply those theories in life or that knowledge will influence those people’s behaviour. A common way to more fully include the complexity of using science in different practices is to focus participation around issues and study how students use multiple sources to reflect critically and ethically on that issue. However, participation is situated in practices and thus becomes something specific within those practices. For instance, shopping for groceries for the family goes beyond reflecting critically and ethically on health and environment since it involves considering the family economy and the personal tastes of the family members. I have consequently chosen to focus my studies on how to assess scientific literacy as participation in civic practices. The thesis describes a praxis development research study where I, in cooperation with teachers, have designed interventions of assessments in lower secondary science classrooms. In the research study I use the theory of Community of Practice and Expansive Learning to study affordances and constraints for assessing communication, source critique and decision-making in the science classroom. The affordances and constraints for students’ participation in assessments are studied through using a socio-political debate as an assessment tool. The affordances and constraints for communicating assessment are studied through peer assessments of experimental design. The affordances and constraints for teachers to expand their assessment repertoire are studied through assessment moderation meetings. Finally, the affordances and constraints for designing authentic assessments of scientific literacy are studied through a review of different research studies’ use of authenticity in science education. The studies show that tensions emerge between purposes of practices outside the classroom and practices inside the classroom that students negotiated when participating in the assessments. Discussion groups were influential on students’ decisions on how to use feedback. Feedback that was not used to amend the designs was still used to discuss what should count as quality of
experiments. Teachers used the moderation meetings to refine their assessments and teaching. However, conflicting views of scientific literacy as either propositional or procedural knowledge were challenging to overcome. Different publications in science education research emphasised personal or cultural aspects of authenticity. The different uses of authenticity have implications for authentic assessments, regarding the affordances and constraints for how to reify quality from external practices and through students’ engagement in practices. The results of the studies point to gains of focussing the assessment on how students negotiate participation in different civic practices. However, this approach to assessment puts different demands on assessment design than assessments in which students’ participation is compared with predefined ideals for performance.

Keywords: scientific literacy, assessment, authentic, communities of practice, expansive learning, argumentation, peer assessment, moderation meetings
Acknowledgments

Above all I want to thank my supervisor, Maria Andrée. Not only are you incredibly wise when it comes to understanding the practices of science education research, but also amazingly supportive and encouraging when the challenges of the research seemed unsurpassable. I further commend you for your patience when I strayed from my research focus or joked too much. I am infinitely grateful for having you as a supervisor during my years at Stockholm University. I hope many future PhD students will be fortunate to learn the craft from you. Additionally, I would like to express my gratitude to P.O. Wickman. With your immeasurable experience, you have kept me on track, navigated me through conflicting feedback from reviewers, and encouraged me to continue when the critique was overwhelming.

A very special thanks to ‘William’, ‘Sofie’ Ms. ‘W’ and all the teachers and students in ‘Apple’, ‘Birch’ and ‘Central’ school who participated in the project. I hope it was just as fun for you as it was for me working together on the research project. You are all great and I wish you good luck with science in future participation in civic practices. I am sorry I cannot credit you for your contribution for a greater understanding assessment of scientific literacy.

Furthermore, I would like to express my appreciation to my fellow PhD students. During my first year, it was valuable to have experienced PhD students like Jakob and Auli to ask for help when engaging in a practice so different from school teaching. Above all I want to thank my roommates. There are many times where an obstacle was overcome in a few minutes simply by discussing it with you. Thanks, Per, for all the pea soup (or perhaps PEA-soup) dinners. It is always nice to visit your family. I could easily have talked all day with you, Zeynep, had we not been bogged down with hard work. Jonna and Malin have proved a great addition to our ‘Dr-And-Team’. Though you have resided in other rooms, I still consider myself lucky to have had such fellows among PhD students as Cecilia, Camilla, Ilana and Veronica. Though I will not be hosting any more PhD student dinners, I hope we can remain in contact with one another. Not least, I would like to thank the mathematics education PhD students for all interesting cross-disciplinary discussions.

I am very grateful for all the helpful advice given to me by the senior researchers: Iann, Jesus, Karim, B.O., Britt and C.J. You are always willing to help with the research. Moreover, I am grateful for all advice received on my
seminar readings, especially Lotta Lager Nyqvist (10%), Anders Jönsson and Margareta (50%) as well as Anders Jakobsson and Astrid (90%).

Thank you very much to all the instructors at the department, Åsa, Carolina, Sofie, Matti, Lotta and Mats, who allowed me to teach about assessment in your courses. Discussing assessment of scientific literacy with pre-service teacher students has been very helpful for sorting out the usability of my research and learning how to talk about it in Swedish. Furthermore, thank you to all the administrative staff at the department, Linda, Ann, Mikael, Tomas, Marcus, Marie, Olga, Olga, Helena, Johanna, Kristina, Hilde, Louise, Toula, Siv, Kerstin and the Director of PhD Studies, Eva, for helping me deal with all practical aspects of a PhD student’s work.

Thanks also to all of my friends and family outside the academia who have provided a necessary intermezzo from the hard work and for putting up with interrogations of how you use science in your work and everyday life. Special thanks go to Mom, Dad and my supportive siblings as well as all my adorable nieces and nephews who reminded me for whom we are doing research in science education. I am sorry I was absent from socialisation spring and summer 2015. I will have to make up for it in future vacations.

From a sunny balcony in Stockholm in August 2015
List of papers


IV. Anker-Hansen, J., & Andrée M. In Pursuit of Authenticity in Assessment of Scientific Literacy. *Manuscript*
## Contents

Acknowledgments .................................................................................................................. vii

Introducing the dilemma ......................................................................................................... 12
  Overarching purpose of the thesis ....................................................................................... 16
  Outline of the thesis ............................................................................................................. 17

Taking a theoretical position ................................................................................................... 18
  Communities of practice ....................................................................................................... 18
    Central ideas in the communities of practice framework ................................................. 19
    Critique against studying schools as communities of practice ...................................... 22
  Expansive learning ............................................................................................................... 23
    Central ideas in expansive learning.................................................................................. 25
    Critique against expansive learning............................................................................... 25
    Expansive learning in relation to CoP ............................................................................ 26

Framing the research questions ............................................................................................... 28
  Affordances and constraints for assessment of scientific literacy as participation in civic practices ................................................................................................................................. 29
  Affordances and constraints for student participation in assessment of scientific literacy ................................................................................................................................. 33
  Affordances and constraints for communicating assessment of scientific literacy ................................................................................................................................. 35
  Affordances and constraints for the expansion of teachers’ assessment repertoire of scientific literacy ................................................................................................................................. 38
  Affordances and constraints for authentic assessment of scientific literacy .................... 40

Finding a path .......................................................................................................................... 42
  Praxis developing research ................................................................................................. 42
  Finding teachers for the project .......................................................................................... 44
    Pilot study ......................................................................................................................... 45
    The schools of the study ................................................................................................. 45
  Studying affordances and constraints for student participation in assessment of scientific literacy ................................................................................................................................. 46
    Analysing the data from the debates ............................................................................... 47
  Studying affordances and constraints for communicating assessment of scientific literacy ................................................................................................................................. 48
    Analysing the data from the peer assessments .................................................................. 49
  Studying affordances and constraints for the expansion of teachers’ assessment repertoire of scientific literacy ................................................................................................................................. 49
    Analysing the data from the assessment moderation meetings ..................................... 50
  Studying authenticity in science education research .......................................................... 50
  Methodological considerations ............................................................................................ 51
  Ethical considerations ......................................................................................................... 52
Presenting and discussing the results.................................................................54

Article 1: Affordances and Constraints for Using the Socio-Political Debate for Authentic Summative Assessment.................................................................54

Article 2: Using and Rejecting Peer Feedback in the Science Classroom: A Study of Students’ Negotiations on How to Use Peer Feedback When Designing Experiments...................................................................................58

Article 3: Challenging and Expanding Science Teachers’ Assessment Repertoires Through Social Moderation .................................................................60

Article 4: In pursuit of authenticity in assessment of scientific literacy........63

Concluding the studies ......................................................................................66

The designed and emergent of assessment of scientific literacy ....................66

Reifying quality in assessment of scientific literacy........................................68

Making scientific literacy discernible in student participation .......................69

Transferability through negotiation...................................................................70

Expanding practices........................................................................................71

Implications for authenticity............................................................................73

What more is to be done?..................................................................................73

It is worth the trouble of assessing scientific literacy as participation in civic practices........................................................................................................74

Svensk sammanfattning ....................................................................................76

Artikel 1............................................................................................................78

Artikel 2............................................................................................................79

Artikel 3............................................................................................................80

Artikel 4............................................................................................................82

References ........................................................................................................84
Introducing the dilemma

An important reason for making science education compulsory has been the need to better prepare people for citizenship, often referred to as scientific literacy (SL) (D. A. Roberts, 2007). With this aim follows a call for making clear and meaningful connections between the science taught in school and science used outside school. Parallel to this exists a necessity for keeping the classroom a virtual arena for training practices without the accountability involved in real out-of-school practices (Carlsgren, 1999). This seems to create tensions for designing education in general and assessment in particular. The more we arranges assessment in education for easy measurements of single aspects of scientific knowledge, the further one risks straying from the complex conditions citizens face when using science in the world outside school. For example, extracting and simplifying information for the students eliminates a necessary step of information selection and coding that students deal with when reading a newspaper or selecting products for purchase. With standardised and atomistic assessment design follows a loss of the holistic complexion that citizenship comprises (Ratcliffe & Grace, 2003). Rather than using precise test models and selecting content that can be measured with those tests, ‘authentic assessments’ are designed from what knowledge the educational system intends the students to be god at (Wiggins, 1989, 1990).

Assessment is authentic when we directly examine student performance on worthy intellectual tasks. Traditional assessment, by contrast, relies on indirect or proxy ‘items’—efficient, simplistic substitutes from which we think valid inferences can be made about the student’s performance at those valued challenges. (Wiggins, 1990, p. 2)

Central aims in the Swedish syllabi of biology, chemistry and physics, concerning the use of science in civic life, are described as the development of abilities to communicate, critically examine information and make decisions on topics and contexts concerned with science. The curricula and syllabi look almost the same in all three science subjects.

Teaching in biology should essentially give pupils the opportunities to develop their ability to: use knowledge of biology to examine information, communicate and take a view on questions concerning health, natural resource use and ecological sustainability. (The Swedish National Agency for Education, 2011, p. 105)
Teaching in physics should essentially give pupils the opportunities to develop their ability to: use knowledge of physics to examine information, communicate and take a view on questions concerning energy, technology, the environment and society. (The Swedish National Agency for Education, 2011, p. 120)

Teaching in chemistry should essentially give pupils the opportunities to develop their ability to: use knowledge of chemistry to examine information, communicate and take a view on questions concerning energy, the environment, health and society. (The Swedish National Agency for Education, 2011, p. 135)

The Swedish syllabi standards are open for local interpretation and implementation, but teachers face demands of reliability; the marks given in summative assessments must be comparative to those of other teachers. Furthermore, the Swedish National Agency of Education stresses that grading of students should be based on rich and varied evidence (Skolverkets Allmänna Råd, 2011, 2012). Swedish science teachers have expressed confusion and a lack of ideas, particularly on assessing communication, source critique and decision-making (Lundqvist & Lidar, 2013). There is consequently a need to aid the development of assessment practices for those abilities in Swedish schools.

One common way to assess the abilities to communicate, critically examine information and make informed decisions are to regard these abilities as propositional knowledge where the application of scientific knowledge in new contexts is considered unproblematic. Such assessments often emphasise the correctness of scientific concepts used (Lyon, 2013; Lyons, 2006), for instance, marking if students are explaining nuclear fission properly when making decisions regarding energy sources. Thus, the assessment of SL becomes a measurement of the scientific core tools that students have at their disposal. Such an approach to assessment is reasonable from a view of knowledge as a possession that students carry with them through different contexts. Ideas of knowledge as possession can be found in theories of learning often referred to as ‘symbol processing’ theories, since people’s understanding of, for example, science is described as symbolic representations that are modified and transferred through new experiences (Bredo, 1999; Elwood, 2008). From a ‘symbol processing’ theory of learning, it is tempting to allow test items in any context to become signifiers of a student’s assumed general ability to use science. If assessment is a matter of extracting ‘symbolic representations’ from students’ minds, the complexity of authentic assessment could be regarded as unnecessary and even disturbing to the measurement. However, predictions about how students use science in various situations are regarded as complicated; science educators cannot assume that students who can explain scientific concepts will apply this knowledge or that this knowledge will influence decisions or other behaviours.
However, how students transfer experiences through contexts are too complex processes to be assumed or neglected (Gruber, Law, Mandl, & Renkl, 1999; Marton, 2006; Roth, 1998b).

Instead of regarding SL as applications of propositional knowledge, SL could be regarded as procedural knowledge situated in social contexts such as a science class. The situated learning perspective can be found in many theoretical frameworks originating from sociocultural theories (Bakhtin, 1986; Vygotsky, 1978). From this perspective, people’s use of science is unique to different contexts and the way people talk about subjects such as nuclear fission differs from situation to situation. Consequently, the assessment of students’ decision-making of energy sources rather concerns how students select between different sources for a specific purpose of the task, make moral, aesthetical and practical reflections, choose to make their ideas clear to a specific audience and negotiate those ideas so that they become functional for specific purposes.

The ‘situated learning’ perspective exposes assessment as a complex social practice demanding a vast and varied set of student performances to profile the student’s knowledge (Bredo, 1999; Elwood, 2008; T. D. Sadler, 2009b). However, there are ways to deal with this complexity. One approach is to design the assessment around a topic where science could be an important aspect. In so-called socioscientific issues (SSI), the assessments centres on how students resolve specific problems or issues together (Zeidler, 2014). In SSI, scientific literacy is understood as using a multitude of sources when discussing and deciding on issues, not only scientific evidence, as well as reflecting on moral and ethics (ibid). Moreover, SSI allow for employing multiple methods in the assessment, such as drama, art and debates (ibid). However, the SSI is predominantly developed as an approach for lesson design where the issue affords student engagement (Klosterman & T. D. Sadler, 2010). Discussions regarding what ways are ‘better’ for dealing with an issue have been peripheral. SSI researchers have yet to explore what counts as good quality in classroom assessment (ibid).

An alternative to situate SL in ‘issues’ is to regard SL as situated in practices such as shopping or healthcare (Lave, 1988; Rogoff, 1984). SL then becomes the ability to participate in a practice where science matters. SSI could involve participating in practices, but the focus in SSI has rather been on ‘the issue’ per se and how students learn to approach it critically and ethically than on how critique and ethics can be enacted in practices (T. D. Sadler, 2009a). A nurse participating in healthcare practices has a professional stake in the participation. Healthcare issues are not only treated as issues that are scrutinised from different angles because they are embedded in the nurse’s work. A nurse must thus learn how to deal with the issues within the boundaries of the healthcare community. However, different theories of learning use different metaphors to describe learning in practices. The ‘traditional’ way of describing learning is through metaphors of knowledge as
acquired possessions; the ‘new’ metaphor of learning is to describe processes of participation in a practice (Sfard, 2008). Learning a practice could be described in various ways with acquisition metaphors, for example, as reception, construction or internalisation of knowledge in interaction with teachers, other students or artefacts like texts and laboratory material (ibid). Assessment of acquired knowledge then becomes restricted to things that are quantifiable or predefined in some way (ibid). Assessment becomes a measurement of how much a student understands or how well a student can follow a defined routine. In the participation metaphor, learning is described through how students change participation and how those alterations in participation include them as members in communities (ibid). This metaphor for learning has special consequences for assessment since ‘doing’ and ‘inclusion’ are ongoing processes and consequently not restricted to what can be measured in a frozen moment. An individual can become more and less active in a practice and thus become more and less included. Nonetheless, this is described as a process within the activity and not as a level of accomplishment. For instance, elaborating an idea after experiencing positive feedback could be described as a process of inclusion. However, it could not be described as a person having acquired a higher level membership status. Consequently, it would not make sense to make a stop in learning and take a measurement with test items of what membership or identity had been accomplished so far (ibid). Assessment thus becomes a study of how students’ participation includes them in the practice. This could, for example, become observable through how students adapt communication with purpose and audience and how they develop ideas from peers. The assessment is thus an evaluation of how students participate with other students. Such an approach is controversial to traditional approaches of assessing SL, like the Programme for International Student Assessment (PISA) in which students’ participation is compared to predefined ideals of performance (T. D. Sadler & Zeidler, 2009).

This thesis centres on some assessments that draw on SL as practices of citizens in general rather than professional scientists in particular. I could have chosen to draw on practices of professional scientists in the belief that those practices are idealistic and people in general would benefit from adopting them. However, practices in society vary and what is valued in a community of professional scientists might not concern other citizens or even be a resource for scientists themselves when participating in non-scientific communities. Indeed, R. L. Bell and N. G. Lederman (2003) found in a questionnaire that even professional scientists built their decisions about public issues on other grounds than only the presented scientific evidence, for example, personal relations to the issue.

Additionally, assessments of SL should take different forms depending on the purpose and conditions for the assessments. Purposes and conditions vary for high-stake tests used for evaluation of a larger population such as PISA,
and teacher classroom assessments used for making decisions regarding teaching. High-stake tests have high demands of standardisation, meaning that questions must be understood the same way by students and marked the same way by different assessors (Moss, 2007; 2003). For example, PISA is designed for the purpose of assessing SL (OECD, 2013). However, PISA is designed to be conducted all over the world in a great variety of countries with different school systems. This involves austere restrictions of conducting globally or nationally ‘fair testing’ which eliminates the possibility of including local, current and topical issues (Sjøberg, 2012). Moreover, standardisation often involves an individual written test format that offers few opportunities for students to engage in negotiations of economic, political and ethical aspects of complex societal issues (T. D. Sadler & Zeidler, 2009; Serder & Jakobsson, 2014). In classroom assessment, on the other hand, teachers and students have greater liberties to form not only assessment tasks that reflect the students’ everyday activities, but also criteria that encompass the multiple views of what is valued as ‘good performance’ in different civic contexts (Ratcliffe & Grace, 2003). It is consequently not necessary for teachers to restrict themselves to traditions of written exams decontextualised from complex circumstances, where there are correct or ideal ways of understanding the task, answering questions or scoring performances. Externally produced text are often carefully designed and tested through scientifically developed methods (Brennan, 2006). Nonetheless, teachers can better assess certain specifics of SL locally. Teachers can explore alternative and more authentic assessments set in complex environments, with open-ended tasks where quality of performance can be negotiated for different purposes. Science education research can thus not only focus on large-scale testing of SL, but also needs to aid teachers in the development of classroom assessment of scientific literacy.

Overarching purpose of the thesis

The purpose of this thesis is to contribute to the understanding of the affordances and constraints for developing practices for classroom assessment of SL as participation in civic practices. Such contributions involve investigating possibilities for teachers to design assessments that are authentic to participation in the practices the assessments are intended to evaluate. This includes studying affordances and constraints for students to participate in ways that are actually valued in different civic practices. It also includes studying affordances and constraints for teachers to articulate and communicate how students’ participation is valued, as well as how teachers can act from the information received from those processes.
Outline of the thesis

A doctoral thesis is commonly developed from a research review guiding the reader towards the research questions, followed by a description of the theoretical framework and the methods used to answer those questions. However, since the thesis is grounded in a theoretical positioning, it will start by introducing the theoretical frameworks in *Taking a theoretical position*. This will be followed by a description of how the theoretical frameworks guided my focus in the fields of science education and assessment research. The pursuit of an area where I could add something new to those research fields and my research questions will be presented in *Framing the research questions*. My choice of methods to answer the research questions will be presented in *Finding a path*. How my studies ended up in four article manuscripts will be presented in *Presenting and discussing the results*. Finally, the studies implications for classroom assessment will be discussed in *Concluding the studies*. 
Taking a theoretical position

As discussed in the introduction, I have chosen to describe SL through participation in civic practices. One commonly applied theoretical framework using a participation metaphor on learning is the communities of practice theory.¹ I first encountered the communities of practice theory in my first PhD course, and this theory provided me with a framework for observing assessment as something made manifest through student and teacher participation rather than simply an artefact to which they related. Communities of practice provided concepts for describing how students and teachers transformed the assessment practices through participation. I thus found the framework excellent for describing how students and their teachers as a group could develop assessments. However, the framework did not help me describe how schools were motivated to refine its practices in relation to the society of which the school was a part. From my understanding of the community of practice theory, the framework only described learning of organisations as ways of sustaining internal communication and coordination so that the organisation effectively delivers what it is supposed to deliver to society (Lave & Wenger, 1991; Wenger, 1998). I then found the theory of expansive learning developed from third-generation Cultural Historical Activity Theory (CHAT) by Yrje Engeström (1987; 2001; Y. Engeström & Sannino, 2010). The theory of expansive leaning helped me describe how the teachers negotiated expansion of its practices as solutions to the contradictions between what they had been doing and new demands they faced.

Communities of practice

The concept of communities of practice (CoP) was first used by Jean Lave and Etienne Wenger (Lave & Wenger, 1991) to describe learning and cultural reproduction in a group of people sharing a craft or a profession. The concept of CoP was initially not framed, especially not in relation to formal education. Moreover, the central concept of Lave and Wenger’s joint work was ‘legitimate peripheral participation’, which was described as the process in

¹ For further reading about the widespread use of communities in practice in education, refer to the anthologies of Kimble, Hildreth and Bourdon (2008a, 2008b).
which newcomers of a workplace learn the trade and become better able to carry out tasks in ways that are more valued by their colleagues.

Wenger (1998) later reframed much of his and Lave’s earlier work and developed the framework to include all forms of education. CoP became a central concept and was defined as people: (a) being mutually engaged in activities in which they form collaborative relationships, (b) through interactions come to share understanding of the activities and becoming part of a joint enterprise, and (c) forming a shared repertoire of resources and values. The mutual engagement, joint enterprise and shared repertoire were not regarded as defined, but in constant flux and negotiation by a community’s members.

Wenger talks about learning as negotiation of membership being dependent on members’ engagement in practices. Furthermore, learning is an integrated part of people’s interaction and can be observed through the changes in the repertoire of the community and engagement by its members (Wenger, 1998). This means that as students and teachers engage in school activities, they come to share values on how to act in the classroom.

Central ideas in the communities of practice framework

Resources and values of a CoP are under constant negotiation by its members – newcomers and veterans alike (Wenger, 1998). The affordability of becoming included in the practices as well as members’ ability to participate in ways that are valued by the community is referred to as the mutuality of engagement. In the science classroom, this involves, for example, using a particular language or providing certain kinds of questions and replies (Lemke, 1990) as well as conducting practical work in certain ways (Andréé, 2007). Membership in a CoP is, however, not only a matter of relating to a predetermined set of rules. The conditions for community members to influence values and resources are described as the negotiability of the repertoire. Introduced artefacts, questions and replies meet reactions from the teacher and the students in the classroom that convey how they come to value participation (Kress, Jewitt, Ogborn, & Tsatsarelis, 2001). Through their participation, teachers and students negotiate what they count as ‘good participation’, that is, the repertoire of the classroom. Some of these values draw on traditions and expectations of what usually happens in the classrooms, and some values are imported from practices outside the classroom (Roth, 1998a). Students in a science classroom are not newcomers or veterans like in a workplace. Nonetheless, discourses are negotiated by groups of students through exchanges with and reproductions from other groups, for example, between students from different school years talking about what happens in science class. Furthermore, teachers bring experiences of what has previously been negotiated in classrooms into new classrooms. As the repertoire of the CoP is negotiated by past and present members of a community, the members
become accountable for the enterprise. However, this accountability may take different manifestations depending on how power is divided in the community.

Learning as meaning and identity
Learning in CoP is described in terms of meaning (learning as experience) and identity (learning as becoming) (Wenger, 1998). Meaning and identity are regarded as processes as opposed to properties or traits. When engaging in civic endeavours, people negotiate meaning of how different aspects such as science are important for accomplishing something through experiences of what works and how the community values those aspects. Similarly, people position themselves through their engagement through identities such as knowers and supporters.

Meaning and identity as participation and reification
Wenger provides two central analytical terms to describe learning: participation and reification (Wenger, 1998). Participation is described as ‘the social experience of living in the world in terms of membership in social communities and active involvement in social enterprises’ (ibid, p.55). Wenger writes that participation both involves taking part and the relations that reflect taking part. Therefore, participation does not necessarily involve direct interaction with other members of the community, but could constitute any activities that have social purposes, for instance, doing homework or writing and reading this thesis. People can thus participate in multiple communities’ practices without being in direct contact with other members of that community. A person may, for instance, think about work when going to bed. Participation could be collaborative as well as conflictual. Not only do people’s experiences change when they participate, but also the communities in which people participate (ibid). Wenger writes continuously that experiences are renegotiated and thus mutable through various engagements. Although participation becomes unique in each context, it also influences other contexts that people have and will experience. Through current actions, individuals make new meaning of past events. What they experience through current actions also opens up new options for future actions.

Reification is used to describe ‘the process of giving form to our experience by producing objects that congeal this experience into “thingness”. In so doing we create points of focus around which negotiation of meaning becomes organized’ (Wenger, p.58). As I understand, reification in the science classroom involves, for example, using a specific language to explain phenomena, or connecting laboratory work with scientific concepts and theories. Reification could be very concrete through, for example, the formulation of rules, routines, or criteria. It could also be subtly embedded in conscious actions, such as using a graduated cylinder instead of a teaspoon for measuring volumes of liquid in the chemistry laboratory.
Participation and reification are described as simultaneous processes, but they can be more or less pervasive. Wenger (1998) gives an example in which he describes the activities of a flower as being extremely participatory, whereas a computer’s description of these activities is extreme reification. The flower ‘knows’ everything there is about flowers by doing what flowers do (participation), but it cannot explain this (reification). On the other hand, regardless of how well a computer can explain what a flower is (reification), it cannot do what flowers do (participation). However, the computer could never reify what it means to be a flower without flower participation, and one would not be able learn what flowers do without some reification. Most often the process of reification draws from a history of participation; routines and rules are organised from what has been experienced to be working well. Furthermore, participation usually rests on tools, such as a vocabulary, that are reified by the community. However, problems are likely to emerge when new rules are introduced that are experienced as strange compared to what community members usually do or when community members have to do new things for which there are no directions to draw from.

If participation prevails – if most of what matters is left unreified – then there may not be enough material to anchor the specificities of coordination and to uncover diverging assumptions. This is why lawyers always want everything in writing.

If reification prevails – if everything is reified, but with little opportunities for shared experience and interactive negotiation – then there may not be enough overlap in participation to recover a coordinated relevant or generative meaning. This helps explain why putting everything in writing does not seem to solve all our problems. (Wenger, p.65)

If the development of the praxis is left to the community members’ participation without reification in the form of policies, instructions or criteria, community members risk finding it difficult to coordinate and compare their participation, resulting in misunderstandings. Correspondingly, if the practice is reified clearly in text but is not grounded in the community’s shared experiences and what is negotiated to be valued participation, the text risks being meaningless to the community’s members.

**CoP as an insider’s perspective**

The CoP research takes an insider’s perspective when describing negotiation of meaning and identity. Trying to study insiders’ meaning-making of their own culture is sometimes referred to as taking an ‘emic’ perspective (Zhu & Bargiela-Chiappini, 2013). In practice, this perspective often conflicts with the view of what counts as good scientific practice, as predefined by a larger community. Such a viewpoint is referred to as an ‘etic’ perspective, since it is based on the notion that communities have a predefined, formal and informal
set of rules that newcomers learn to follow (ibid). From an etic perspective, people’s actions could consequently be judged as more or less in alignment with those rules. One could, for instance, make a model for how scientific argumentation should be conducted and measure to what extent students or teachers follow this model (cf. V. Sampson & Blanchard, 2012; V. Sampson & Clark, 2008; V. D. Sampson & Clark, 2004). Contrariwise, analysis from an ‘emic’ perspective is based on what is locally reified as valued participation through how advantageous or disadvantageous the participation becomes (Tatli & Özbilgin, 2012). This does not necessarily mean that everyone agrees with what one person does or says in order for participation to become advantageous. From the perspective of, for example, the CoP framework, values and resources are in constant negotiation by the community; advantageous participation is rather what is included in this negotiation. This could be observed by how the members in a community are building their argument on prior statements and thereby use other members’ participation.

**Prior operationalisations of CoP in science education and assessment research**

Further operationalisation of Wenger’s theory into a theoretical framework applicable in assessment or science education has been presented in several research projects. Roth and McGinn (McGinn & Roth, 1999; Roth, 1998a; Roth & McGinn, 1998) used CoP early to frame science students learning how to become scientists. Munby, Taylor, Chin, and Hutchinson (2007) used Wenger’s concept of border brokering, that is, how repertoires from one community become central in the practice of another. They describe how students integrate their participation of classroom education and professional training. Kisiel (2010) used the concept of participation to analyse how the practice in in the two communities of a classroom and an aquarium differed and were integrated into each other. Willis (2011) utilised the framework to describe how students developed an autonomous sense of assessment by participating in peer- and self-assessment practices. She studied inclusion through the processes of peripheral legitimation in student participation.

**Critique against studying schools as communities of practice**

Using CoP to describe formal education has been argued to be problematic for several reasons. Haneda (2006) criticises using CoP to explain the practices of the classroom since the framework does not consider different forms of learning and the concept of the class as a community has not been thoroughly framed. She further argues that learning in school and workplaces cannot be studied analogously because of the unequal power structures of the teacher and the pupils in schools, as well as pupils’ participation not being voluntary as in workplaces. A question also arises concerning whether students can start as apprentices and become masters in a classroom; Haneda argues that
students cannot become teachers by participating knowledgably. The kind of apprentice-master relationship she is referring to is described in an example of tailors in West Africa, one of the five practices used as examples in Lave and Wenger’s (1991) initial joint production. However, in another example from the same work, the apprenticeship of sobering alcoholics in AA meetings does not involve aspirations to become sponsors (masters), but rather to receive advice and support towards sobriety. This means that negotiating membership in a community’s practices (through legitimate peripheral participation) does not necessitate replicating veterans, teachers or other people in charge. Furthermore, Roth (1995, 1998a) found that innovative students might be granted an informal status as teachers and resources of knowledge for fellow students.

Roth, Hwang, Mafra Goulart and Y. J. Lee (2005) argue that the division of labour is actually an important part of what defines a community. Though the labour is divided in a school community, the school becomes a community of (learning) practice because the activities are reproduced toward the object of the activities, which in a school is learning. Roth et al. use a terminology from Cultural-Historical Activity Theory, where Leontiev (1978) describes that the labour is divided in an activity of a hunt with the object of gaining food. In school, these reproductions towards the object of learning could be a repetition of activities where learning is observed, and an amendment of those that do not produce signs of desired learning are amended. However, the object of learning can change, for example, with new curricula, and teachers may still reproduce the same activities towards the new objects. Roth et al. (2005) claim that the CoP cannot be isolated to a classroom where the community is formed in the beginning of the school year and then dissolved again at the end of the year. However, there is continuity in connecting activities in the school system as a whole. It is a school and not a class that is the community of practice from Roth et al.’s perspective.

Expansive learning

Operationalisations of CoP have mostly been used to describe short learning encounters. A framework for studying activity over time is Engeström’s (1987; 2001; Y. Engeström & Sannino, 2010) theory of expansive learning. Felstead et al. (2005) claim that the theory of expansive learning uses the participation metaphor, but Engeström himself (Y. Engeström & Sannino, 2010) argues that neither the acquisition metaphor nor the participation is sufficient for describing creative and transformative processes, as they restrict themselves to describe learning as a one-way directional move from incompetence to competence. Engeström claims to use the ‘expansion metaphor’ for learning. Expansive learning is based on Leontiev’s (1978) division between action and activity. Leontiev criticised contemporary
Russian psychologists of trying to create ‘cybernetic approaches’ by adding up explanations of the actions of individuals when analysing activities in a community. Leontiev argued that human activity is embedded in systems of culture, artefacts, etc., which need to be taken into account when understanding the activities. Activities are motivated, evaluated, reproduced and transformed through the object of the activity. My understanding of this is that, both teacher and student participation comprise SL in school science. If one, for instance, only compares teachers’ intentions with students’ perception of science education, one misses how students and teachers interact to constitute the object of SL. In CHAT, the whole activity system is important for understanding what happens. An activity system such as science teacher education is composed of subjects, for example, students and educators being part of a community sharing an object, learning to become science teachers, dividing the labour of the activities between them, teachers introducing didactic issues and students discussing solutions according to certain rules such as the curriculum, using mediating artefacts like discussion questions and laboratory equipment, producing an outcome of science teachers.

Engeström used third-generation Cultural-Historical Activity Theory (CHAT) to explain learning where what is learned is not being predefined by someone who has already mastered what is to be learned. Digitalised routines, staff changes, new policies, etc. involve changes in the whole activity system; people continuously learn to deal with these changes even though experts are not instructing them what to do. Engeström (1987) considers schools to chiefly have been studied as a subject-production activity (what teachers or students do) and science as an instrument-product activity (what one can say from inquiry). In contrast, expansive learning is studied as an activity-product activity (how the activity and the outcomes of the activity change). This refers to studying how the schools transform as different components in the system changes. For example, a school’s mission might change, or the introduction of computers in the classroom may change how the activities are mediated. In schools, teachers try different approaches to facilitate the development of the students’ knowledge and evaluate these activities from an object of educating them for further studies or handling endeavours outside school. Expansive learning thereby involves the movement from action to activity (Y. Engeström & Sannino, 2010).

Expansive learning happens in the zone of proximal development (ZPD). ZPD is a concept taken from Vygotsky (1978) that describes the distance between what people can do without help and what people can do with help. It is here, in the ZPD, where the community makes meaning and sense of what they are doing in relation to the object of the activity. It is here where contradictions emerge between what has historically been done and what is necessary to do in the present. How the community collectively chooses to resolve the contradiction can be observed as expansive learning.
Central ideas in expansive learning

In expansive learning, activities are described as part of an *activity system*. In this system there are *multiple voices* about, for instance, what to assess and how to value students’ performances in science class. However, teachers share traditions of assessment activities to which they relate when they engage in new assessments. Therefore, a shared *historicity* connects the multiple voices. Through that historicity, the overarching purposes for the practices remain continuous over time (Y. Engeström, 2001). The varied forms of assessment essentially have common goals of evaluating student knowledge for formative, summative or accountability (for the authorities) purposes (Black & Atkin, 2014). In an activity system, various views, mediating artefacts, rules, ethics, division of labour or new subjects can create *contradictions* between, for example, old assessment emphases and reformed syllabi standards. As the contradictions are translated and negotiated, changes occur in the activity systems described as *expansive moves* of the activity, repertoire etc. (Y. Engeström, 2001). These movements emerge as people question old habits and thereby expand the activities of the system in new directions that are meaningful for those people (ibid). Studying the historicity of activity systems does not need to involve ethnographic studies over long periods. People continuously bring up how things used to be and what has commonly been done in discussions with other people in a community and thereby provide information about the history of the activity system. Engeström (2001) conducted a study on Finnish healthcare where data on historicity were gathered through peoples’ narrations.

Critique against expansive learning

Expansive learning has not been criticised as much as CoP. Contrariwise, it has been used in critiques against other frameworks (A. Edwards, 2005). Critiques have been proposed from other CHAT users in complex discussions regarding how to look at the activity system and the significance of different components for the transformation of the activity system (Engeström & Sannino, 2010). Langemeyer (2006) claims that Engeström’s form of expansive learning is limited to situations where people can look critically at their own practices in a willingness to improve. I think that Langemeyer have a point here. The whole point of Engeström’s theory is that practices expand as a result of contradictions and conflicts. However, changes in repertoires might evolve without apparent conflicts. Changes might, for example, arise out of curiosity; what would happen if I did this instead of what I usually do? Young (2001) makes a similar notation that the theory of expansive learning does not properly describe changes that do not arise from knowledge lying in the practice itself. Expansive learning could, for instance, not explain how medical doctors become motivated to utilise new medical treatments.
developed in research environments. If one considers medical research and healthcare as different practices (which is not an obvious division), one can argue that there are border brokers, such as pharmaceutical representatives, who visualise contradictions between old treatments and new needs. Although new pharmaceuticals may not be the result of negotiated contradictions by medical doctors adopting new treatments, the contradictions still come from shared needs to treat the medical problem.

Expansive learning in relation to CoP

According to Wertsch (1998), sociocultural theory (like CoP) and cultural-historical activity theory (CHAT) have much in common. Both sociocultural theory and CHAT researchers have reacted disapprovingly to the social reductionism of prior research ascribing accounts of human actions to social forces alone (ibid). Concepts such as learning and identity can not simply be explained as products of society, family environment, etc. Furthermore, sociocultural theory and CHAT both oppose descriptions of appropriation of new knowledge as merely a matter of students adopting new actions from the teachers, but rather as a generation of new meaning between people interacting, according to Wersch. However, in sociocultural theory, as formed by Vygotsky (1978) and Cole (1996), the unit of analysis becomes a mediated action and in CHAT, it becomes an activity. In Leontiev’s classical example with the hunt, sociocultural theories can consequently be used to study what meaning the hunting tools and procedures have to an episode of a hunt. From CHAT, the focus would be on how the cultural history of the activity system is oriented around the objective of killing game (Wertsch, 1998). How this difference in focus can result in different conclusions can be seen in Chaiklin’s (1999) research, where he deduced that activities in the classroom could not motivate the children to learn atom physics, per se. Chaiklin did not see that a lesson activity included (or failed to include) the students becoming practitioners of atom physics. However, by analysing how the entire science educational system could be transformed to become closer to the practices of physicists, he could see how education afforded opportunities for students to be included as practitioners.

Nardi (1996) is most critical of what she calls ‘situated action models’ (like CoP), which she claims can only be used for analysis of an activity of people in a setting. The arena is described through situated actions that are ‘edited’ from the personal needs of students and teachers. Thus, the objects of the activities become retrospective reconstructions in the situated action model. Contrariwise, activity theory offers a richer description of activities over time by explaining the motivations for people’s actions through the object of the activity. Nardi claims that objects are stable over time and do not change in a moment-by-moment basis, as situated action models imply. Objects do, however, transform over time, and such transformation can change even the
nature of the activities. For instance, in teacher classroom assessment, the object may grow from simply reporting ‘results’ to authorities in the form of a mark, to an investigation of students’ meaning-making of the lesson activities for the purpose of improving the lesson design. How the longevity of the study can make it possible to discern unpredicted emergent meaning of science is described by van Eijck & Roth (2009). An internship in science did not make an aboriginal science student feel included in the science practice or motivated to become a scientist, but expanded the student’s view of what science encompasses. Thus, the student could find new ways to pursue studies in science without becoming a scientist.

However, what is gained through sociocultural and situated learning theories is also lost to some extent in CHAT, like a binocular zooming out, gaining a wider angle but less magnification. Consequently, there are reasons to let the theories complement each other and gain both a wide and deep insight of learning activities. Andrée (2012) did this in a study about how a specific object of the activities in the science classroom could emerge that differed from the object of the educational system as a whole. A student who took the position of a failed science student not being included in the culture of the science classroom was able to make her own expansion of an experiment. Through a mistake in the experiment, the experimental task became a personal problem for the student that she could handle on her own terms.
Framing the research questions

Drawing on the CoP framework, I discern two major challenges with developing assessment of SL as participation in civic practices:

1) Assessment is an evaluation of quality, and thus involves a challenge to reify what good participation means in relation to civic practices so that improvement can be communicated. From a CoP perspective, communication about quality and improvement involves negotiations toward alignment in the community. These negotiations are continuous processes, since students’ work changes. However, there is a question to which extent alignment really is necessary in assessment. Can students and teachers not be allowed to reify diverging values of what it means to participate in civic practices? Alignment is not the same as conformity of opinion. There are no problems with people finding the same novel intriguing or dull, but a student may get upset if one teacher gives an A and another a D on the same essay. The reason is that the two teachers reify quality differently in a context where the reification of quality is claimed to be comparable. Alignment in CoP is a coordination of multiple localities, competences and viewpoints (Wenger, 1998). As I understand this, alignment can just as easily be a negotiation towards influence or consensus as disempowered yielding to a majority of peers or the management of the teacher. However, from a CoP perspective, negotiability is important and enforcing alignment might be counterproductive. If teachers and students value participation in conflicting ways, that is, students believe they are supposed to do something but the teacher assesses something else, feedback or marks risk being misunderstood or rendered useless. For instance, students could be led to believe that they are supposed to draw from their everyday life experiences, when the teacher actually wants them to apply the scientific theories covered in class. The feedback and marks then risk only becoming signifiers of how well students can do in science class, implying that students experience what they do in science class as having little to do with their participation outside the classroom. However, communicating the reification of quality is complicated. From a CoP perspective, students shape the practice by participating, and ambitions to predefine quality before students have acted might alienate the assessments from students’ experiences of society. Let us for instance imagine that a teacher designs an assessment of students’ decision-making and in that design defines a template for the ideal decision-making process. The teacher then scores the students on how close their performance comes to that template. Though that model might seem
transparent and fair, it is rigid in its design; the teacher will be less able to assess the variety of ways in which people can actually make decisions than would be possible with a more open-ended assessment. However, it is equally problematic to treat assessments as if there were nothing to draw from when reifying quality. Citizens have made relatively informed decisions on civic issues concerned with science before a teacher designs exercises and assessments about decision-making, and the students have made relatively informed decisions before entering the science classroom. This challenge involves making such experience a resource for negotiating alignment of what works better when making a decision on a civic issue.

II) Assessment is constrained to what an assessor can discern from students’ actions, and teachers are thus faced with the challenge of designing assessments so that students’ abilities to participate in civic practices can become observable. Grading may be (and unfortunately sometimes partially is) based on what teachers believe or guess that students can do. However, such grading is neither valid nor reliable since it is not based on the assessment of evidence (Harlen, 2007). Assessments, on the other hand, involve decisions made on evidence gathered from students’ activities. Besides providing evidence for grading and making decisions about teaching, the assessments provide students with experiences of inclusion or exclusions, confidence or insecurity, positive or negative dispositions to activities valued in the assessments. It is thus not possible to design assessments that can objectively infer students’ abilities to participate in practices without simultaneously affecting the students. The consequences of assessments thus constrain what a teacher can make discernible in an assessment. The previously imagined assessment designed with a predefined ideal for decision-making might discourage students from exploring different functional ways of making decisions, because one way has been introduced as better than others. That imagined assessment might give evidence about how well students can learn to follow a given model. Nonetheless, not much information is gained about how the students would make decisions when participating in a civic practice, since students are discouraged from exploring their own ways. Consequently, this challenge involves designing the assessments to gather information with respect to how that information is intended to be used and the desired consequences of experiencing the assessment.

Affordances and constraints for assessment of scientific literacy as participation in civic practices

The affordances for developing practices for assessing SL as participation in civic practices are subject to the potential of designing activities and articulations of quality with which participants can identify. Researchers have
come to various conclusions not only regarding how to describe the necessary qualities of a scientifically literate citizen, but also how to discern or measure those qualities in assessments (D. A. Roberts, 2007). Assessments may vary either in how the assessed abilities are constructed, for example, as procedural participation, or as propositional recollections and how those constructs are observed, for example, through cooperative open-ended tasks or focussed test items with ideal solutions. Furthermore, ideas in research and policy documents often require further operationalisation into concrete classroom assessment practices. Teachers thus face challenges not only concerning how to navigate conflicting messages about what SL can be and how it can be assessed, but also concerning how to transfer those messages to their particular purposes with their current classroom assessment.

The approach to SL has taken different paths in the European and American fields of science education. Whereas science education in America has mainly concerned how to reform the curricula from different perspectives or visions of SL, the European didactics traditions have developed analytical instruments to support teachers’ professional decisions in the classroom (Wickman, 2014).

The need to reform science education curricula arose from multiple instances in the United States in the late 1950s as a response to Russian technological advances (Bybee, 2007). The public’s understanding of science had to be developed to speed up the technological advances of the nation. The solutions suggested centred on making citizens adopt repertoires from professional scientists. This approach was later named Vision I (D. A. Roberts, 2007; D. A. Roberts & Bybee, 2014). Another approach was to look at what activities citizens engaged in and work with science among a multitude of other aspects relating to these activities. This was called Vision II (ibid). Though different research fields work differently (Knorr Cetina, 1999; Latour, 1999; Pickering, 1995), the Vision I approach seemed to try to define sets of discourses with ideal forms of participation to draw from when assessing students’ performances (c.f. Miller, 1998; Sandoval & Reiser, 2004). Promoters of Vision II, however, struggled with using values from a complex web of knowledge with a vast variation of how science was used and valued in different communities (Ratcliffe & Grace, 2003; Roth & Barton, 2004; Zeidler, T. D. Sadler, Simmons, & Howes, 2005). Millar (2002) posed the question of whether it is possible to create assessment that both works as an evaluation of readiness to use science in everyday contexts and to prepare for the education of future scientists. Such assessment faces the challenge of finding criteria that reflect both goals. He argues that first a framework has to be established from studies of how to conduct scientific work. Secondly, test items have to be constructed that focus on the particular elements described in the framework (Reiss, Millar, & Osborne, 1999). However, as described earlier, such an approach presupposes that the students’ scientific activities are already defined and that the students are to adapt to existing rules (Zhu & Bargiela-Chiappini, 2013).
In the European didactics tradition, the selection of scientific content is more central than in the American curriculum tradition. The research field of science didactics centres on four questions: ‘What content is taught? How is this content going to be taught? Why teach this content and why with these methods?’ (Wickman, 2014, p.146). The same questions can be asked for the assessment of SL specifically. From a didactics tradition, the assessment of SL is just as much the choice of scientific content as the design of the assessments. The selection of content includes predictions of what knowledge might be relevant for students to develop and what is possible to assess. The selection also involves transpositions of knowledge from different communities of practice into activities in the science classroom (ibid) and in so reflecting on how that content manifests in student performance in assessments. In so doing, teachers consider why this knowledge is necessary to assess and why a certain assessment design is optimal for assessing this. Furthermore, it comprises making the purposes for engaging in the assessments understandable for the students (Johansson & Wickman, 2011) so that it becomes clear in the assessments why students should perform and why some forms of performances are better than others. For example, the assessment is designed to clarify if the purpose of the assessment is to give an elaborate explanation of photosynthesis or if photosynthesis is merely a concept that is to be used to make a decision about energy sources. However, the four questions summarised by Wickman are not restricted to a specific perspective on learning. Ekborg, Ideland and Malmberg (2009), on the other hand, specify didactic questions concerning SSI. What is the starting point? What school science curriculum goals are approached? What is the nature of scientific evidence? What is the most important social content? Why is scientific knowledge important? What kinds of conflicts of interest are there? From a CoP perspective, answering the didactic questions becomes a process of negotiation in continuous interplay between students’ participation and reification and the teacher’s purposes. The didactical questions are rather expressed as how to distribute a design between reification and participation; what to reify, when and in relation to what participation (Wenger, 1998). The answers to those questions may be the same as to the question summarised by Wickman (2014) or Ekborg et al. (2009). However, from a CoP perspective, the practice cannot be seen as a result of design, but rather a response to design. Practice and meaning cannot be designed, but the affordance for negotiation of practice and meaning can (Wenger, 1998). Unpredicted responses to teaching emerge from students’ participation and are thus important elements of the practice; they should not be understood as disturbances that need to be removed. For example, a student might ask a question that, from the perspective of the teacher, is irrelevant to a topic, but is highly relevant to several students. Furthermore, nobody can claim sole ownership of the practice. Though authorities like the Swedish National Agency for Education might give directions for lesson design, the Agency can
never fully organise what happens in classrooms. This does not mean that
teaching cannot be designed to be dogmatic and restrict negotiability.
However, reactions to such teaching will likely be superficial compliance (or
silent protests).

Though the selection of relevant scientific content and the transposition of
that content are central in the European didactics tradition, those selections do
not necessarily involve drawing from civic practices. However, a concept that
has been used in environmental education to describe students’ preparedness
to take action to preserve a sustainable environment is ‘action competence’
(Breiting & Mogensen, 1999; Jensen & Schnack, 1997). These ideas of action
competence align well with the CoP perspective in that they describe
processes negotiated by the students. However, action competence has been
used as an educational ideal, rather than a goal to be assessed, and criteria for
‘action competence’ were initially not constructed (Mogensen & Schnack,
2010). Mogensen and Schnack (ibid) suggest that the assessment of action
competence should be a self-evaluation by an educational system and not a
measurement of whether students’ actions are correct. Rather than
constructing criteria, assessment becomes a search for ‘indicators’ that the
learning activities are characterised by democratic views and where citizens
are held accountable as agents. The indicators they suggest are that students
are encouraged to: a) work with power relations and conflicting interests, b)
consider different perspectives and identify themselves with others, c) present
different arguments and d) explore alternative actions. Consequently, the
teacher performs the assessment of action competence to improve the lesson
design. However, the focus, as described here, is on the tasks rather than on
what students do with the task. Teachers cannot only evaluate their intentions
with their lesson designs; they must also observe what the students do in the
classroom. Almers (2013) claims that action competence can be observed
through: a) students’ desire to change conditions, b) students’ expression of
values and contrasting perspectives, c) creative representation of role models
rather than mimicking ideals, d) confident contributions from the students, e)
students feeling entrusted to work freely without control and strict demands,
and f) students being included in the activities. However, Almers gives no
descriptions of how these characteristics can be observed in students’ actions,
leaving the operationalisation to teachers. In sum, research on assessment of
‘action competence’ gives rough directions about how content selection and
transpositions can be done and evaluated. In my opinion, there is consequently
still a need to investigate teachers’ affordances and constraints for designing
classroom assessment, where quality can be reified and discerned in students’
participation in civic practices.

The overarching research question I aim to answer with my studies is:

- What are the affordances and constraints for classroom assessment
  of scientific literacy as participation in civic practices?
Affordances and constraints for student participation in assessment of scientific literacy

From a CoP framework I regard assessment as a negotiation of meaning between teachers, students, steering documents, etc. Design, performance, evaluation and their consequences are in constant interplay with prior and future engagement in assessments as well as with the activities the assessments represent. Central to the assessment of SL as participation in civic practices is that it affords the kind of student participation in which citizens actually engage. In the CoP framework, this is referred to as mutuality of engagement. If students are led to believe that they can engage in a task like citizens usually do, but their participation is then judged through a scoring system representing a constructed ‘school science’ reality, the assessment becomes a poor signifier of student’s abilities to perform out-of-school tasks (Andrée, 2005; Serder & Jakobsson, 2011). It is furthermore important to observe how participants influence the assessments. From a CoP perspective, student engagement in the assessment is not only the individual interaction between the student and the task, though it might appear so when students are sitting by themselves scribbling answers on a paper. Through participation, both the teachers and students form and transform the assessments. This becomes much more obvious in oral examinations in which part of the task is to relate to other students or the teachers’ utterances. The affordability for students to influence the assessments is referred to as the negotiability of the repertoire in CoP. With the negotiability of the repertoire comes accountability for the enterprise. Each utterance from a student or a teacher influences the opportunities for further participation in the assessment. It is thus important to be observant of what responsibility is placed on the students in the assessments so that it represents realistic and fair civic practices (c.f. Gipps, 1995; Murphy, 1995). In sum, important aspects of the affordances and constraints for classroom assessments of SL are the mutuality of engagement, negotiability of the repertoire, and accountability for the enterprise.

There are different assessments where student mutuality, negotiability and accountability become salient in students’ interactions with each other. In group project work, students are depending on mutual negotiations for producing a collective performance. Group project work can afford communication about the need for improvement on a group level, but constrain the assessment of an individual student’s needs. Different oral examinations in which students discuss topics are thus more affordable for discerning the performances of each student. A socio-political debate is a form of assessment where students interact and, at least in theory, have opportunities to influence the course of the discussion. Furthermore, socio-political debates represent a civic endeavour that might be concerned with science. Scientific endeavours outside the classroom often differ in character from activities in the science classroom (Aikenhead, Orpwood, & Fernsham,
2011). For instance, the chief purpose for participating in a debate outside of school is to convince an audience of the benefits of one idea in favour of competing ideas. Assessments in schools have the overriding purpose of making knowledge discernible for the assessors. The challenge for using socio-political debates as assessments is thus to transform the experiences of debate practices into a school practice of articulating meaning-making of the topics debated. In this transformation, conflicting purposes emerge between experiences of what is desired in science classroom assessments and what is functional in the out-of-school contexts where debates can also be found (Sambell, McDowell, & S. Brown, 1997). Consequently, experiences from civic life outside the classroom can both work as resources, affording participation and constraining possibilities to display specific scientific knowledge to the teacher. A. L. Tan, Kim and Talaue (2013) showed that this dialectal movement between science and everyday experiences can be problematic for students. Aiding students with this movement is thus an important instructional goal. However, to make it an instructional goal, teachers must learn to discern those struggles in students’ participation.

More than a decade of analysing student discussions on topics concerned with science (Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Newton, Driver, & Osborne, 1999) have resulted in the development of assessment models for argumentation with defined levels of quality of argument composition (Erduran, Simon, & Osborne, 2004; Grace, 2009). Though studies exist on the impact of social structures of the discussion groups, argumentation research has primarily focussed on ways to articulate the quality of students’ argumentation skills (Jiménez-Aleixandre & Erduran, 2007). However, some research on role-play in science education describes how the roles can work as a mediator to broaden perspectives in science education (Kolstoe, 2000; Simonneaux, 2001; Ødegaard, 2003). This indicates that the participation of students is very likely to alter as a discussion turns in favour of a certain viewpoint. Consequently, there are resources from which to draw when articulating the quality of students’ arguments and conditions for social interactions to consider when setting up the socio-political debate as assessments. However, a gap seems to remain regarding the affordances and constraints of transferring debate practices from out-of-school into the classroom.

The research question for my first study is:

- What affordances and constraints for student participation emerge when using a socio-political debate for authentic assessment of scientific literacy?
Affordances and constraints for communicating assessment of scientific literacy

Another important aspect of affordances and constraints for assessment concerns the communication of the assessment. As explained, the communication involves reifying quality of student performance. Quality can be reified explicitly through written criteria or more implicitly through actions that are valued in apparent ways by the community. What is valued as ‘good quality’ is reified through experiences of what usually works well or has worked well this time. However, sometimes criteria, standards or feedback that can be reified both too generally or too specifically afford the improvement of participation.\(^2\) Criteria or standards can be reified from one task or one imagined performance, for example, comparing reaction speed in a concentrated and a diluted acid. However, they may become very strange when applied to another task such as identifying an unknown liquid, since that is a much more open and complex task. Criteria and standards then become poor descriptors of what actually works well. The same applies to feedback. If feedback does not reify quality as experience of what works in practice, the usability of the feedback risks becoming poor. Students might make alterations to their work because the teacher requires it, which is necessary for obtaining a high mark. Nonetheless, the student does not necessarily share the belief that the alteration actually makes the work better.

Opportunities for students to work with feedback can be found in the activities of peer assessment. Peer assessment can work as a tool for students to negotiate the quality of participation in various practices (Jönsson, 2013; Van Zundert, Sluijsmans, & Van Merriënboer, 2010). In assessment, drawing from various practices in civic life, there are no single right or wrong solutions. However, various choices might be more or less functional for different expressed purposes (Roth & Barton, 2004). An example of this is when scientific research gets published and read by an audience. What is valued as the more functional scientific research study to respond to a civic issue is negotiated between citizens, for instance, within the scientific community in a given field or between the scientific community and the general public. Though teachers or peers can afford the development of source critique in a desired direction through feedback on student work, students’ use of feedback is constrained by the negotiability of how to value preferable performance, and correspondingly, how feedback can improve the work (Gamlem & Smith, 2013). To study these negotiations of quality, the author chose to include students as assessors. The second focus of the thesis became

\(^2\) I use D. R. Sadler’s (1987) distinction between criteria as characteristics or properties used to assess knowledge, for example, the critical examination of sources, and standards as a defined levels of quality, for example, to ‘apply developed informed reasoning to the credibility and relevance’ (The Swedish National Agency for Education, 2011, p.112).
students’ discussions of quality and choices on how to use or not to use given and received feedback in peer assessment. The content the students’ teachers and I chose for this study of peer assessment was critical reviews of experimental work. The rationale is that citizens are constantly exposed to the results of scientific research; the ability to scrutinise scientific research can be considered an important aspect of scientific literacy. Furthermore, the scientific community continuously evaluates the credibility of scientific research. Thus, it could be useful for all citizens to experience how scientific research is produced and reviewed.

From a didactic perspective, the content of any activity, such as peer assessment, in the science classroom is important for how one understands the learning activities. Experimental and investigative activities are often made central in science education (Abrahams & Millar, 2008; Högström, Ottander, & Benckert, 2010). However, the importance of developing scientific experimental and investigative skills for citizenship is not obvious. Mak, Mak and Mak (2009) have studied examples of how citizens make use of what the authors referred to as ‘the scientific method’ in everyday life. However, the people in Mak, Mak and Mak’s examples never mentioned that they were using scientific working methods. Rather, they described how they approached issues systematically or logically. On the other hand, an expressed aim of science education is to develop students’ ability to critically examine scientific processes (Kolsto, 2001; Sandoval & Reiser, 2004). For example, in the Swedish curriculum standards, more advanced experimental skills are expressed as an increased understanding of the systematics in scientific experiments (The Swedish National Agency for Education, 2011). The development of these abilities to critically examine scientific processes is motivated both by a need to develop capabilities to deal with everyday problems as well as critical thinking. From a citizen’s perspective, it might be useful to know how scientific knowledge comes to be and thereby be able to view media with a scientific content more critically (Bybee & DeBoer, 1994; Klosterman, T. D. Sadler, & J. Brown, 2012; Nicolaidou, Kyza, Terzian, Hadjichambis, & Kafouris, 2011; Sandoval & Reiser, 2004).

Scrutinising experimental work concerns knowledge about the nature of science (NOS), as well as knowledge of scientific inquiry (SI) (Ibrahim, Buffler, & Lubben, 2009; Khishfe, 2008; N. G. Lederman, Abd-El-Khalick, R. L. Bell, & Schwartz, 2002). What NOS encompasses varies, but it usually concerns developing knowledge about science as tentative, empirically-based, and subjective, it necessarily involves human inference, imagination, and creativity, and it is socially and culturally embedded (N. G Lederman, J. S. Lederman, & Antink, 2013). Closely connected, but with a different focus than NOS, is SI. SI concerns learning the processes of making scientific discoveries. However, it also refers to scientific reasoning and critical thinking (ibid). Though NOS and SI comprise different aspects of scientific knowledge,
Schwartz, N. G. Lederman, and Crawford (2004) showed that the NOS and SI can be combined in education.

It is argued that NOS and SI should be taught explicitly and reflectively, since NOS will not necessarily be developed automatically through students’ engagement in scientific activities (N. G. Lederman & J. S. Lederman, 2014). If the purpose of peer assessment is to develop knowledge of NOS and SI, it is important that tasks in the peer assessment are made continuous with the NOS and SI purposes (c.f. Johansson & Wickman, 2011). This means that the students’ evaluation and feedback becomes reification of scientific reasoning and critical thinking in a tentative, empirically-based, subjective, creative and culturally embedded practice. Through the critical review of scientific experiments, students become included in the scientific reviewing process; this ensures the legitimacy of scientific research (Kolsto, 2001; Sandoval & Reiser, 2004).

Peer assessment in science education has mostly been studied with students in tertiary education (Huann-shyang, Zuway, Hsin-Hui, & Sung-Tao, 2011; Nicol, 2009). The focus of tertiary science education studies has mostly concerned what feedback university students provide, or to a much lesser extent, how students use feedback. However, the dialogicity of those processes have mainly been left unstudied (Van Zundert et al., 2010). Research mainly reports positive results from using peer assessment (e.g. Dixon, Hawe, & Parr, 2011; Willis, 2011). Nonetheless, some problems have been observed concerning the divergences of what students assess (Hovardas, Tsivitanidou, & Zacharia, 2014; Poon, McNaught, Lam, & Kwan, 2009; Tal, 2005; Tsai, Lin, & Yuan, 2002). Another problem observed is students’ inability to use the feedback received (E. Brown & Glover, 2006; Jönsson, 2013; Tsai et al., 2002). This can partly be explained through feedback not containing concrete suggestions on how to improve the work (Gamlem & Smith, 2013; Jönsson, 2013). However, students have also been shown to reject concrete suggestions (Hovardas et al., 2014). This indicates that students might not share the same values regarding what counts as quality of scientific work and thus not regard the suggestions as improvements (ibid). This calls for studies regarding how students’ use of feedback relates to the feedback given to other students, and how students can be given opportunities to negotiate what counts as quality of scientific work through peer feedback.

The next research question is:

- What are the affordances and constraints for using peer assessment as a tool for reifying quality and improvement of experimental design?
Affordances and constraints for the expansion of teachers’ assessment repertoire of scientific literacy

As previously mentioned, there are indications that many Swedish science teachers regard communication, source critique and decision-making as new aspects of the science curriculum (Lundqvist & Lidar, 2013). When new aspects are introduced into the syllabi, contradictions may emerge between what was possible to assess through prior assessment repertoires and what becomes a necessity to assess with the new curriculum reform (Lyon, 2013). The Swedish National Agency for Education offers examples of assessment tasks through compulsory national assessments. However, teachers are required to use a variety of evidence for grading. As the national assessments are constrained by demands of standardisation, those tests can only work as support for grading. Alternative solutions to the national assessments are being offered by the Agency so that teachers can collect a richer set of evidence. Such solutions involve models to use the experiences of colleagues for affording the expansion and coordination of the assessment repertoire.

Apart from offering some guiding material, the Agency recommends that teachers continuously discuss assessment with colleagues for moderation and development purposes (Skolverket, 2014). Previous research (Adie, 2012; Adie, Klenowski, & Wyatt-Smith, 2011) on moderation indicates that discussion groups could afford possibilities for teachers to negotiate the expansion and coordination of repertoires for the challenging assessment of communication, source critique and decision-making. Consequently, the third study focusses on assessment moderation meetings as a means for teachers to negotiate assessments concerning decision-making on various civic issues with a scientific content.

Assembling teachers in meetings has been used for the moderation of marking and grading, with positive results in the sense of increased agreement on the quality of students’ performances, that is, inter-rater reliability (Adie et al., 2011; Black, Harrison, Hodgen, Marshall, & Serret, 2011; Klenowski & Adie, 2009; Wyatt-Smith, Klenowski, & Gunn, 2010). Such assessment moderation meetings have been especially important in performance-oriented assessment where quality standards are difficult to articulate (L. Roberts, 1996). However, some problems have been reported with assessment moderation meetings. For example, Bloxham (2009) found that teachers tended to give more average marks when asked to share their assessment with colleagues due to fears of diverging. Additionally, teachers are often restricted to limited evidence when assessing colleagues’ students (L. Roberts, 1996).

Assessment moderation meetings are, nevertheless, argued to be a preferable solution to external moderation though issued tests and comparison with expert raters. The reason is that teachers have more autonomy over their own assessment in the discussions (Cumming & Graham, 2004; Klenowski & Wyatt-Smith, 2010; McMahon, 1999). To be given time to discuss and
compare the assessment of student achievements with colleagues has not only been appreciated (Bolt, 2011), but may also have significant moderation effects. This is the case even if teachers sometimes agree on abandoning external directions (Adie, 2012; Adie et al., 2011). Though the positive effects on inter-rater reliability have been argued extensively, the potential for expanding the assessment repertoire into new challenging areas remains to be studied.

The Swedish teachers’ frustration regarding the assessment of communication, source critique and decision-making could be the reflection of a wider international curriculum reform from the testing recollection of scientific theories to assessing how science is being used by students in different contexts (Lyon, 2013). Many efforts to situate scientific knowledge in societal contexts have ended up emphasising ‘hard scientific knowledge’ that can potentially be applied, rather than assessing how students use multiple resources when making decisions (Zeidler et al., 2005). Zeidler et al. (ibid) criticise a technocratic view of the more informed decision-maker as the person with the most extensive scientific conceptual knowledge. Furthermore, the opportunities for students to express their personal meaning risk becoming constrained by educational demands to perform and adapt (Hasslöf & Malmberg, 2015).

However, it is important to point out that science teachers do not have one discourse for assessment. Rather, they seem to change both the content and form of assessment with their understanding of science in different contexts (F. Edwards, 2013; Magnusson, Krajcik, & Borko, 1999). For example, in education for sustainable environment, Swedish teachers have been described as taking a fact-based, normative or pluralistic approach (Lidar, Lundqvist, & Östman, 2006; Lundqvist, Almqvist, & Östman, 2012; Öhman & Östman, 2007). Teachers have continuously shifted between emphasising that students should express a certain scientific explanation in a specific way, taking a normative stand to preserve nature, or encouraging students to consider multiple aspects and make their own opinions regarding issues concerned with science (ibid). This indicates that teachers can find alternative approaches to SL and potentially develop new assessments for assessing procedures like communicating, critically examining information and making informed decisions.

The following research question is:

- In what ways can assessment moderation meetings afford and constrain the expansion of repertoires for teacher assessment of source critique and decision-making in science education?
Affordances and constraints for authentic assessment of scientific literacy

From a situated practice perspective, participation, and consequently, student performance varies with the contexts. If assessments are used to make assumptions about students’ use of science in a context, the assessments must be as accurate, that is, as authentic to that context as possible (Wiggins, 1990). However, the research literature approaches the ambition of authentic assessment in many ways. Assessment is sometimes stated as being authentic because it reflects practices outside school, for example, the practices of chemists (Schwartz et al., 2004). Sometimes the assessment is claimed to be authentic because it is based on students’ own choices or interests (Swaffield, 2011). Both these perspectives pose challenges for teachers’ assessment of scientific literacy. If authenticity can only be found outside school, then teachers and students need to know where to find such practices and how to bring them into classroom assessments. If authenticity is based solely on the choices students are interested in making, then questions arise as to why only these choices were made and how the students discerned which choices were meaningful to make. Another way is to regard scientific literate participation as being authentic when people are contributing to practices with a scientific content (Roth, van Eijck, Reis, & Hsu, 2008). From this approach, it is argued that authentic assessment has to extend beyond students providing the teacher with evidence of knowledge. Rather, it needs to evaluate the usability of students’ processes and products in the practices of different communities. When the object of the assessed activity is to produce something useful to other people, that participation becomes meaningful not only for the individual students, but also for the communities in which the student becomes included. However, with actual contribution to the practices of different communities comes accountability to those communities (Aikenhead et al., 2011). Such accountability conflicts with the school tradition of affording environments where students can practise without society depending on their production (Carlgren, 1999). Consequently, different theoretical frameworks both afford and constrains ways to discuss authenticity and design authentic assessments. This in turn affords and constricts how the assessments can be used to make statements about students’ abilities to participate in different practices concerned with science.

Specific studies regarding the authenticity of assessment in science education seem to be very rare. B. Bell (2007) wrote a section of authentic assessment in a review of science classroom assessment, based almost entirely on research that did not deal explicitly with science education (Darling-Hammond, Ancess, & Falk, 1995; Cumming & Maxwell, 1999; Wiggins, 1989). B. Bell wrote her review many years ago, but conducting a Google Scholar search on July 15, 2015 using a combination of ‘science’ and ‘authentic assessment’ and publications after 2007 only produced six studies
among the first 100 results that specifically dealt with authentic assessment in science classrooms. Herrington, Reeves and Oliver (2014) payed little attention to assessment; they only argued that assessment should be integrated with the learning exercises to be considered authentic. Yılmaz-Tuzun (2008) mentioned that pre-service teachers who were interviewed named assessments emulating scientists’ activities as being authentic. Kaya (2008; 2009) argued that concept mapping was an authentic assessment tool due to being student-centred. Similarly, Akçay (2009) used journals and notes on laboratory work that were argued to be authentic as they were student-centred activities. Siegel and Wissehr (2011) expressed disappointment that pre-service science teacher students resorted to multiple-choice questions instead of more authentic assessments. Only one of the publications (Kaya, 2008) gave directions on how to design authentic assessments in science; pre- and post-laboratory conceptual maps.

Plenty of publications can be found about research on authentic assessment generally or the authenticity of science education. For example, reviews have been conducted regarding different definitions of authentic assessment (Palm, 2008; Wiggins, 1990) as well as articulations of the critical aspects of authentic assessments (Ashford-Rowe, Herrington, & C. Brown, 2013; Herrington & Herrington, 1998). Since the scientific content is so important in science didactics, it would be imprudent to neglect the specificities of science education in authentic assessment. The authenticity of assessment in science education is not independent of what is considered authentic science education. However, authenticity is frequently used in science education research with very different meanings (Murphy, Lunn, & Jones, 2006; Roth et al., 2008). Therefore, what implications different uses of authenticity have for the affordability and constrictions of designing assessments for SL do not seem to have been thoroughly investigated. The fourth study is a review of the three highest-ranking journals in science education on how different researchers have used the concept of authenticity to describe science education. The implications these uses of authenticity have for designing authentic assessments are then discussed. The last research question is:

- How may authenticity as used in science education research contribute to a discussion about authentic assessment in science education?
Finding a path

The contribution to understanding the affordances and constraints for developing classroom assessment of SL as participation in practice can be approached in different ways. I could choose to be a quiet observer and study the ongoing assessment of communication, source critique and decision-making. However, I have chosen to actively intervene and study the affordances and constraints while working with the teachers in the development of assessment practices. Due to the inexperience expressed by teachers in a survey conducted by the Swedish National Agency for Education (2010), I mainly chose to design my research as intervention studies in which I co-operated with teachers in developing assessments, rather than mappings teachers’ struggles. I chose a Praxis development approach to help teachers expand their own development of assessment practices. Just like in clinical research of medicine, I aimed to obtain results that could be applied directly in teachers’ practices. I also assumed that teachers and students would find participating in the research more meaningful if they were allowed to manage and experience the improvement of their own practice.

Praxis developing research

There are many different research forms in which the researcher can be directly involved in developing an educational practice. The Action research tradition has existed since the 1940s (Lewin, 1944, 1946). In this tradition, researchers work with teachers in iterative cycles where analyses of prior cycles are used to help teachers change their practice in the following cycles (Elliott, 1991). Traditional action research studies preserve much of its original characteristics as emancipatory sociological research in which participants themselves should break free of old problematic habits (Adleman, 2010). In educational action research, students and teachers do not repeat the same lessons in the cycles. Rather, data collection and development focus on the same ‘critical aspects’ judged necessary for change (Carr & Kemmis, 2005; Elliott, 1991). Such a critical aspect could, for instance, involve forming investigative questions during the lesson to elicit students understanding of scientific concepts rather than controlling whether they can remember the correct formulations from earlier lessons.
Similar, but with more specific learning objectives, Lesson study (Stigler & Hiebert, 1999) and Learning study (Marton & Ling, 2007) are methods where teachers and researchers can try and refine a lesson design. Lesson studies can be very diverse and are often performed by groups of teachers wanting to develop their teaching, for example, regarding how to better make students understand the concept of electrical currents. The design and redesign of the lessons are built on experiences and observations, but are not necessarily theory-driven in the sense that the designs are grounded in pedagogy (Fernandez, Cannon, & Chokshi, 2003; Lewis, 2000). Learning study was originally developed to develop variation theory, a theory to explain learning as discerning variations of phenomena (Marton & Pang, 2006). Learning study is more technically theory-driven than lesson study and often involves professional researcher participation (Marton & Ling, 2007).

A much more researcher-led form of praxis developing research is Design Experiment. Design experiments were developed as a form of research in which behavioural scientific experiments were to be conducted in natural environments rather than controlled laboratories (A. L. Brown, 1992; Schoenfeld, 2006). The research is ‘designed’ to, for as long a time as possible, investigate the effects of interventions and collect a broad set of data, for example, both large statistical and close ethnographical data (Collins, Joseph, & Bielaczyc, 2004). However, design research has changed in diverse directions since it was first introduced. The present-day design-based research (DBR) still aims to develop empirically-grounded theories by studying the process of learning and the support of that process (diSessa & Cobb, 2004). DBR has sometimes been misunderstood as an evaluation of teaching models, since researchers often implement new lesson plans. Rather, the aim is to develop theories by applying them in different contexts (The Design-Based Research Collective, 2003). Researchers use a theory on how teaching can be improved and try to understand what happens in the environment of learning from as many angles as possible when this theory is put into practice (Collins, 1990; Collins et al., 2004; Schoenfeld, 2006). Van den Akker, Gravemeijer, McKenney and Nieveen (2006) claim that the essential characteristics of design-based research are that the research is: 1) interventionist 2) process-oriented 3) utility-oriented 4) based on a theory and 5) iterative. Data are analysed from each intervention and used to design the next intervention turn, or phase as they are called. Unlike action research or learning/lesson studies, interventions do not need to be very similar or even conducted with the same teachers and students. Cobb, Confrey, diSessa, Lehrer and Schauble (2003) explain that the main objective is to understand the learning ecology, defined as several different elements interacting on different levels in education.

DBR opens up for studying various aspects of multiple domains and is thus the most suitable of the four traditions for my research questions. Although large-scale quantitative studies were not called for by the research questions, certain elements from designed-based research proved useful. I conducted
three empirical intervention studies from the principles of DBR. My studies were designed with interventions in which assessment tasks were constructed with the purpose of reflecting real out-of-school activities and how participation is valued in those contexts. Rather than measuring the result of the interventions with pre- and post-tests, I studied the processes of negotiating meaning of the assessment. The studies were utility-oriented in that the teachers and I developed tasks that were used directly, and the outcome could be used for further construction of assessment tasks. The first and second research studies were based on the theoretical frameworks of CoP in that: a) development aimed to be based on the shared experiences of students and teachers rather than on externally defined principles of how things should be, b) the repertoire of the classroom was regarded as continuously being negotiated by students and their teachers rather than predefined in that environment. The third research study was based on the theory of expansive learning in that I investigated how teachers could expand their assessment repertoire by uncovering contradictions and negotiating new ideas. Interventions were iterative in the sense that similar interventions were conducted in at least two schools, and experiences from one school could be used to amend the intervention of the other.

The articles produced in the research focussed on different research questions of the entire research project. Therefore, in the articles it did not become necessary to present how the data material was part of answering an overarching research question to make the analyses understandable. The DBR can thus not be found in the article manuscripts. However, in this thesis, it is necessary to describe how the interventions were designed to answer the overarching research question.

Finding teachers for the project

An initial issue was to find teachers willing to develop their assessment practice towards a situated learning approach. Through recommendations from colleagues at the department, an e-mail was sent out to inquire if the teachers would be interested in working with the assessment of SL in civic contexts. The positive replies were followed up with information meetings in their school during which the purpose of the research project was explained. At this stage, three teachers from three different schools and a team of six teachers were interested and had received the approval of their headmasters. The plan was to work with the first and second research questions with the three independent teachers and the third research question with the team of teachers.
Pilot study

I probed for what would be meaningful and possible to work with by designing different tasks with the three teachers that we tried out with their students. One task regarded a textbook’s description of gender roles and reproductive behaviour among animals. Another task was an investigation of changes in the frequency of species and environment of a local lake. The third task was a group discussion regarding ‘family planning’ through history. The team of teachers was asked to discuss the assessment of some PISA items and corresponding student replies as well as how they dealt with the assessment of some aspects in the science subject curricula.

In these pilot studies, it became evident that neither the students nor the team of teachers had developed repertoires for dealing with the tasks or the assessment regarding the use of science in civic contexts. Most students did, however, engage in the negotiation of such repertoires with their peers and their teachers. Furthermore, the prospects for the praxis developing research looked promising. However, one of the teachers in the team of teachers ended his employment, another changed her working tasks, and the remaining four teachers dropped out of the project.

The schools of the study

The three schools remaining here referred to with the fictive names, Apple School, Birch School and Central School, were situated in diverse neighbourhoods. I have chosen to describe the ethnic, socio-economical background as well as academic achievement of the schools (SiRiS/SALSA, 2013), since teachers made an issue of this in the third study.

Apple School is located in a local suburb of Stockholm. The ethnicity and socio-economical background of the students was very diverse (22% of the students were registered as ‘newly arrived’). A total of 68 per cent of the students who graduated in 2013 had passed all of their school subjects. The male teacher who participated in all three phases of the study had approximately ten years of experience teaching science in secondary school. The class was attending school year seven and eight for the duration of the first and second research studies, with students being between 12 and 15 years old (depending on their birthday). In the third research study, four more teachers participated in the study. The headmistress of the school had expressed appreciation for involving her school in the research project and that the praxis developing research activities were needed to reform the assessment practice of the school.

Birch School is located in another local suburb of Stockholm. A total of 89 per cent of the students who graduated that year passed all subjects. The female teacher who participated during all three phases had approximately ten years of experience teaching science in both primary and secondary school.
The three classes participating were attending school year eight and nine during the time of the first and second research studies (13–16 years old). The school had been in contact with the university, expressing an interest in working with research and development projects. In the third study, two more teachers joined the project.

Central School was an inner city school in Stockholm. Eighty per cent of the students received a pass in all school subjects. Two classes participated from school year eight (13–14 years old) in the first research study. The female teacher had five years of experience teaching science. However, she went on maternity leave after the data collection of the first study and did not participate in subsequent studies.

**Studying affordances and constraints for student participation in assessment of scientific literacy**

The first research study for understanding affordances and constraints for developing assessment practices of SL concerned the mutuality of engagement, negotiability of the repertoire, and accountability for the enterprise. I intended to accomplish this by studying what purposes emerged when socio-political debates were introduced in classroom assessment and how students negotiated how to address conflicting purposes. The teachers and I chose to set up the debate as a role-play. Prior research has demonstrated how roles mediate participation and meaning-making of the scientific content (Kolstoe, 2000; Simonneaux, 2001; Ødegaard, 2003). I predicted that assigning roles to students would provide data on students’ negotiability of the repertoire through enactment of the roles and accountability for the enterprise by defending stakeholders’ interests. The roles could potentially provide students with a position from which they could become part of the debate and thus negotiate the repertoire. Making the students accountable for those specific civic interests, for example, clean drinking water, were also addressed. Students were appointed roles from different stakeholders in society whom the teachers and I imagined would have an interest in the issue.

For the sake of students’ mutuality of engagement in the debate, the roles were appointed to pairs or groups of three students. The students could thus prepare for the debates together and were not left to defend a stakeholder on their own. In the first intervention, both classes from Central School participated, and the students were divided into two debates per class. The second intervention was conducted with the class from Apple School, where the students were divided into six debates. The third intervention involved all three classes from Birch School, and the students were divided into two debates per class. Each class was given two lessons to prepare for the debate and one lesson to conduct the debates. The second time I used socio-political debates (Apple School), the
teacher and I introduced a session after each debate where students could provide feedback to the debaters. Our intention was that this would further increase students’ mutuality of engagement and negotiability to the repertoire, as they could highlight and articulate how certain arguments worked better for the stakeholders in the debate. This feedback session was later modified in the last debates (Birch School) into half-time breaks during which students could coach the debaters, who then received a second chance to hone their arguments in the second half of the debates.

Topics of the debates were chosen by the teachers and me from what scientific content the teacher had planned to cover in the lessons before the debates. Consequently, the assessments would not only function as an assessment of general abilities to construct arguments, but the debates were also simultaneously used as an assessment of the scientific content in the context of a civic issue. The debate was first conducted as part of chemistry education in Central School about a claimed decline of fish in a lake. The second debate was conducted in Apple School as part of physics education and involved the purchase of a refrigerator, a pair of ‘thermodynamically designed’ shoes or cars with hybrid engines. The third debate was conducted in Birch School and concerned chemical additives in food and cosmetics. The first debate in Central School provided sufficient material to write a manuscript, and the data from the debates in Apple and Birch School were not used in the article.

In all cases, the students were provided with information with texts taken from authentic reports. The teachers and I chose to make some adjustments of the language in the texts to increase the mutuality of engagement, as we believed the original texts would be too difficult to read. However, in so doing we lost some authenticity to the challenges that citizens actually face when engaging in debates outside school. Students were also allowed to use sources of information that they could find themselves.

For the assessments, we chose to use criteria and standards already developed by Erduran et al. (2004), which we interpreted to correspond well with the less concrete standards of the Swedish syllabi in biology, chemistry and physics. These standards were presented to the students as they received the task, roles and information, so that the students were informed of how the teachers would judge their participation in the debate.

Analyzing the data from the debates

I initially used the concepts of mutuality of engagement, negotiability of the repertoire, and accountability for the enterprise to analyse the data. However, it proved too difficult to communicate this analysis in the compact format of journal articles. Instead, I only used CoP to explain the emic perspective of the analysis in the first article. The research questions were instead answered with a detailed content-oriented analysis, taking into account the dialogicity
at the level of utterances (Linell, 1998), in other words, how the students used and built upon each other’s utterances. Conflicting emergent purposes were analysed in episodes where students changed (negotiated) their argumentation (purpose) after facing opposition (conflict). The data selected for the first article have also been discussed in seminars of my own department. The identified purposes and conflicts were compared to other identified conflicts of the same and other categories. Through continuous comparison, characteristics of the categories were redefined and restructured until the authors agreed on the robustness of the categories (Corbin & Strauss, 2008).

Studying affordances and constraints for communicating assessment of scientific literacy

The second research study concerned the negotiation of meaning of quality as participation and reification. I studied this through peer assessment designed as a learning tool for developing the scrutiny of experimental work. This became a complex design for the teachers and me; we had to combine the educational goals of developing NOS and SI with the research goals, making students’ negotiation of meaning of what counted as good experimental work observable. Research in NOS and SI argues for teaching NOS and SI explicitly (N. G. Lederman & J. S. Lederman, 2014). Thus, the teachers and I decided to present the educational goals of NOS and SI through a historical research example. We then introduced the experimental design task students were about to investigate and assess. The experimental task concerned the connection between breakfast and physical prowess in the morning. Topics about health and fitness have been reported to engage students (Jidesjö, Oscarsson, Karlsson, & Strömdahl, 2012).

Students received the assessment format from open-access national assessments for biology, chemistry and physics. However, using peer assessment as a learning tool involves negotiating the reification of quality of the designs in relation to experiences of having participated in conducting and reviewing the experiment. It is consequently important to observe how students reify experiences of having eaten a breakfast and performed a physical exercise (participation) as quality of the experiments. By asking the students to provide written feedback to other students, such reification could become manifest in their comments. To make this more explicit, the teachers and I asked the students specifically to give suggestions on how to improve the design and not merely point out what was functional or problematic. Reviews of peer assessment have asked for research studying students as both the providers and receivers of feedback (Jönsson, 2013; Van Zundert et al., 2010). Consequently, I decided to collect both written design and written feedback from all the students so that I could compare the given and received
To be able to collect data from student negotiations, the teachers and I had the students sit in small groups of 3–7 students where they could discuss the usability of the feedback received. These group discussions were audiotaped. In Apple School, which conducted the peer assessment first, the students were not required to alter their experimental design (since the teacher estimated this would put unwanted strain on the students). However, E. Brown and Glover (2006) have discovered that students can express their appreciation of feedback, and still not use it to amend their work. Thus, when Birch School did the peer assessment, I felt it was necessary to collect data on what feedback students ultimately chose to revise. The students were also required to submit a revised version of their experimental design.

Analysing the data from the peer assessments

Analysis of the data was conducted in three steps. The first analysis looked at different types of suggestions the students provided in the feedback. The second analysis studied how students negotiated the usability of the received feedback. The third analysis studied how the changes that students made to their design were linked to the group discussions. I used the NVivo computer software to link students’ written text with the audio recordings. Thus, when listening to a discussion, the students’ provided and received feedback as well as original and revised design could be accessed directly. The CoP concept of participation was useful for describing how students positioned themselves (e.g. as researcher or teenager). The concept of reification was also useful for describing what students articulated as the quality of the experimental design.

Studying affordances and constraints for the expansion of teachers’ assessment repertoire of scientific literacy

The teachers in both schools wanted to select the scientific content of the assessments that they were planning to cover in the lesson prior to the third intervention study. The moderation meetings were held at each school separately. Since the teachers were inexperienced with the assessment of source critique and decision-making, they asked for help with constructing the assessments. I found civic contexts that could be used for source critique and decision-making. The teachers defined what they wanted to assess regarding those abilities. They also defined the appropriate tasks for assessing their students. I predicted that contradictions between the historicity of assessment and the new syllabi requirements would emerge in the construction process due to the novelty of designing assessments for these abilities. I therefore started to collect data during the construction of the tasks.
I used two methods for designing the assessment moderation meetings. First, each teacher used the same assessment task as their colleagues, with one of his/her own classes and joined the discussion group to discuss his/her individual assessment of the students’ performances (Adie et al., 2011). This model was initially chosen so that each teacher could use a wide variety of evidence gained from being present when students were participating in the assessment task (L. Roberts, 1996). However, during this meeting, teachers at Apple School spent a lot of time defending why their students had not performed as well as usual, rather than helping each other to assess the student material brought to the meeting. I therefore changed the final meetings so that only one teacher from each school, who felt comfortable displaying their students’ work to their colleagues, was asked to conduct the task with their students and bring the students’ work back to the group for collaborative assessment. Thus, the group of teachers discussed the work of the same group of students (Adie et al., 2011). By altering the design, I also changed the parameters for the discussions. This had to be taken into account when analysing the expansive moves the teachers made during the meetings.

Analysing the data from the assessment moderation meetings
Episodes in the transcribed data were digitally coded and labelled with the Transana computer software, using teachers’ own words for describing the historicity, contradictions and suggestions for expansive moves. This digital labelling system made it easy to access and compare different episodes in the meetings, for instance, finding all the episodes in which teachers discussed ‘scientific concepts’.

Studying authenticity in science education research
The fourth research study used a model for conducting systematic reviews developed by Bennett, Hogarth, Lubben, Campbell and Robinson (2010). The model consists of seven steps: (1) identification of the review topic, (2) development of inclusion and exclusion criteria, (3) systematic searches in databases and other sources, (4) screening relevant studies against criteria, (5) coding against agreed characteristics, (6) providing a systematic map of the review area and (7) in-depth review of the selected studies in relation to the research questions. The method is developed as a result of prior critiques against more narrative reviews being more at the discretion of the authors, with an increased risk of bias (Bennett, Lubben, Hogarth, & Campbell, 2005). The only inclusion criterion was that the journal article should contain some form of the word ‘authentic’. The review was limited to the three highest-ranking journals in science education according to SCImago Journal and Country Ranking (SJR). Coding was primarily based on the journal authors’
own vocabulary. That is, a word could have divergent meanings in different articles, and different articles could use different words to describe similar empirical events. The review was used to discuss the implications different uses of authenticity in science education had for designing authentic assessment of science. For this I borrowed concepts from Murphy, Lunn and Jones (2006), who operationalised authenticity into personal and cultural authenticity. Cultural authenticity deals with science learning as participation in an activity of a social context, where understanding the nature of the scientific context and skills, like critical evaluative abilities, is needed to participate in these activities. Personal authenticity is the individuals’ meaning-making and evaluation of the intended learning objects. For a task to be considered personally authentic, Murphy et al. (ibid) argue that it has to be perceived as relevant by the students when engaging in activities relating to science and society.

Methodological considerations

My choice of methods implies certain limitations for validity and reliability of the results. All studies involve some kind of categorisation or grouping. I have tried to keep the coding transparent and close to the students’ and teachers’ own utterances. Nonetheless, all coding involved some interpretation and thereby a risk for having misrepresented the informants’ actions. My supervisor has verified the coding, and other researchers have also verified some of the data material. However, my supervisor and most colleagues share research traditions and are used to approaching data in similar ways. Therefore, it is possible that a larger group with more varied research experience would have categorised the material in a slightly different way. To make the research as transparent as possible, I have tried to include as many quotations and excerpts of students’ work as were practically possible. All data material needed translation due to being in Swedish, but care has been taken to keep the original sentences as much as possible (including preserving incorrect grammar).

The situated and emic approach limits the generalisability of the results since the analyses become a narrow description of particular events where the activities are unique to the specific scientific content and people involved. Changing the topics of the tasks or the composition of the groups would have resulted in completely different data. However, as Biesta (2007) explains, people do not respond instrumentally to education. The purpose of qualitative educational research is, thus, not to produce general rules for how optimal assessment should be designed.

Another methodological consideration concerns my participatory involvement as a researcher in my praxis-developing approach. The purpose of praxis-developing research is to facilitate the development of practices in a
certain direction, and in DBR also the development of the theory used. Through this participation, the researcher affects the outcome of the result. However, the purpose of praxis-developing research is not to describe how conditions are, but rather to investigate the potentiality for conditions to change (Carlgren, 2005). As theory-driven as this study is, I consider it important to abductively use the interventions to observe new and unexpected aspects of the theories when applied to the assessment of SL (c.f. Stiles, 2009). Considerations concerning myself as a researcher affecting the results are grounded in what claims I aim to make about the generalisability of the research results. It is impossible to speculate whether the observations would be more transferable to other classroom assessments if I had left teachers and students to find ‘their own’ ways. However, the applicative value of my research lies in the reader’s ability to find many similarities between the contexts described in my research and other classroom contexts they will encounter. I have thus tried to give thick descriptions of students’ and teachers’ participation and leave the judgment of generalisability to the reader (Larsson, 2009). Nonetheless, researchers in interventions must take care not to direct or control the students’ and teachers’ actions. I must also point out that my research was not designed to investigate to what extent teachers could follow pre-constructed assessment models. Such an approach would conflict with my theoretical framework; the practices of a community lie in what happens when the members participate. From an emic perspective, affordances and constraints for developing practices cannot be studied as comparisons between what teachers did and what they should do. Rather, they are understood through what teachers and students negotiate together.

Furthermore, when present in the classroom during the data collection, I heard and witnessed discussions between students and teachers that I did not understand. I then asked the students and teachers to explain what they meant. Corbin and Strauss (2008) point to the difficulty of analysing people’s behaviour without asking them to explain it. Since this was an intervention study where I already had intervened in the design of the assessments, I saw no major problem with my presence. However, when as a researcher I asked questions about an utterance, these questions might draw attention to that utterance in the discussion and thus influence the outcome.

Ethical considerations

The research was designed from the ethical principles from the Swedish Research Council (Vetenskapsrådet). Students and teachers were informed in person of the purpose of the research project and their role in it. A letter was sent home to all students describing the project and asking for written parental consent. The letter also informed the individuals about their right to withdraw from participation at any time and that their anonymity would be guaranteed.
in publications and presentations. The participants were informed that the data material would be used for anonymised scientific publications and presentations. In one school (Birch School), most of the students and parents consented to the material also being used for other educational purposes than teacher assessment.

The right to withdraw from the research project put special conditions on the research since it concerned assessment and marking. It was important to inform the students in every intervention that if they did not want to participate in the assessments, they would be offered other opportunities to display knowledge to the teacher. However, all students agreed to participate.

Anonymity also created a dilemma. Teachers were very active in designing the research setting, creating tasks and criteria as well as using results from the interventions to suggest improvements for future interventions. The insurance of anonymity did, however, prevent me from crediting their contributions by name.

Care has also been taken not to expose any students or teachers as ‘bad examples’ and to ensure that all activities displayed in the articles are properly set in a context not to be misinterpreted.
Presenting and discussing the results

Article 1: Affordances and Constraints for Using the Socio-Political Debate for Authentic Summative Assessment

The first article investigates affordances and constraints for using socio-political debates for assessment. The investigation draws from a known trade-off between validity and reliability in assessment becoming apparent when redefining those concepts in authentic assessment (Moss, 2007; 2003). The more one increases validity by making the assessment more similar to the activity the assessment is aiming to measure, the less reliable student participation and teacher judgment become. The first debates in Central School offered sufficient material, and the analysis presented in the article was only based on data from the debates in which students were given scientific reports about a lake, introduced to an issue and appointed character roles to represent stakeholders of the lake. In the debate, new purposes emerged due to the authenticity of the task; this conflicted with the purposes stated in the predefined criteria. The debate not only became a discussion about what conclusions could be drawn from the scientific reports, but students’ dialogues also exposed a negotiation on how to address conflicting purposes emerging from sustaining participation authentic to the context of the debate. Students had to negotiate how to address the emergent purposes of:

I) Putting scientific knowledge on display versus staying true to one’s role
II) Putting scientific knowledge on display versus expressing social responsibility
III) Putting scientific knowledge on display versus ‘winning’ the debate
IV) Using sources tactically versus using sources critically

The emergent purposes of staying true to one’s role, expressing social responsibility, ‘winning’ the debate, and using sources tactically constricted the students from putting scientific knowledge and source critical knowledge on display for teacher assessment. However, the socio-political debate enabled the teacher to assess how students made conclusions about the scientific
material in relation to the environmental issue of the debate and the concerned stakeholders they were representing. Much evidence of how students constructed and modified arguments while dealing with opposition and support were made available to the assessing teacher. However, with increased negotiability of the repertoire followed increased accountability to the enterprise through collective decisions about what topics to discuss. The teacher could thus neither ensure that all topics the teacher desired would be covered in depth nor that all students would say something regarding every topic discussed.

Students were consequently able to use their roles and their stakeholders’ interests to discern what were relevant conclusions to draw and what were efficient ways to compose and counter arguments. It is very likely that students, in the same way, were constrained by their roles from bringing topics to the table that they personally believed were important for the debate. One girl did express this in the last debate in Birch School. The practices of engaging in socio-political debates proved to be complex; they involved the purposes of expressing social responsibility and ‘winning the debate’, which were not stated as issues the teacher would assess. Thus, in light of the purpose of the teacher’s summative assessment, these emerging purposes, although negotiated by the students, potentially interfered with the predefined criteria of the assessment task. The students were constantly challenged to address the tensions between conflicting purposes. In addition, the failure to quickly adapt to the fluctuations of the debate could result in students becoming less included. For instance, students who had prepared arguments from the predefined criteria might be constrained from defending their stakeholder in the debate due to being unable to follow the negotiation and adjust their arguments to the discussion. This finding is in line with previous research, pointing to the importance of taking into account the social dynamics of role-play and how students’ opinions are influenced by the debate setting (Kolstoe, 2000; Simonneaux, 2001; Ødegaard, 2003). The salient purpose of a debate is to persuade other participants. Thus, there is always a risk that stakeholders might resign from the discussion if their arguments lose influence in the debate.

Through the CoP framework, I focussed mainly on the socio-political debates as a group endeavour. Mutuality of engagement, negotiability of the repertoire and accountability for the enterprise were understood through the group interactions rather than as characteristics of individual student’s participation. In this study, mutuality, negotiability and accountability were used similarly to Mogensen and Schnack’s (2010) assessment criteria to understand the affordances and constraints for students’ participation and not as an assessment of the students’ performances. Students were constantly exchanging ideas (in fact, the debates were difficult to stop). However, some students were more included in the debates than others. Therefore, they had more influence on the direction of the debate and made their stakeholders’
interest important to the discussion. Though it was not possible to determine if or why students were holding back when they were quiet, it became possible to identify which students needed support when they spoke. It became obvious that students sometimes needed help learning how to interpret the scientific facts and concepts from the reports and how to use them correctly and productively in the discussions. In line with prior research on argumentation in science education, the debates afforded special conditions for the instant response from other students. This is difficult to offer in written argumentation examinations (Säljö, Mäkitalo, & Jakobsson, 2011). The quality of students’ arguments also frequently became clear to the students as their discussions made apparent what scientific conclusions became advantageous and disadvantageous to the different stakeholders. Consequently, the debate afforded opportunities for assessing the quality of students’ arguments in action, how students respond to one another, and the arguments put forward in the debate (Åberg, Mäkitalo, & Säljö, 2010). The emergent purposes of ‘staying true to one’s role’, ‘expressing social responsibility’ and ‘winning the debate’ indicate that the students authenticated their engagement to represent the interests of stakeholders such as fishermen and farmers beyond what was required from the criteria.

A major question is if the affordances of assessing students in interactions outweigh the constraints that are involved with activities in which students’ affordability heavily depends on the actions of other students. One expressed purpose for using alternative assessments to standardised tests is to collect evidence of how well the communication of conclusions, source critique and construction of arguments is negotiated in groups (T. D. Sadler & Zeidler, 2009). If one separates students in performance assessment, one reduces the opportunities for these negotiations. Alternatively, group project work can be used to assess interactions, since group projects are conducted as collaboration rather than competition. Project work reports will provide the result of a negotiation. However, to assess the process, the teacher has to observe how students work and discuss (Boud, Cohen, & J. Sampson, 1999). Video recordings can be used to recapture and discuss the events with students (Tan & Towndrow, 2009). However, the individual marking of such group discussion still requires the same teacher attention as the assessment of debates. Using peer assessment to complement what the teacher can observe has been suggested in the assessment of collaborative work (Cheng & Warren, 2000). Nonetheless, the unfairness of group dynamics cannot be completely avoided if the negotiation process is included in the assessment.

There is a question of whether the openness of the debate format involved places too much accountability on the students. For example, the teachers could have directed the debates and ensured that all students were given equal opportunities (e.g. in time slots) to develop their standpoints and interpretations of the scientific reports. Furthermore, teachers could have ensured that students stayed on the scientific topic rather than focussing on
Hasslöf, Ekborg and Malmberg (2014) argue for affording student development and appreciation of the multivoicedness in the debate. However, if a socio-political debate is to become authentic, students need to be given the creative freedom and accountability to decide on the focus of the discussion (Åberg et al., 2010). One advantage of student influence on the socio-political debates is that this influence affords developing scientific knowledge in contexts that are personally authentic to them. Consequently, stating obligations for students to display descriptions of predefined scientific concepts may restrict students from discussing what becomes meaningful to them in the particular debate. The assessment criteria and standards used for assessing student participation in a debate need to be constructed to reflect what will actually be valued in the socio-political debate. The indicators for action competence that Mogensen and Schnack (2010) suggest are that students are encouraged to: a) work with power relations and conflicting interests, b) consider different perspectives and identify themselves with others, c) present different arguments and d) explore alternative actions. I would evaluate all these indicators as salient in the socio-political debate. Students are likely to find themselves in situations in which they somehow represent the interest of a stakeholder, where they have to challenge rhetorical strategies or evaluate the importance of science in a context and the usefulness of sources. Consequently, it is advisable to inform students that they are assessed on how they mutually engage, negotiate a repertoire and deal with their accountability for these challenges. By studying and discussing conflicts between emergent purposes of the debate, students and teachers can better prepare for such conflicts. This could be observed when the students frequently used suggestions from peers to improve their argumentation when the coaching element was added in Apple School and Birch School. Moreover, it is recommended to allow students to work with a variety of contexts, engage in a multitude of debates, and offer possibilities to try participation from diverse stakeholders (Ratcliffe & Grace, 2003). Such alternations involve exploring varying contexts and roles, and they show how the different characters afford different kinds of participation. Engaging more than one assessing teacher would also ensure that more evidence was gathered from all the students participating in the debate.
Article 2: Using and Rejecting Peer Feedback in the Science Classroom: A Study of Students’ Negotiations on How to Use Peer Feedback When Designing Experiments

The second article discusses the affordances and constraints for using peer assessment as a learning tool for developing critical experimental design. The study draws on a stated problem that students are seldom utilising peer feedback. This might be due to diverging values of quality (Tsivitanidou, Zacharia, & Hovardas, 2011). In light of the controversy in prior research, a research design was set up to study the dialogicity of the process of both giving and receiving feedback in order to understand what afforded and constrained the students’ use of feedback. Data materials were used from both the Apple and Birch Schools’ tasks to design an experiment concerning breakfast and exercise and exchange experiment designs with each other. After conducting the experiment, the students could draw on experiences when providing feedback to the peer who designed the experiment. The discussion groups that the students formed when discussing received feedback provided data on how students negotiated the usability of the feedback. By connecting written student work, feedback and the audiotaped discussions, it became possible to investigate upon which resources student drew when altering their design.

First, it became apparent that students gave feedback that was reified into three different aspects of quality:

A) Concerning personal experiences and preferences
B) Concerning scientific inquiry (SI)
C) Concerning confirmation of knowledge about nutrients and health

Whereas students almost always used feedback of type B, only half of the feedback of type A and C was used to make changes in the design. Students used feedback they had given to others as often as feedback they had received when making changes. The study confirmed what had been implicated in prior research: students were constrained in their use of feedback that diverged from the feedback they had offered themselves. However, by following students’ discussions, it became clear that the group not only afforded support for students’ decision to reject, but also convinced peers to use received feedback. Feedback that was not used to make changes was still used in the discussion group to afford reification of what counted as good experimental design and what did not.

First, the specific scientific content became very important for how I understood the peer assessment processes in the second study. For instance, the three categories of feedback provided by the students were closely
connected to the particular experimental task. These categories and the distribution between them are likely to have differed if the students, for example, had been given a task to design experiments concerning lightbulbs in serial and parallel electrical circuits. The study results should therefore not be used for drawing conclusions about what students generally regard as quality of experimental work. The different categories of feedback, however, bear witness to different reifications of quality for the specific experiment of investigating physical exercise and breakfasts. Suggestions concerning personal experiences and preferences reified quality of the experiment as addressing the participation of citizens like the students themselves. Experiments about the effect of a breakfast that a person does not like, or an exercise that the person does not enjoy doing, are not very relevant to that person. Suggestions concerning SI reified quality from participation in a scientific community. Controlling variables are important for the credibility of the research results. Suggestions concerning the confirmation of knowledge about nutrients and health were not surprising. Prior studies have presented that practical work is commonly used in science class to confirm what should already be known to the students (Högström, 2009; Högström et al., 2010). It is therefore not unexpected that students reify quality of the experiments from their experiences of such activities.

Discrepancies between the assessments performed by science teachers and students as well as between the students are well known (Hovardas et al., 2014; Poon et al., 2009; Tal, 2005; Tsai et al., 2002). These kinds of discrepancies in what is valued as quality have been proposed as one reason for students rejecting peer feedback (Tsivitanidou et al., 2011). Since students articulated why they found feedback usable or not, the addition of discussion groups proved significant for explaining this. Data from the students’ group discussions illuminated how students negotiated quality. The discussion helped the students to articulate what counted as good quality of the experiment. These reifications could be rearticulated through negotiation with the other group members. The quantitative distribution of how the students used the feedback might have differed if the students had worked individually through the entire process. However, these interactions seem to be an important aspect that has been neglected in prior peer assessment research (Jönsson, 2013; Van Zundert et al., 2010). It is possible that this is the reason why previous research has not been able to explain students’ choices on how to use feedback (E. Brown & Glover, 2006).

The positive effects of giving feedback have been argued in prior research (Nicol, 2009; Nicol & Macfarlane-Dick, 2006). Nonetheless, this study concretely shows how students used the suggestions given to others to improve their own laboratory design. In language classes, students have been observed to copy strategies from other students’ texts (Lundstrom & Baker, 2009; Min, 2005). In this case, students actually used the suggestions they had given to others. The second study shows that given feedback that was not used
by the receiving students could thus still be used by the students who provided the feedback. It would have been interesting to study whether students had discussed the quality of their own feedback with their peers. However, students hardly made any references to their own suggestions. Students were not asked to discuss the given feedback, and an improvement for future peer assessment studies could be to investigate how students negotiate the quality of given feedback in relation to each other.

Students’ peer feedback and their discussion about feedback also provide the teacher with information about students’ meaning-making of quality criteria and the purpose of a lesson activity (Black, Harrison, C. Lee, Marshall, & William, 2003). In this intervention, the written peer feedback provides information on different purposes for participating in the experimental work, and what values of good experimental design the students discern. A teacher may use this kind of information to initiate discussions on both the purposes of scientific experiments and different aspects of quality (Jönsson, 2013). It also became apparent in the group discussion that received feedback which was not used for altering the design could still be used by the students to reify what counted as good quality of experimental design and feedback on this.

Finally, it is worth mentioning that although students frequently discussed the SI aspects of the reviewed experiments, they rarely connected this to the NOS of experiments in general. Students often participated in the discussions from the position of scientists designing, conducting and reviewing experiments. Nonetheless, they distinguished their work from the activities of ‘real professors’.

Article 3: Challenging and Expanding Science Teachers’ Assessment Repertoires Through Social Moderation

The third article follows two groups of science teachers taking part in assessment moderation meetings about source critique and decision-making. The study design drew on a problem expressed by Swedish science teachers, of not receiving sufficient support when assessing some aspects of the knowledge expressed in the national syllabi (Lundqvist & Lidar, 2013). The idea was that the assessment moderation meeting would provide an arena in which contradictive aspects of the assessment could become apparent to the teachers and enable them to invent and reform assessment repertoires for challenging aspects of the syllabi. The research design was set up to observe what contradictions emerged between old assessment practices and new demands and how teachers negotiated how to expand their practices. Teachers’ expansions of the assessment repertoire were observed between the task construction meeting and the two following assessment moderation
meetings. In the first moderation meeting, teachers assessed a task with their own students and brought the material to the meeting for discussion. In the second meeting, one teacher from each school brought student work to the meeting for collective assessment.

The assessment discussions in the two schools differed. The Apple School teachers were always considering that the students had difficulties with the Swedish language and needed support. Contrariwise, the teachers at Birch School put demands on their students to express themselves clearly and include desired information, so that the teachers could understand how students had made their decisions. The steps taken by the teachers to address problems with students’ performances or with the assessments also took different directions in the two schools. Apple School suspected that students risked exposing themselves as ignorant for exploring different options. Thus, its participants chose to change the task format, from asking for the students’ decisions to asking the students to describe three alternative decisions. They also chose to decrease the emphasis on propositional knowledge and practise procedural skills of selecting information and making informed decisions. Contrariwise, Birch School chose to increase emphasis on assessing students’ propositional knowledge of scientific concepts since the teachers judged that the procedural accounts students had made often lacked scientific content. They furthermore argued that students were used to providing their replies to a teacher who already knew the theories behind the replies. The teachers suspected that the students’ decisions were not warranted enough, since the assessing teachers were an already informed audience. Hence, the teachers chose to give the students a real audience beside the assessing teachers so that students would be motivated to describe why some decisions were less preferable than the ones they had made. Three teachers at Apple School did not attend the final moderation meeting. Therefore, I was unable to discern if those teachers had changed their assessment after the first meeting. Some of the absent teachers’ arguments in the first two meetings had been challenged. In addition, the remaining two teachers expressed concerns that their colleagues were unlikely to change their assessment. This indicates that there is reason to question to what extent the absent teachers were included in the expansion. The assessment moderation meetings afforded collective alterations of the assessment as long as the teachers discussed the assessment from the same reification of knowledge. However, it was not possible to determine how the meetings afforded and constrained teachers in overcoming their conflicting views of what counted as scientific knowledge during the studied intervention.

The assessment moderation meetings afforded the development of repertoires for assessing source critique and decision-making seen as expansive moves during the course of the meetings. Although essentially discussing the same contradictions concerning new desires for evidence of the students’ decision-making process and how to deal with the assessment of
procedural knowledge, the discussion in the two schools resulted in completely different expansive moves. Prior research has shown that what teachers agree upon in moderation meetings is difficult to predict, even with clear criteria and standards (Adie, 2012; Adie et al., 2011). In this study, the teachers were dealing with an unaccustomed aspect of the syllabi. Towndrow, A. L. Tan, Yung and Cohen (2010) compared the responses of teachers in Hong Kong and Singapore to central direction. They found that teachers in Singapore discussed the possibilities for implementing new direction, whereas the teachers in Hong Kong took a more critical position. Significant for teacher discourse in Towndrow et al.’s study was whether the teachers felt obliged to follow directions or if they sensed that they were free to negotiate what was best for their students. This could be a possible explanation for why three teachers at Apple School did not participate in the last meeting. Another possible explanation could relate to Bloxham’s (2009), Klenowski and Adie’s (2009) discoveries. The teachers may have adjusted their markings during the meetings to comply with the general discourse, but not necessarily have shared the view of source critique and decision-making as processes needing special attention.

The assessment moderation meetings shed light on some differences between teachers’ interpretation of the syllabi standards. Nonetheless, the teachers were constrained by their ability to discern conflicting views of knowledge. Several utterances described source critique and decision-making as an application and explanation of scientific concepts. This concurs with other studies showing that science teachers tend to emphasise scientific hard facts when situating assessment in societal contexts (Zeidler et al., 2005). If knowledge was regarded as possessions and those possessions were not transferred to another context, there was reason to question whether the students really possessed the knowledge the teachers thought. In the discussion, this problem was addressed by regarding transfer as either unproblematic or problematic. Regarding transfer as unproblematic admitted a focus on students’ use of scientific concepts in the assessments. Regarding the transfer of knowledge as problematic required scaffolding from the teachers. Alternatively, a suggestion given by one teacher was to avoid assessment involving transfer and instead only assess recollection. The results from the study point to a constraint of moderation meetings: if the teachers could not recognise the contradictions, they could not negotiate a need for change. Langemeyer (2006) pointed out this shortcoming of the expansive learning framework.

Furthermore, though the moderation meetings afforded the exchange of ideas, that sharing was constrained by what part of their assessments teachers wanted to expose to their colleagues. The teachers sometimes faced opposition and requirements to change their opinion when making a statement about how a student’s performance could be valued or how to change something from a collective assessment. Klenowski and Adie (2009) argue that exposing
assessment to colleagues in moderation meetings could be perceived as threatening to teachers and their positions in the teaching community as ‘good teachers’. For example, when teachers in this study experienced that students who had performed well on prior tests did not perform well on the tasks assessed in the moderation meetings, either the prior or the new assessment format was questioned. However, this exposure is essential to make contradictions between the old and the new apparent so that expensive moves can be negotiated.

Prior studies on assessment moderation meetings have mainly investigated how teachers use marking manuals for specific tasks (Adie et al., 2011; Wyatt-Smith et al., 2010). In this study, teachers had no such marking manuals and had to reify general syllabi texts into the concrete articulation of quality for specific assessments. This opens up for more diverging interpretations and misunderstandings; there was still an option to avoid assessment of the challenging aspects of the syllabi and emphasise the familiar assessment of conceptual knowledge. There is therefore a further need to study how assessment moderation meetings can be designed to afford the recognition of diverging views of knowledge and how to negotiate these divergences.

**Article 4: In pursuit of authenticity in assessment of scientific literacy**

The fourth article is a systematic review of how authenticity is used in science education research and a discussion of the implications these uses have for designing authentic classroom assessments. The article argues the importance of the scientific content for designing authentic assessments. What is considered authentic science education is therefore essential for how to understand assessment of scientific knowledge as authentic. All articles from the three highest-ranking journals in science education published in 2013 and 2014 were scanned for what the authors referred to when claiming science education was authentic. The authors described science education as being authentic as:

- Comparable with practices of professional scientists
- Grounded in the world of students
- Involving inquiry practices
- Contributing to out-of-school practices
- Involving a challenge of transferring knowledge
- Comparable with practices of citizens (other than scientists)
- In alignment with curricula and stated purposes
- Pedagogy of caring between students and teachers
Based on the findings, the concepts of cultural and personal authenticity were used to suggest important aspects involved with designing science classroom assessments authentic to the different references. As a conclusion, the article suggests a strategy for balancing assessment between cultural and personal authenticity.

Cultural and personal authenticity (Murphy et al., 2006) proved useful for describing different implications for assessment in science education. When authenticity was described as comparable with practices of professional scientists, practices of citizens (other than scientists), or as in alignment with curricula and stated purposes, the authenticity was described as comparisons with something defined outside the classroom. There was thus a strong emphasis on cultural assessment, that is, what students are required to do in the scientific activity. Culturally authentic assessment should thus be designed to afford participation in accordance with externally reified criteria. Furthermore, the students’ performances are assessed from what is usually valued by scientists, other citizens or the curricula.

Contrariwise, describing science education as authentic, meaning grounded in the world of students, as well as involving a challenge of transferring knowledge involving a pedagogy of caring between students and teachers stressed personal authenticity. All of the references, in some way, concerned what the students did in science. Before students engaged in the science, there was nothing through which authenticity could be evaluated. Designing personally authentic assessment becomes difficult because personal authenticity is evaluated through how meaningful and challenging engagement in the assessments becomes to the students and is consequently difficult to predict beforehand. Therefore, personally authentic assessment designs involve ambitions to avoid students’ performances to become artificial, in the sense that students are pretending for the sake of the task. Yet, since there could not be authentic participation before students have engaged in the task, circumventing these scenarios becomes highly problematic.

Assessments emphasising cultural authenticity might not be very personally authentic. For instance, drawing nearer to the cultural authenticity of scientists’ practices might distance the assessments from issues that may be personally authentic to students. The greatest challenges are therefore put on classroom assessments aiming to be both culturally and personally authentic, such as when authenticity is referring to contributions to out-of-school practices or involving inquiry. Teachers do not have the same opportunities to try out and refine assessment items like in the construction of high-stake tests. Consequently, teachers have to construct assessments after some model of what culturally authentic participation usually comprises. They also have to simultaneously predict how the assessments can be balanced so the students will participate in personally authentic ways. One strategy to deal with this dualism is to use a variety of alternative assessments authentic to a variety of contexts and communities. Among the different assessments, students can
explore and try different practices. The variety of assessments can mediate negotiation between personally authentic identity and culturally authentic membership in different communities using science (Jackson & Seiler, 2013), personally authentic cognitive development, and culturally authentic repertoires of scientific inquiry practices (Slavin, Lake, Hanley, & Thurston, 2014), as well as personally authentic self-expression and culturally authentic contributions to students’ communities (Birmingham & Calabrese Barton, 2014; Price & H. C. Lee, 2013).
Concluding the studies

Each of the four studies contributes to the understanding of affordances and constraints for participation, communication, expansion and authenticity in the assessment of SL. When connecting the four studies, it is important to remember that from a CoP perspective, all learning, and thus the assessment of SL, is primarily viewed as the development of identity and modes of belonging:

…we need to think about education not merely in terms of an initial period of socialization into a culture, but more fundamentally in terms of rhythms by which communities and individuals continually renew themselves. (Wenger, 1998, p.263)

This means that in the studies, the interventions became processes of learning how to negotiate participation and reification in various practices for the students and the teachers. The engagement was too temporally bound and short to claim that the assessments afforded learning specific practices of, for example, those of a CEO or a scientific peer reviewer. However, the conditions set for the studies were that the assessment of SL would be on how to negotiate participation and reification and thus membership in different communities. Though situated, such negotiations become part of a lifelong learning of negotiating participation in the many different practices in which students will engage both inside and outside the classroom (Wenger, 1998). Through negotiation, people change who they are and where they belong. Assessment for learning negotiation thus not only becomes formative, but also transformative assessment. I will conclude this thesis by discussing the implications for assessing students’ negotiation of participation in civic practices.

The designed and emergent of assessment of scientific literacy

In the framing of the research questions, I discussed the importance of the scientific content in the European didactics tradition (c.f. Wickman, 2014). The didactic questions from a CoP perspective are: What content to reify, when and in relation to what practice (c.f. Wenger, 1998)? Multiple policy
documents and other publications present lists of abilities and skills to be developed as part of SL (D. A. Roberts & Bybee, 2014). Those abilities can then be connected to different scientific domains such as ecology, nuclear physics or environmental chemistry. Designing the assessments of SL as participation in practices is more complicated than reducing them into lists (though such lists might be helpful in assessment design). My studies show that there are always unpredicted emergent aspects of assessments that become important in practice. I do not, however, argue for merely presenting a scientific topic to the students and observing what happens. I have argued for in this thesis that it is just as important to coordinate participation around reification as grounding reification on participation. In the socio-political debate, the students had a model for argumentation and scientific reports on which to base their argumentations. However, they had to negotiate them in relation to participation as different roles. In the peer assessments, students had a model for designing experiments, but they had to negotiate the use of the model in relation to participation in the review process. In the assessment moderation meeting, teachers had national syllabi. Nonetheless, they had to negotiate them in relation to what was possible to discern in students’ participation. An initial didactic reflection in the design of SL assessments then concerns what kind of participation would afford the learning of a desired scientific subject matter. Participation could be centred on a socio-scientific issue. However, this does not necessarily involve participation in different practices concerned with the issue. The assessment of SL as participation in practices also means reflecting on different stakeholders’ interests in that issue and their purposes for engaging in negotiation of the issue. For example, what interests do farmers have in the ecological balance of a lake? The following didactic reflection would then be how to present a task that affords students’ negotiation of those interests in relation to other stakeholders. However, this also involves reflecting on what constrains students from participating in the assessment. Due to unforeseen emergent aspects of participation and reification, there is a need for flexibility in the task concerning student negotiability.

In this thesis I will give a concrete example on how this works using the socio-political debate about the lake. The scientific reports given to the students contained data from measurements on pH, heavy metals and phosphates in different parts of the lake. The students were also given a model for argumentation (Erduran et al., 2004) that pointed to the importance for warranting or rebutting their own or other students’ arguments. Through this materiel, students had something around which they could coordinate their participation. However, participation in the debate was not only about expressing different conclusions made from the material through the argumentation model. Students also reflected on which conclusions drawn from the reports would be beneficial or disadvantageous to the stakeholder they were representing. They prepared arguments that would put
responsibility on other stakeholders and prepared defences for accusations that could be thrown against them. In the preparation lesson, the teacher aided students in this reification by telling them to work like solicitors preparing a case. How those prepared arguments worked in the debate were negotiated in interactions in which the students responded and modified their arguments to steer the discussion in different directions. Students were thus negotiating participation in the debate from reification of the material, and reifying arguments from participation. However, since tensions arose between emergent purposes that were challenging for the students, the following socio-political debates were modified to include couching discussions where the negotiation of participation could be discussed among the students.

Reifying quality in assessment of scientific literacy

Several examples emerged of how grounding the assessments in practices helped students and teachers to reify quality of participation. Trying other students’ experiments provided concrete examples from which to reify the qualities of experimental design. Similarly, participating in discussions with colleagues and bringing students’ work afforded the teachers an opportunity to reify quality not only of the students’ performances, but also the quality of the assessments used to make the students’ performances discernible. It is difficult to discern to which extent reification was coordinated towards alignment in the long term, since data were collected within a short period of time. I cannot determine whether students provided new feedback that was more usable for the peers the next time they engaged in the peer assessment of experimental design (if they did so). Similarly, I was unable to analyse how the assessment moderation meetings affected the teachers’ future assessments. I did, however, receive plenty of data where students and teachers discussed and agreed on how to assess the quality of a performance and use feedback. The moderation effect of the assessment moderation meetings was reported to be short-lasting (Klenowski & Adie, 2009). Contrariwise, prior studies on primary school students have shown that students can learn to become better at giving feedback (Dixon et al., 2011; Willis, 2011). I would therefore suggest that both peer assessment and assessment moderation meetings should be more frequent for developing assessment practices. When the assessment of SL as participation in civic practices focusses on the negotiation of participation, it is important to have arenas in which the quality of reification can be reified in open discussions.
Making scientific literacy discernible in student participation

From the three empirical studies I would draw the conclusion that it is not reasonable to aim to authentically assess students’ abilities to participate in practices per se. That is, the aim is not to assess how well students can act like, for example, farmers and politicians. First of all, the assessment of participation in practices is constrained by the opportunities to participate authentically in the assessed practices. If, for example, participating like a CEO becomes disadvantageous for the assessments, then the assessment is not authentic in that sense. It has been observed in tertiary education that new purposes emerges in authentic assessments when students try to envision what it means to participate in professional practices for which they are training (Sambell et al., 1997). On the one hand, teachers could therefore consider it a challenge to deal with the confusion caused by transferring out-of-school practices to classroom assessment. On the other hand, Wenger (1998) points out that the emergent in the practice is as important as the designed, since it emerges from community members’ participation. When transferring practices into the classroom, additional problems of reification emerge that do not necessarily exist in the authentic practices (ibid).

There were always some scholastic elements that caused tension between the practices inside and outside of school. These tensions were not only obvious in students’ affordances and constraints for participating in the socio-political debate. Tensions between the school context and out-of-school contexts became obvious in students’ feedback and use of feedback. The feedback concerning personal experiences and preferences, which were often disregarded by other students, was a reification of the applicable value of the research for the assessing student. The tensions between school and out-of-school practices also became apparent in the contradictions between how teachers had assessed scientific knowledge and how they came to understand the syllabi standards regarding source critique and decision-making. The students had not written enough about science in their descriptions of how they had made their decisions. The complication of transferring science learnt in school to the outside world has been discussed in the introduction of this thesis (c.f. Aikenhead et al., 2011). However, there were consequently also constrictions for bringing the science outside the classroom into classroom assessments. Therefore, the challenges lie in the negotiability of authenticity in the assessments.

If one views the assessment of scientific literacy as the ability to negotiate participation in practices, the tensions caused by the authenticity of the task become less problematic. From this view, the studied assessments afforded numerous opportunities to negotiate participation. As previously argued, a debate offers unique opportunities for the students to negotiate which arguments work and which do not (Säljö et al., 2011). The discussion groups
in the peer assessment similarly provided an opportunity for students to negotiate participation. In a way, the recordings of these discussions are more valuable sources for assessing students’ SL than the written designs; they provide information on students’ abilities to negotiate experimental designs. Tsivitanidou et al. (2011) could not discern how students decided to use or not use feedback. However, the recordings of the group discussions made this clearer.

Putting a greater focus on the negotiation of participation rather than whether students’ actions are correct also strengthens students’ influence over participation. The negotiability is greater, or at least more obvious, when assessment tasks are open-ended and involve explicit interaction between students. In the concept of ‘action competence’ (Mogensen & Schnack, 2010), educational ideals are characterised by democratic views and citizen accountability as agents. In assessments where students have to negotiate power relations and conflicting interests, different perspectives and identify themselves with others, participation becomes much more at the liberty of the students. This liberty is granted students as a group, and it does not necessitate the negotiability of individual students. Based on CoP and empirical data from the debate, teachers cannot solely rely on the design to ensure the mutuality for engagement and negotiability of the repertoire; there is always an unpredicted emergent aspect of assessment. There is thus a need to intervene and alter the assessments to ensure that each student is afforded negotiation in the assessments, without losing too much authenticity. An example of this was to include coaching pauses in the debates.

Transferability through negotiation

This thesis began with arguments for considering the situatedness of participation in practices. The research study has convinced me that the complexity of the transfer of repertoires between practices makes assessment of SL challenging. Students and teachers participate in multiple practices. Participation in those practices might be more or less integrated with and exclusive to each other. For example, A. L. Tan et al. (2013) showed how primary science students were in need of much help when integrating everyday experiences and science, lest they focussed too much on personal preferences and emotions. Bhattacharyya and Bodner (2014) pointed to the need of helping students transform experiences from science laboratories to authentic endeavours in their own ‘real world’. E. Tan, Calabrese Barton, Kang, and O’Neill (2013) as well as Carlone, Scott and Lowder (2014) described the challenges for aiding students in identity transformations in conflicts between home and science class cultures. There is thus a need to aid students in transformation between practices. Assessments help teachers who are making decisions how to aid students (Airasian, 2001; Black & Atkin,
2014). Focussing on students’ negotiation in assessments makes assessment of SL a little less difficult. Wenger writes:

The ability to apply learning flexibly depends not on abstraction of formulation but on deepening the negotiation of meaning. This in turns depends on engaging identities in the complexity of lived situations. (Wenger, 1998, p.268)

As I understand this, the transferability does not lie in learning strategies for negotiating meaning that can be applied when participating in different practices. Towndrow et al. (2010) argued that the local diversity between different school and everyday contexts makes ‘fit-for-all solutions’ unsuitable. Transferability in CoP is about finding and reflecting on one’s place in a community and exploring new possibilities for participation (Wenger, 1998). I am not advocating assessments focussing on students’ identity transformation in different communities, but rather on how students decide how to participate and what to reify when engaging in different practices. I therefore recommend assessing and providing feedback on how students in socio-political debates negotiate not only their role in the debate, but also how their role will influence the course of the discussion. In this assessment, the emergent purposes, which could be regarded as distractions from the stated purposes of the assessment, could be discussed as important elements for negotiating authentic participation in the debate. Similarly, I initially agreed on students’ choices to reject peer feedback that concerned personal experiences and preferences as unscientific since it was subjective. A little wiser, I would now argue that negotiating personal relevance is an important aspect of positioning oneself in relation to science and could thus be discussed with the students from that perspective.

Expanding practices

Prior research has mainly studied teachers’ moderation and alignment of assessment in discussion groups (Klenowski & Adie, 2009; Klenowski & Wyatt-Smith, 2010; Wyatt-Smith et al., 2010). Though the immediate moderation effects were apparent in the studies, the alignment of teachers’ assessment was temporal (Adie et al., 2011; Wyatt-Smith & Klenowski, 2012). I used assessment moderation meetings to make contradictions apparent to teachers and afford expansion of teacher assessment repertoires as a response to apparent contradictions. Y. Engeström, R. Engeström and Suntio (2002) found three obstacles for teachers to negotiate the collective expansion of their repertoire: A) Most teaching is conducted in classrooms that are isolated from each other, and teachers thus have difficulty coordinating their practices with each other. B) Teaching is temporal (teachers instruct students about something in a series of lessons and then move on to something else),
making continuity challenging. C) The dominant motive of school activities is to measure the success of students, leading to a division of students as more or less abled. This makes it difficult for teachers to analyse the activity system. It is my impression that many professional practices involve responsibilities where people are alone with clients and where they deal with different projects for short periods. Many professional practices also have relatively formal procedures for coordinating these isolated activities through regular conferences and habits of asking colleagues for assistance. Introducing assessment moderation meetings has the same function. However, the dominant motive for most professional practices is not to divide clients or customers into different categories from measured success. Measuring success is central in assessment moderation meetings, but only in summative assessment are teachers required to assign marks to students’ collected performances. Due to pressured schedules, teachers often choose to combine formative and summative assessments (Black et al., 2003). In the third study, the teachers went beyond dividing students into categories. The expansion moves presented concerned modifying the assessments for increased affordance of desired performance, or at least a more valid and reliable division of students’ performances. The analysis revealed those expansive moves to be grounded in a historic reflection by the teachers of their own practices. Contradictions of which teachers were unaware would not appear in the analysis, since I was only analysing contradictions that the teachers themselves presented.

From my third study I draw the conclusion that an additional major obstacle for expanding the assessment repertoire in science classroom assessment was two diverging views on SL. The three times I observed the discussions at Apple School was not always sufficient for overcoming those differences. As Langemeyer (2006) pointed out, the theory of expansive learning assumes that people want to expand the practice as a response to the contradictions. However, if the teachers do not acknowledge contradictions between what has previously been done and what is now asked of them, there is not much reason to join the expansive move negotiated by their colleagues. Young (2001) also argued that people might join the expansive move of the community out of fear of being excluded. It is unclear how the theory of expansive learning could differentiate between expansive moves as compliance and as a more engaged movement. Kärkkäinen (1999) studied elementary school teachers who were working with an interdisciplinary project. The teachers used twelve meetings where they went through eight turning points of misunderstandings and conflicts before expanding their practice in the following year. There is thus reason to argue the need for regular reoccurring assessment moderation meetings to be able to discern transformations in the assessment repertoires of a faculty of science teachers.
Implications for authenticity

The review concerning different uses of ‘authenticity’ in science education research is not intended to be used to pit the different uses against each other to decide which one is preferable. The implications different uses have for the design of authentic assessment is, however, noteworthy. When I claim that an assessment such as the socio-political debate is authentic, I am claiming it is authentic to something. This could mean that it is authentic to the practices of farmers, fishermen, etc. However, it could also mean that it is authentic to what students find to be relevant issues to address. Despite the complexity, I would argue for the negotiability aspect of both cultural and personal authenticity (Murphy et al., 2006). When negotiating membership in a community, students become part of what is reified as cultural authentic participation in the community. From this perspective, the assessment becomes culturally authentic because students influence what is reified as authentic participation. I would argue that personal authenticity is also negotiated by the community because what one believes is meaningful to pursue concerns where one positions oneself in social contexts and one’s sense of belonging. From this perspective, the assessments become personally authentic when students are included.

What more is to be done?

Though many changes in participation could be observed in the short intervention studies, there is further need to study how the whole activity system transforms over time. I consider this to be particularly necessary regarding peer assessment and assessment moderation meetings. How do the negotiations about how to use feedback change as students become more familiar with these negotiations? How do negotiations about feedback transform with different types of scientific activities, for example, between peer assessment of experimental design and debating? It is also necessary to study the effects of reoccurring assessment moderation meetings on teachers’ assessment when they are assessing on their own and the feedback discussions they have with their students.

Furthermore, my review of the three highest-ranking journals revealed that most research published in those journals in 2013 and 2014 concerned secondary science education. Therefore, it would be interesting to study how primary science students negotiate participation in civic practices. There is reason to assume that primary students have less experience participating in civic practices. However, that does not necessarily mean that they would be less able to negotiate participation in civic practices. For instance, Byrne, Ideland, Malmberg and Grace (2014) showed that 9- to 10-year-old students
were able to make relatively advanced discussions about climate change, taking into account several aspects of the discussed issues.

It is worth the trouble of assessing scientific literacy as participation in civic practices

It is easy to focus on the constraints for developing assessment of scientific literacy. Students were constrained from putting their scientific conclusion on display for the teacher in the socio-political debate. Students were also constrained from using peer feedback that differed from their own. Teachers were constrained from following their colleagues’ expansion after conflicts emerged in the assessment moderation meetings. The schools’ diverging expansion and researchers’ diverging use of authenticity constrain science educators from forming a joint enterprise of shared values and repertoires of authentic assessment of scientific literacy.

However, the affordances and constraints depended largely on how the assessments actually related to SL as situated in practices. The studies did, for instance, consequently show that assessments designed for assessing SL as situated processes became problematic for assessing scientific knowledge as possessions being transferred and applied in the assessments. They were, however, much more functional for assessing students’ negotiation of participation. Furthermore, socio-political debates provided limited evidence of how students could explain scientific concepts and theories, since students had to struggle to put such knowledge on display. Nonetheless, the debates afforded gathering of unique evidence concerning how students connected conclusions of scientific content to different interests of various stakeholders as well as how they negotiated this with each other. As Jönsson (2011) explains, making assessment authentic is not about investigating how facts and theories learned in one situation are applied in another situation. Rather, it involves how students deal with situations they might face in society. Similarly, assessment moderation meetings could afford expansion of the assessment as long as the teachers could come to a conclusion whether they should assess propositional or procedural knowledge. The teachers could then amend the assessments to afford student participation for the agreed assessment purpose. The importance of negotiating a shared purpose for the assessment was also important for affording the students’ use of peer feedback; this was seen in the influence the discussion groups had on students’ choice to reject or use the feedback. Peer feedback that was judged unusable could still be used in the discussion of what counted as good experimental work. I would therefore conclude this thesis by claiming that a lot can be gained in terms of authenticity by assessing scientific literacy as participation in civic practices. However, when taking this approach, teachers and students
have to abandon demands for pre-set ideals for participation, or that specific scientific theories should be central for participation in a civic practice. Instead, they need to let the quality of the participation be reified through negotiations.
Svensk sammanfattning


För att undvika problematiken med att göra antaganden om elevers förmåga att överföra kunskap mellan kontexter, har man ofta valt att låta eleverna engagera sig i samhällsfrågor med ett naturvetenskapligt innehåll. Eleverna får diskutera fall utifrån olika naturvetenskapliga, ekonomiska, etiska och estetiska aspekter i s.k. Socioscientific issues (SSI) (Zeidler, Sadler, Simmons, & Howes, 2005). I praktiken är t.ex. beslutssättande, emellertid inbäddat i deltagande i olika praktiker, t.ex. yrken. Dessa praktiker ställer villkor för vilka beslut som kan tas. Ett alternativt sätt att se på kunskap är därför att betrakta lärande som utveckling av identitet och medlemskap i olika gruppers

I denna avhandling har jag valt att utforska möjligheter och begränsningar för bedömning av SL som deltagande i olika medborgerliga praktiker. En sådan bedömning ställs inför två stora utmaningar: För det första så finns en stor utmaning i att kommunicera vad som kan betraktas som prestationer av hög kvalitet så att eleverna kan få någon återkoppling om hur deras prestationer skall förbättras. En problematik ligger här i att detta är något som i praktiken förhandlas i elevernas deltagande och därmed blir svårt att konkretisera på förhand. För det andra måste kvaliteterna i SL bli synliga i elevernas deltagande så att bedömningen vilar på bevis. Deltagande i bedömning har emellertid också konsekvenser i form av uppfattningar och erfarenheter av inkludering och exkludering, ökat eller minskat självförtroende samt positiva och negativa attityder till den typ av deltagande som bedöms. En problematik i denna utmaning ligger därmed i att designa bedömningar som både tydliggör SL och har önskade konsekvenser för elevernas framtid deltagande. Lärande (inklusive bedömning) är emellertid inte bara en konsekvens av undervisningens design, då det alltid uppstår oförutsedda aspekter som påverkar vad som blir av elevernas deltagande i praktiken (Wenger, 1998).

Forskningsfrågan för hela uppsatsen är: Vilka är möjligheter och begränsningar för klassrumsbedömning av naturvetenskaplig bildning (SL) som deltagande i medborgerliga praktiker? Denna forskningsfråga har besvarats genom fyra delstudier som tar upp möjligheter och begränsningar för: elevers deltagande (artikel 1), kommunikation av bedömning i kamratbedömning (artikel 2), utveckling av lärares bedömning i sambedömning (artikel 3) samt autenticitet i undervisning i naturvetenskap (artikel 4).
Artikel 1


Två klasser från årskurs åtta fick delta i en debatt om framtidens sju där beståndet av fisk påstods ha minskat. Läraren och jag kunde använda autentiska data om syrehalt, surhetsgrad och mängden av andra föroreningar från Länsstyrelsen för att konstruera vetenskapliga rapporter. Vi skapade sedan roller utifrån vem som skulle kunna tänkas vara orsak till föroreningarna. Eleverna blev tilldelade roller som fiskare, bönder, politiker, medlemmar i en miljöförening och representanter för en intilliggande PVC-fabrik. Eleverna fick sedan två lektioner för att förbereda argument och på den tredje lektionen genomfördes två videofilmade debatter i varje klass. I analyserna framkom det att eleverna förhandlade de konflikterande syftena: 1) Visa upp naturvetenskaplig kunskap eller behålla sin rollkaraktär. 2) Visa upp naturvetenskaplig kunskap eller ta ett socialt ansvar. 3) Visa upp naturvetenskaplig kunskap eller vinna debatten. 4) Använda källor taktiskt eller använda källor kritiskt. De uppkomna syftena begränsade på så sätt elevernas möjligheter att visa upp det kunnande som läraren hade begärt. Debatten gav emellertid mycket information om elevernas möjligheter att förhandla fram ett autentiskt deltagande för sin rollkaraktär i debatten (jmf...
Åberg, Mäkitalo, & Säljö, 2010). Sådan information är värdefull för att bedöma elevernas handlingskompetens för deltagande i medborgerliga praktiker där naturvetenskap spelar roll.

Artikel 2


Fyra klasser från årskurs åtta och nio planerade undersökningar för att jämföra vilken effekt två olika typer av frukost hade för effekt på en fysisk aktivitet utförd på morgonen. Elevernas planeringar gavs till en annan elev som genomförde undersökningen och gav återkoppling på hur planeringen kunde förbättras för att jämförelsen mellan frukosttyperna skulle bli bättre. Eleverna placerades sedan in i grupper där de fick diskutera användbarheten.

Artikel 3


konflikter angående om man verkligen kunde kräva att elever skulle kunna använda naturvetenskap i nya sammanhang. Två av lärarna valde att öva källkritiska och beslutsfattande processer inför den andra uppgiften, men tre lärare uteblev från det sista sambedömningsmötet.

Sambedömning kunde i viss mån användas för att utveckla bedömning för delar av kursplanerna som inneburit osäkerhet för lärare. Det fanns emellertid möjlighet att undvika att följa kollegornas förändringar. I sambedömningsmötet utsätter sig lärare för exponering inför kollegorna vilket kan leda till att de känner att nackdelarna väger över fördelarna (Klenowski & Adie, 2009). De tre lärarna som inte var närvarande i sista sambedömningsmötet hamnade i flera konflikter angående sin syn på hur SL skulle bedömas under de tidigare mötena. Det finns därför anledning att misstänka att möjligheterna för att övervinna divergerande kunskapssyn med hjälp av sambedömning är begränsade åtminstone på kort sikt.

Artikel 4

I den fjärde studien gjorde jag en litteraturanalys av de tre högst rankade tidskrifterna inom utbildning riktad mot naturvetenskap. I artiklar publicerade 2013-2014 undersöktes hur autenticitet användes i beskrivningar av utbildningen. Naturvetenskaplig undervisning behandlades som autentisk på grund av att den var: 1) jämförbar med professionella forskarens aktiviteter, 2) grundad i elevernas värld, 3) involverade undersökande arbetsätt, 4) bidrog till praktiker utanför skolan 5) innebar en utmaning att överföra kunskap mellan kontexter, 6) jämförbar med medborgares (förutom forskarens) praktiker, 7) i linje med läroplaner och definierade syften eller 8) involverade en omvårdande pedagogik mellan elever och lärare.

Kategori 1, 6 och 7 lade stor vikt vid kulturell autenticitet. I utformning av bedömning blir kvalitet något definierad externt på förhand. Kategori 2, 5 och 8 lade istället stor vikt vid personlig autenticitet. Om bedömning blir autentisk eller inte avgörs då först när eleverna engagerar sig i bedömningsuppgifterna så det på förhand blir svårt att förutse vad som kommer hända. Kategori 3 och 4 involverar både kulturell och personlig autenticitet då det både finns aspekter som beskrivs som externt definierade samt blir till i elevernas verksamhet. Utmaningen blir då att skapa bedömningar där elever kan utforska deltagande i olika praktiker för att finna sin identitet som kunnig i naturvetenskap.

References


The Swedish National Agency for Education. (2011). *Curriculum for the compulsory school, preeschool class and leisure-time centre 2011.*


