Perspectives on intended learning outcomes in the field course and their assessment

How do field instructors assess learning outcomes in the field course

Jørgen Haugvaldstad

This thesis is submitted in partial fulfilment of the requirements for

the degree of Master of Science

The institute for bioscience

The Mathematical science institute

The University of Bergen

20.01.2020

Preface

As a certified electrician I never imagined I'd ever go do higher education, much less so that I would ever write a master's degree. I'd like to give thanks for my time at the University of Bergen.

I would like to thank everyone who has made this thesis possible. My wonderful supervisors Torstein Nielsen Hole, Vigdis Vandvik and Pernille Bronken Eidesen. As well as the Field Pass project for enabling me to visit Svalbard, twice! All my friends who have supported me, listened and given feedback especially Rebecca Storevik & Are Pedersen

And off course the one none off this would have been possible without: Oda Bugge Kambestad. Planning a wedding, getting married and then immediately writing a master thesis was intense, but worth it, because I get to be with you.

Jørgen Haugvaldstad

20.01.2020

Research question

Do the assessments biology teachers employ to measure achievements of set learning goals capture what students are supposed to learn in the field?

Abstract

This thesis examines the assessment practices of biology courses at the University level that feature a fieldwork component (field course). Seven professors who teach and administrate such courses were interviewed about two primary areas of research relating to field courses: How they are led, and how the professors assess student mastery of the course's Intended Learning Outcomes (ILO) that pertain specifically to practical skills associated with field work. The study uncovered several original arguments in favor of field courses as a method of teaching, but the most significant finding is a number of discrepancies between what the students were expected to learn in terms of practical skills, and the ability of largely theoretical and written forms of assessment to properly evaluate mastery of these skills.

Table contents

1.0.0 Introduction

A field course is a structured excursion away from the normal educational setting, typically for the express purpose of exploring phenomena in their natural setting. It allows an investigation of phenomena not by bringing it into the classroom context, but rather by bringing the students into the natural context of the phenomenon (Orion & Hofstein, 1994). Field courses are considered a fundamental part of a biological education (Orion & Hofstein, 1994 Rickinson et al., 2004). The Norwegian government consciously placed a greater emphasis on such active learning methods for higher education in its white paper St. Meld. 16 titled "Quality Culture in Higher Education" (Kunnskapsdepartementet, 2016). At the same time the concept of "quality" in higher education has undergone something of a renaissance. Theories such as 'constructive alignment' seek to align intended learning outcomes (ILOs), learning activities and assessment to make what is expected of students clear (Biggs & Tang, 2011, p 12-15). This thesis examines how seven different professors of biology at two institutions of higher education - The university of Bergen (UiB) and The University Centre at Svalbard (UNiS) - approach the teaching and assessment of field courses. More specifically, the thesis aims to investigate how they work to combine constructive alignment theory with their field course, how they assess their ILOs in the field, whether everything is assessed, and their reasoning behind the choices that shape their field course teaching practices. A related, but separate line of questioning focuses on their views on the social and educational value of field courses.

This thesis contributes to a growing body of research and knowledge relating to quality in higher education. Examining the teaching and assessment practices of field courses is of particular interest, as field courses are expensive, work-intensive - and potentially greatly rewarding. For these reasons, the teaching practices found in field courses ought to be as thoroughly researched as any other aspect of higher education.

2.1.0 Background & Theory

This section deals with the theoretical underpinnings of the study, which will later be employed in interpreting the data. Though there are many other theoretical approaches to viewing field courses, such as Dewey's Theory of Experience - or one of its spiritual successors such as the Kolbs 'Learning spaces' (Kolb & Kolb, 2005) - it was decided to view field courses through a primarily sociocultural lens. This was decided because the sociocultural view of learning is the theory that lends itself best to describing and analyzing both the phenomena that arose during the interviews and the assessment strategies employed in the field by the professors interviewed.

2.1.1 The constructivist learning philosophy: Constructivism & Sociocultural learning

Constructivist learning philosophy is a compound of several theoretical perspectives to create one theory that addresses not only how knowledge is 'constructed' (internalized within students), but also its social and situated contexts. The original basis of constructivism, concerning how learning is achieved, was formulated by the Swiss psychiatrist Jean Piaget (1896-1980). His theory was based on allowing students the opportunity to construct their own individual ideas and opinions, based on relating their previous experiences to new ones. (Biggs & Tang, 2011, p22-23; Olitsky & Milne, 2012). Piaget's view is that the onerous task of "educator" falls on the student - they must be engaged with the material and performing learning activities that are directly applicable to achieving the desired learning goal. For instance, if they are supposed to learn how to pluck chickens, they shouldn't *just* read a chicken-plucking manual, they should also *pluck chickens*. In such a scenario, the teacher's primary role is to be a facilitator of well-designed and effective learning activities (Biggs & Tang, 2011, p. 22-23). Through such direct learning action, the student either builds on preexisting knowledge or changes her perspective on old knowledge (Biggs & Tang, 2011, p.22-23; Olitsky & Milne, 2012).

It should be noted, however, that this does not mean that the student must discover *everything* by themselves. From the perspective of constructivist learning philosophy, everything a student learns is built cumulatively upon some form of pre-existing knowledge, whether through a self-discovery-based classroom experiment, or through a lecture. This seemingly missing component of constructivism - the social relations between educator and pupil and between pupils themselves - makeup Lev Vygotsky's later contribution to the theory of constructivism, as well as the importance of the cultural context of learning and knowledge.

Lev Vygotsky (1896-1934) was a contemporary and admirer of Piaget's. Nevertheless, he came to disagree with some aspects of Piaget's theory. Vygotsky agreed that learning could not happen without active engagement, but in his view the method of learning that Piaget prescribed was too "individualistic"; it did not properly take into consideration that teaching activities across classrooms, universities and field courses are mediated through social and cultural planes (Scott et al., 2007). Before the student can reflect on theory or practice, they would - according to Vygotsky - first must encounter it in a sociocultural context, through language. The social context, in Vygotsky's terms, refers to the information mediated from an instructor to the student, but also between students, the class, the school and society itself (Scott et al., 2007). The cultural context describes how various "knowledge cultures" (such as botanists, zoologists, Norwegians, or a high school) emphasize objects, practices, phrasing and together imbue these with meaning and value (Scott et al., 2007). This means, in practical terms, that learning - how knowledge is mediated to the mind and constructed - cannot neatly be separated from the environment in which it took place (Wertsch, 1991, p. 18).

From the perspective of sociocultural learning theory, one first meets an idea on the social plane, where it is explored. Following this process, an individual can then reflect on and give meaning to the new information (Scott et al., 2007). This means that per the theory of sociocultural learning, learning itself is facilitated through different semiotic resources, such as pictures, models, diagrams, graphs and so on. The most important semiotic resource, however, is language. In order for a student to learn science, they must first be introduced to and learn the "social language" of the science field. This she will necessarily have to be introduced to by a 'teacher'[1] (Scott et al., 2007).

The aforementioned process implies a situated cognition in the sense that it considers knowledge not only as an internal factor, but also externally, maintained by and within authentic practice communities. Thus, in order to begin learning - absorbing and internalizing - knowledge from within a given field, one must first know and understand the "social language" that constitutes the field. In our example, this means that any student or pupil seeking to learn science, must first achieve a degree of fluency in the "social scientific language" (or put more elegantly, the language of science). On this basis, Scott (et al 2007) argues that one way to gauge student learning is to ignore their invisible, internal processes and instead simply focus on analysis of conceptual systems, to investigate the thematic patterns of their language usage (Scott et al., 2007). Finally, it remains important to not consider these theories, constructivism and sociocultural learning, or perspectives on learning to be mutually exclusive, or indeed to put too much stock in their absolute efficacy - or lack thereof - in predicting or explaining learning outcomes. One should instead view these theories as complementary and, perhaps, as engaged in a kind of constant interaction with one another (Scott et al., 2007).

2.1.2 Constructive alignment

Constructive alignment is a theory of learning that aims to focus on to the broader context of the learning environment, specifically on how educators design their courses. Though intended for the macro-scale examinations of how universities build their courses, it can also be a useful tool for the individual teacher as they plan for what students are supposed to learn as part of a course (Biggs & Tang, 2011). The theory's originator, John Biggs, began developing the theory of constructive alignment in 1994 (Biggs & Tang, 2011, p 96). It was - and remains - based on the constructivist notion that when students "construct" learning, they do so based on prior knowledge, and that they construct new knowledge primarily by performing activities centered on that which they are supposed to learn (see 2.1.1.).

Biggs's chief innovation was that courses should be practically designed and planned on the

[1] Teacher in this case in the absolute broadest sense of the word. Basically any person or medium that can communicate the right language to the student. The alternative being that the student would have to (re)invent the language themselves.

basis of this theoretical understanding. Furthermore, he suggests that any course description should explicitly state in the courses' Intended Learning Outcomes (ILOs), up to and including concise descriptions of the desired level of understanding within the framework of any given ILO. Then, as the course progresses, the students should be presented with opportunities to perform these tasks, and finally, students would then be evaluated based on their ability to successfully perform the given set of tasks. This seems like a simple point, but most university courses to date still conclude with an exam where the students must explain or define what they did declaratively, rather than performing activities, as suggested by Biggs (Biggs & Tang, 2011, p. 16-32). This method of evaluation (a written exam) might be appropriate for gauging the acquisition of certain skills, but not for all (Abrahamsen & Reiss, 2015).

2.1.3 Field courses: Real world application of Constructivist learning philosophy

Field courses is work done by students outside of the university or school to get first-hand experiences and knowledge of working with real life phenomena, to do science through observation, practical tasks and experiments. field courses have been shown to be greatly rewarding for students if "done right" (Boyle et al., 2007; Eaton, 1998; Orion & Hofstein, 1994; Rickinson et al., 2004, SEER, 2000). But what is the 'right' way of doing field courses?

A successful field excursion with one's students, requires a significant amount of effort towards preparation. Not only by the teacher, whose preparatory work extends beyond academic and educational preparation, but also must include addressing any logistic and safety concerns relevant for the excursion. Additionally, the students are also required to prepare in their own right: They must prepare for where they are going, what is required of them practically, mentally and physically, as well as for the relative novelty of the situation (Kent, Gilbertson, & Hunt, 1997; Orion & Hofstein, 1994). Actively preparing the students typically leads to a greater focus on the task at hand, rather than allowing the novelty of the excursion to become dominant. Keeping focused on the task has been shown to increase the learning potential of students on excursions (Orion & Hofstein, 1994). Preparation also synergizes effectively with practical and safety concerns, as thorough preparation will also help reduce any potential anxiety, which students might feel if they are unfamiliar with field courses in general (Boyle et al., 2007), or if they are entering an unfamiliar biome such as the Arctic's (i.e. Svalbard) (Hole, 2018). This preparatory work should also directly address concerns related to safety, which is relevant for all field courses -but especially so when it comes to land as alien and wild as Svalbard (Hole, 2018). There are many ways of preparing, and the specifics appear to be left up to the individual course leaders. For instance, one of my interviewees would -as part of the preparation for the field course - give a half-hour lecture outside in Longyearbyen, describing the local flora.

Usually, the harsh temperatures outside would quickly make the students realize the importance of proper cold-weather attire suitable for field courses in Svalbard.

Both during and after the actual field courses, time must be reserved to actively reflect on the experience and the work done (Kent et al., 1997; Orion & Hofstein, 1994, Harland et al., 2015). Students generally report finding field courses rewarding, with a majority remembering their field experiences for years, and being able to actively recall at least one of the scientific exercises they performed (Falk & Dierking, 1997; Harland et al., 2015). The students also receive opportunities to develop practical skills, as well as theoretical knowledge (Rickinson et al., 2004). Field courses helps maintain student interest in the subject, not only through witnessing the majesty and complexity of nature, but also by becoming more interested in complex minutia, which - when presented in an exclusively theoretical manner - might be hard to engage with (Dillon, 2012). Students and educators also report field courses creating a more tightknit class environment (Boyle et al., 2007; Hole, 2018). A tightknit class environment that is driven and positive towards doing well in academia serves as a positive feedback on learning (Manger, 2013).

Some of these benefits might seem "esoteric", and hard to quantify. Remembrance and interest are well and good, but what about (measurable) academic performance? Specifically, is there a corpus of more stringently evidence-based research that also emphasize the benefits of field courses? The "relative benefit" of an activity is difficult to test, both with regards to methodology and ethics. A stringent and evidence-based study would have to involve putting some students in a control group, potentially for years, risking adverse effects potentially detrimental to their education. However, some studies do exist. Eaton (1998), compared students learning about beavers. In his study half of the students spent half a day learning about beavers in the field and the other half would spend the day learning about beavers purely from a theoretical perspective. Eaton found that both the treatment and control group reported gains in cognitive achievement, but the treatment group achieved significantly higher results than the control group. In addition, Eaton found a positive link between student age and field courses. The older students benefited more from the field course than the younger students (Eaton, 1998). The latter implies that field courses is even more important for university students, though further study is required to substantiate this claim.

Another example is the State Education and Environment Roundtable (SEER, 2000), which compared eight pairs of school programs over a three-year period. One of the pairs would do field courses while the other served as a control group. Six of these were pairs from different schools, while two of them were made up of different classes from the same school. Programs which incorporated field courses improved their scores on average across multiple subjects, compared to their respective control groups. Additionally, classes that employed field courses showed increased school attendance amongst pupils (SEER, 2000).

Field courses have also been shown to increase students' interest in nature and the environment (Bogner, 1998; Rickinson et al., 2004). At least longer field courses (more than one day) (Bogner, 1998; Eaton, 1998). This combined with students remembering field courses as a positive experience they remember for years (Falk & Dierking, 1997) seem to imply that field courses can act as a strong force towards fostering interest in nature and its conservation among pupils and students. However, although people generally remember their field experiences, this does not necessarily translate directly to a permanent attitude change towards a positive interest in nature, the environment and its conservation (Dillon et al., 2006).

2.1.4 Field Courses and Pedagogical Theory

Based on the aforementioned theories and studies it seems like the following would be a reasonable hypothesis: The learning theories of Piaget and Vygotsky will manifest themselves in a variety of ways in field courses. I postulate that the basis of field courses efficacy in teaching is that the learning activities that the students perform in the field are closely aligned with what they are supposed to learn (the ILOs). Through this mediation with artifacts[2] of importance of use to scientists, taking part in and engaging with a scientific sub-community, and by speaking the language of their knowledge community, the students construct new meaning on the basis of their learning.

A fundamental feature of field courses is that they literally move students to new contexts (Hole, 2018). Thereby the students are truly immersed in a scientific community, and work under conditions intended for absorbing and internalizing the language of their scientific (knowledge) community. Which may have the added benefit of combatting scientific misconceptions too. Some concepts such as 'energy' is deeply rooted in everyday culture, making it harder to understand in a scientific context (Olitsky & Milne, 2012). Thereby, being in the field surrounded with people all striving towards using scientific terms precisely might help internalize their meaning.

2.2.0 Current views on Assessment

Assessment is typically viewed in three ways: A summative psychometric of understanding and intelligence, as a formative experience that gives feedback to both student and teacher and finally for accountability to the school owners, the state, politicians, parents, advocacy groups and other interested influential participants (Bell, 2010). The extent to which these views and their sometimes-disparate aims overlap, frequently depends on the manner of assessment used. In this segment we will investigate assessment, what it assesses and equitable assessment.

[2] Artifacts in this sense being objects of importance to scientific work. Think measuring instruments, malaise traps, using microscopes and telescopes etc.

2.2.1 Summative and Formative Assessment

There are two main branches that view assessment through different lenses based on what they consider to be the *purpose* of assessment, rather than describing what the assessment *is*, or *how* the student is assessed (Bell, 2010; Donnelly & Jenkins, 2001; Hattie & Jaeger, 1998). *Summative assessment* aims at being a psychometric measurement, its purpose is to measure both the student's mastery over an intended learning goal as well as her ability. *Formative assessment*, on the other hand, purposefully employs assessment in order to provide feedback to both students and teacher about misconceptions and were they can improve (Bell, 2010; Biggs & Tang, 2011,p. 64-66, 195-197; Black, 1993).

Formative and summative assessment is sometimes unclear. For instance, a final exam in a subject would give a student some feedback, but it wouldn't be useful as the course is over therefore, this would be a case where the assessment isn't functionally formative (Biggs & Tang, 2011,p. 64-66) In the present thesis, the view of what constitutes formative assessment follows Black (1993) in asserting that assessment is only formative if it leads to action by both the student and teacher to improve student learning. Consequently, as a course is a limited time period, this leads to formative assessment being something that is done throughout a course, not at the end.

Historically the summative purpose of assessment has been the dominant view. The summative assessment exists to measure the perceived capabilities of students, which provided important feedback to teachers and all people with a vested interest in schooling, such as politicians, parents, school shareholders, lobby groups and academics (Bell, 2010). These results could then be used to sort students, to give or withhold access to higher education, and to hold students *and* educators to account before the interest groups. Summative assessment is still used for these exact purposes and is therefore a major component of the administrative view on the purpose of assessment namely the concept of accountability[3] (Bell, 2010).

Today, the importance of formative assessment is evident, and, it has been repeatedly demonstrated that offering feedback is one of the most powerful learning tools at a teacher's disposal (Bell, 2010; Biggs & Tang, 2011, p. 64-66; Hattie, Biggs, & Purdie, 1996), For the student feedback through formative assessment show where they should focus their efforts in order- to master the ILOs (Biggs & Tang, 2011, p. 195-198). Formative assessment can be considered an active, conscious employment of meta-cognitive perspectives. For the educator, the feedback they receive through formative assessment enables them to effectively gauge whether they are communicating the subjects

[3] Accountability assessment is used as a tool to hold educators and schools to account and will not be discussed further, but it is major influence on education on all levels and it, combined with the power of interest groups, is one of the most significant reasons why implementation of new praxis and theory in schools take a long time (Bell, 2010)

clearly. Furthermore, formative assessment practices allow the educator to discover student misconceptions, missing prerequisite knowledge, and the students' relative level of mastery towards ILOs. This dialogic feedback loop between educator and student is essential for formative assessment (Biggs & Tang, 2011; Hattie & Jaeger, 1998; Perrenoud, 1998). This dialogue informs both parties and enables them to pick specific courses of action to improve student learning. Formative assessment additionally relies less on traditional forms of assessments such as written tests. There are many ways to manage formative assessments, such as performance assessments, demonstration assessments, diaries, dialogue meetings, portfolios and more (Bell, 2010; Biggs & Tang, 2011, p. 279-365)

2.2.2 What do assessments assess?

The precise nature of what students are supposed to learn in Science and science-related subjects and courses has changed over the years (Bell, 2010). The goals of Science education have expanded from mainly focusing on understanding concepts of Science, to include their procedural understanding, as well as knowledge of the process and the practical skills required to perform or demonstrate their knowledge (Abrahams & Reiss, 2015; Bell, 2010; Fusco & Barton, 2001). For example, to demonstrate a chemical reaction, *conceptual understanding* would entail knowing the chemical formula and being able to explain why energy must be added for the reaction to occur. *Procedural knowledge* encompasses theoretical knowledge such as how to set up and perform the experiment safely. *Process skills* would then be the ability to correctly follow instructions in setting up such an experiment and understanding the generic issues of measurement errors and standardized tests. Finally, *practical skills* include the actual setting up of the experiment with the provided equipment. This division of skills is to an extent artificial; practical and process skills are commonly grouped together as one when discussed in broader terms. (Abrahams & Reiss, 2015).

Demonstration and performance assessments were introduced with this in mind (Bell, 2010; Fusco & Barton, 2001). Educators feared that while traditional "pen and paper tests" measured students' ability to recall and present knowledge about science, these tests would nonetheless fail to validly demonstrate the students' combination of knowledge, or understanding, and practical skills (Bell, 2010; Fusco & Barton, 2001). Even so, the indirect assessment of skills - which in many cases[4] could be considered a fair and practical assessment - dominate how students are assessed, especially compared to methods of assessment aimed at more directly assessing their ability to perform practical skills. A poignant example of such practices would be asking, students to *describe the process of retrieving* a standardized sample, and not - as one might expect - demonstrate that they themselves *are*

[4] In chemistry for instance many off our simpler examples of reactions are reactions that would require quite a bit of skill to do safely. The theoretical exam for ones driver's license is a wonderful combination with practical skills which adds to the students ability. There are benefits to "normal" tests, we should just be aware off what they measure and what they do not.

actually capable of retrieving a standardized sample (Abrahams & Reiss, 2015; Abrahamsen, Reiss & Sharpe, 2013).

2.2.3 A fair assessment and a hospitable climate for learning

A related consequence of the shift in focus on the purposes of assessment, and increased consciousness regarding whether our assessments are direct or indirect, is the posing of the following question: Are our tests fair?

Gipps (1998) suggests that fairness is an aspect of equity and validity. To be fair, tests should offer students an equal chance of success, but students vary in backgrounds and experience. To combat these differences, Gibbs then suggests that students should be provided with multiple opportunities for assessment to provide fairness and comparable treatment (Bell, 2010; Gipps, 1998). Various studies have shown this notion to be worth considering. For instance, it has been demonstrated that the higher the perceived stakes of the final exam (i.e. how much of the course's total summative assessment - grading - does it account for), results were consistently less favourable for women studying the subjects (Ballen, Salehi, & Cotner, 2017; Cotner & Ballen, 2017; Neill, Cotner, Driessen, & Ballen, 2019). From this follows that multiple assessments, and the utilization of assessments that directly measure mastery of intended learning goals, are necessary for a course to be fair and equitable. Biggs & Tang (2011, p. 39-50) point to the necessity of creating a fair and trusting climate for learning. Furthermore, they argue that if formative assessment is based on spurring action by teacher and student in order to improve student learning (Black, 1993) it follows that students and their teacher need to establish a social relation characterized by honesty and trust. The teacher will be unable to provide effective formative feedback to a student if the student does not trust the teacher's intention, and thus partake in counter-productive behavior such as hiding their mistakes from the teacher. Therefore, the emphasis of the teaching and assessment should be effective, engaging learning experiences, and not on externally motivating or even penalizing factors such as "do or die" exams and deadlines (Biggs & Tang, 2011). In addition, the students should not be constantly quizzed and assessed, as that indicates you're trying to cover too much. If you're trying to cover too much you cannot achieve deep learning or as Gardner (as cited in Biggs & Tang, 2011, p 43):

> The greatest enemy of understanding is coverage – I can't repeat that often enough. If you're determined to cover a lot of things, you are guaranteeing that most kids will not understand, because they haven't had time enough to go into things in depth, to figure out what the requisite understanding is, and be able to perform that understanding in different situations.

Essentially, if you can't assess everything you need to without your students feeling like they're "drowning" in coursework as a result, you're probably covering too much.

To give a practical example of everything covered in the theory: What, if the purpose of a course is to teach a student a skill that integrates practical and theoretical knowledge, such as recognizing the common insects on the west coast of Norway? They would have to practice capturing and identifying insects from that particular area, and, in keeping with Bigg's philosophy, their exam should then be to identify these common insects. This does not exclude declarative knowledge. To be able to perform this task the student would still have to be able to define an insect, as well as describe their taxonomy and anatomy. Rather it would mean that if the ILO is to be able to identify common insects, they should practice the identification of common insects. Achieving this learning outcome then necessitates their participation in a scientific community, where the students enter and approach the field, as well as practice methods of capture in the manner of an "apprenticeship". Here, they will be gently guided by "masters of the craft"[5], with a degree of constant supervision and immediate feedback not often employed in academic pursuits. Our imagined course should offer formative feedback that gives opportunities to reflect on the ILOs off the course. The course grade would be built on multiple assessments and the final exam should be the students identifying insects.

[5] In the above example by experienced entomologists used to field courses in western Norway

3.0.0 Method

This section addresses my choices when it comes to the study design. I will give a brief overview of the benefits and challenges of qualitative and quantitative research approaches and reflect on why I chose a qualitative design for my study. Following this, I will briefly discuss concerns and perspectives on scientific validity, the nature of qualitative knowledge, and methodological critiques. The participants and my relation to them is explained and the chapter concludes with a discussion on qualitative ethics.

3.1.1 Choosing a method: Qualitative vs quantitative

During my time as a student at the University of Bergen my coursework and research experience have both favored quantitative methods. Within my chosen fields of study - biology and chemistry - this is a common approach. Qualitative research methods largely fell out of favor following Lavoisier's work[6] which fundamentally changed chemistry as a discipline, turning it into a so-called 'hard science' by re-orienting the preferred research methodology from qualitative toward quantitative methods (Perrin, 1989). In time biology followed suit. Qualitative research fell further out of favor with the release of Archie Cochrane's[7] groundbreaking work *Effectiveness and Efficiency* (1972), wherein the various types of evidence are ranked by a system of tiers, and Cochrane rated qualitative research as the lowest tier, along with hard-to-verify claims based on expert opinion. (Kvale & Brinkman, 2017, p. 81-82). Given this context and background, I was surprised when preliminary examination showed that only a qualitative approach could potentially provide answers to the questions I wanted to pose with this project.

When we consider and reflect on the ways we teach biology, one of the more unique features is the field course. Through the lens of constructivism, field courses appear perfect because students actively do what they are supposed to learn. When viewed through the additional lens of constructive alignment, however, one quickly discovers that the specific work students routinely perform in field, differ notably from the ways in which they are actually assessed.

Therefore, I decided on a qualitative method going forward, as this study's goal is to increase understanding of how professors assess their field courses. Simply knowing if there is literally a test at the end of the field course, just after or how the field course and the work they do there affects their grade could be revealed through other methods. The present study, however, seeks to gain insight into

[6] Antoine Lavoisier (1743-1794) Through his work with taxing the peasantry he was able to afford to work as a scientist and the most sophisticated measuring equipment of the time and could therefore accurately show changes in substances in measurable quantities not in changes of quality. For his participation in taxation he was later executed in the French Revolution (Donovan, 2019).
[7] Archie Cochrane (1909-1988) is known as "the father of evidence-based medicine".

the teachers' views on the value of assessing field courses. Is it typically done in an official capacity involving summative assessment, and if so, how? Or if there is no such formal assessment, how exactly do teachers in the field assure themselves that their students possess the knowledge they are supposed to? Hence, as these questions concern the subjective opinions, thoughts, and perspectives of people following their own professional judgement underlying logic, a qualitative approach was chosen, in accordance with the suggestions made by Kvale & Brinkman (2017, p. 135). Qualitative studies, such as this one, primarily investigate people's actions, opinions, knowledge, thoughts, emotions and experiences, and the method therefore appears well suited for providing answers to my research questions. (Nilssen, 2014, p. 30). This study will also investigate specific work methods and the considerations given to aesthetic experiences during field courses. On this basis it was determined, by myself and my supervisor, that the best way to achieve the desired insights and a level of generalizability was to do an interview study featuring professors from two institutions. The university of Bergen (UiB) and the University Centre on Svalbard (UNIS) were chosen as they are connected to each other through the Center for Excellence in Biology Education (BIOCEED), and as a contributing part in the ongoing field pass project.

3.1.2 Believability, validity and objectivity in qualitative research

Semi-structured interviews are ephemeral by nature and cannot be repeated. To assure the validity of the chosen methodology, several measures were taken.
Initially, there must be a well-documented audit trail. I have thoroughly documented the data-gathering process, how I arrived at certain categories, which decisions were made through the study process, and why (Nilssen, 2012, p. 137-152). It should be apparent to the study's reader for how long the researcher was in the 'field', what materials were gathered and what the relationship between the researcher and the subject is. Different methods of analyzing the material also raises believability, for example by showing that the codes and categories are arrived at by triangulation.

One of the more powerful ways to raise believability in qualitative research is to have an open line of communication back to the participants of the study. This 'member checking' is described in Nilssen by a quote from Guba & Lincoln as "the most critical technique for establishing credibility" (Nilssen, 2012, p. 137-152). Thus, the transcriptions were brought back to my subjects so that they could verify the transcript. Additionally, they were given the opportunity to clarify or expand upon comments. Obviously, this leaves my research open to criticism, but at the same time it bolsters my interpretations and ensures a significant degree of transparency and accountability. Both from an ethical point of view and for the sake of the integrity of my collected data, it is in my interest to ensure that I have not misinterpreted or misrepresented my subjects, by allowing them a degree of influence over and access to the interpretations themselves. By having several points of reference for

the evaluation of my research in this way, I aim for a degree of transparency that, according to Nilssen (2012) will strengthen the validity of this study (Nilssen, 2012, p. 137-152).

3.1.3 The nature of knowledge and its impact on qualitative research data

What seems at first glance like a simple process of asking questions and follow-ups guided by theory, turns complicated when one considers the nature of the answers, the interview process and its context. Are they accurate representations of objective reality? There are several distinct theories about the nature of the information gathered through the qualitative interview. The phenomenological approach based on the philosophy of Husserl, Heidegger, Sartre and Merleau-Ponty used to be the most widespread perspective on qualitative interviews (Kvale & Brinkman, 2017, p. 44-46). Phenomenology is the philosophy of studying phenomena; it addresses both consciousness and subjectivity. In a qualitative interview it would mean that every person has their own subjective interpretation of reality and through the interview we access this unique perspective (Kvale & Brinkman, 2017, p. 44-46).

The epistemological perspective represents a different view, arguing that the interviewer and the interviewee together *create* a shared perspective on reality (Nilssen, 2012, p. 25). Therefore, all knowledge gained during an interview is a joint construct by the interviewer and the interviewee, which can only from that point on be applied to the world. It is not something that can happen independently of the interview (Nilssen, 2012, p. 25).

However, the post-modern perspective disagrees with both these views. The post-modern perspective is defined as skepticism toward all "grand narratives" (Lyotard, Bennington, & Massumi, 1984). Be that grand narratives such as "all human history can be viewed through the lens of class struggle" or simply that the information garnered in an interview can have a generalizability. Regardless of the interviewer and interviewees perspectives and whatever data is attained isn't applicable for all people in the world (Kvale & Brinkman, 2017, p. 75-76). The focus in a post-modern perspective is that knowledge is bound by language and by relations, which the context truly becomes the most important perspective. This also applies to the post-modern perspective of knowledge, from which follows that "knowing" isn't simply contained in the minds of individuals, but also in the relations and context of people, as well as the various contexts of location, environment, culture and so on (Kvale & Brinkman, 2017, p. 75-76; Hole, 2018).

My perspective, in short, is that these varied views on the nature of knowing, knowledge and the relationship of knowledge to (perceptions of) reality, while valuable individually, also constitute a framework of understanding. The disparate views they present naturally leads one to question the validity of one's own results. It is clear to me that while care and precaution have been taken to ensure

that the results of this study are valid and generalizable, they are not going to be universally applicable to all perspectives or subjective perceptions of field-related education assessment of field activities. However, this research process and the findings that result from it does provide access to access to *a* reality (Miller & Glassner, 2016, p. 50-63). That particular reality offers a glimpse into the perspectives of educators who lead courses with field courses, and what they find important in regard to the value of field courses, the aesthetic experience of field courses and how they assess the field course.

3.1.4 The interview as a data gathering method: Approaches and criticisms

In several books, David Silverman (2007; 2015) has critiqued the interview as a method of gathering data. Specifically, he critiques the very nature of arranging an interview: His argument is that by hand-picking a small group of people, who are only connected by being selected and answering predetermined questions, we are in fact not collecting 'natural' data, but manufacturing it, akin to doing quantitative research (Silverman, 2007). In addition, he points out that studies show that people tend to present different aspects or versions of themselves at different times of everyday life, but also in the interview setting (Silverman, 2007).

So, given these different perspectives on the knowledge gathered from interviews, their generalizability, their authenticity and depth, what are the arguments *in favor* of doing interview-based qualitative research? Firstly, no data is inherently unsatisfactory: even if the above critiques were "the truth", the data gathered from this study would still be useable (Silverman, 2007). More important than the way in which the data is gathered is the quality of the analysis. Dividing data into opposite poles of 'naturally occurring data' and 'manufactured data' might be satisfactory and present a valuable tool for further thinking - but taken to extremes it will simply hinder progress (Silverman, 2007). Interviewees might present different aspects of themselves in an interview, but at the same time, to assume that the realities beyond the interview is somehow completely inaccessible is to give the interview a sort of mythical 'black box' status where it exists in a state separate from all known "realities" (Miller & Glassner, 2016). As interviews access such a reality, it remains important to understand that the above criticisms can be applied to almost any kind of research. If we truly wanted objective research, it should solely be done by computers in the cold void of space, the only place and way research could be truly free from both context and outside influence.

Therefore, I posit that all methods of research have inherent limitations and potential pitfalls. The interview as a method is in this sense quite clever, as everyone is already acutely aware of its limitations, meaning that the process and the results are honest and equitable.

3.1.5 Ethics in qualitative research

With any form of research certain ethical principles must be followed. History is rife with atrocities committed in the name of 'research'[8]. In preparation for my research work, I applied to the Norwegian Center for Research Data (NSD) and received a formal approval before any data was collected. Nonetheless, I wish to briefly discuss some of the potential ethical pitfalls within qualitative research. One of the most important guidelines for ethical research is openness between the researcher and the interviewees. There cannot be any deception; one has to be completely honest and forthright about what one is attempting to investigate and gain insight into.

By the very nature of the process of qualitative research ethical dilemmas might occur (Orb et al., 2000). If, for instance, I was to observe a class of students over time I might witness inappropriate behavior on their part. It could be a minor issue, such as tired students not paying attention or browsing on their phones or computers. However, other instances might include serious matters where it would be an ethical infringement *not* to interfere, such as with observed bullying. As these emergent situations cannot realistically be prepared for, the best practice for a researcher is to be well versed in different ethical principles. Orb, Eisenhauer & Wynaden (2000) suggest three guiding principles: autonomy, beneficence & justice.

Autonomy refers to the prioritization of the subject's rights, their right to informed consent and to at any point to withdraw from the study without fear of punishment. *Beneficence* refers to doing good and protecting others from harm. I perceive researchers as driven, and I believe most earnestly desire to do good in some way. While admirable, the desire to good can potentially cause a crisis of conscience, leading to pitfalls such as a researcher deciding not to study some aspect of the human condition based on the notion that the prospective objects of study 'suffer from enough problems' - or that the subjects would in some way be compromised by participating in a study[9]. Though well intended, such a decision would rob the potential subject of *their* autonomy and choice. Marginalized and otherwise vulnerable populations could potentially offer perspectives not available anywhere else, but special care must be taken. The principle of beneficence also covers the need to anonymize the participants, as well as the researcher's moral obligation to manage the consequences if the level of anonymization is compromised. The third principle is *justice.* I have a moral obligation

[8] For example: The Tuskegee experiment where African-Americans were never informed that they had contracted syphilis; The Vipeholm experiments, in which scientists fed sugar to patients at the Vipeholm Hospital for intellectually disabled to prove the link between sugar and cavities; and the Stanford Prison experiment, a famous psychological experiment on the nature of power and human cruelty, now just as famously debunked.

[9] For example communists under the heights of McCarthyism, LGBT individuals in Saudi Arabia, and so on.

to be fair. First and foremost, this translates to not in any way abusing the trust or confidence of my participants (or subjecting them to undue physical or mental discomfort), but also to give credit where it's due. If one of my interviewees were to proposes some brilliant turn of phrase or insight that I later want to use, I cannot use that without proper attribution. This, after all, is a fundamental aspect of science: to stand, figuratively, on the shoulders of giants, by adding to a cumulative wealth of knowledge. It would be unethical to think one could refrain from proper attribution simply because the information came about from a participant in a study rather than from a paper (Orb, Eisenhauer, & Wynaden, 2000).

3.2.1 Participants

All participants were working at either the University of Bergen (UiB) or the University Centre in Svalbard during the time of the interviews. The participants had to meet the following criteria: Leading one or several courses with field courses component, while being responsible for the course description and the stated intended learning outcomes. Possible subjects who lead courses that I had previously participated in as a student or assistant were excluded, except in cases where they also lead a separate course that I had not previously participated in.

To avoid a gendered mismatch in perspectives it was attempted to have an even ratio of men and women. This turned out to be a practical impossibility due to the pragmatic realities of trying to arrange a series of interviews in late August and early September, as the new semester starts - which also coincides with the high season for doing field courses. Kvale & Brinckman (2017, p. 148) tells us that you should interview as many people as you need to answer your question. An early target was to interview six people. However, I ended up interviewing a total of seven individuals, two men and five women. One of these interviews were regrettably cut short due to time constraints. My prior connection to the participants was that I was a student at the institution they are working at (I was a student at UiB and a guest master at UNIS), and that I introduced myself and the project through an email or one of my supervisors. In addition, I had previously worked with one of the interview subjects as an assistant in a field courses, where they acted in a leadership role, but they were not responsible for the course itself. However, during our interview we focused on a different course that they oversaw, which I have never participated in. Thus, my relationship to the selected participants for interview was regarded as neutral. Seven participants were enough to achieve the necessary saturation as there was, to me, a noticeable reduction in new information as I progressed through the interviews (Kvale & Brinckman, 2017, p. 148-150). Therefore, I at this time believe that the study could have potentially benefited from fewer participants, which would have left more time for analysis.

All participants have been duly anonymized, and names referenced and presented in excerpts are

pseudonyms. One problem that might arise after publication is that my participants were all professors who lead field trips at UiB and UNiS. That means even with anonymization a large amount of people will quite easily be able to deduce the identities of my participants. What I did to mitigate exposure risk was to adhere to guiding ethical principles and giving the interviewees full access through the previously mentioned method of member checking. By giving my participants full access to my transcripts I preserve their autonomy, I gave them the option of providing feedback or to clarify their statements. In this way my participants were assured of a fair and just treatment of their words, their experience and their time (as previously outlined).

The field course the participants lead vary in length, typically 3-5 days. The size of the associated courses varied from 5 to 20 ECTS (European Credit Transfer System) points. The courses typically had 15 to 25 students. Though we focused primarily on their courses at a bachelor or master's level, many also had courses with field courses for PhDs. The pseudonyms given to the participants are as follows Laila, Philip, Marianne, Henriette, Rebecca, Karoline and Henrik

3.3.2 Conducting the interviews

The interviews were conducted during the final two weeks of August and the first two weeks of September. They were held in the interviewees' offices and lasted between forty minutes and one hour and forty minutes. Audio as well as video recordings were made in order to ascertain that each interview was properly recorded with a degree of redundancy in case of data loss. This enabled me to focus on the interview as a social process; to pay attention and engage in a dynamic dialogue, instead of devoting part of my attention to rapid note-taking. In addition, this enabled me to revisit the interviews both visually and audibly (Kvale & Brinkmann, 2017, p. 205; Nilssen, 2014, p. 31). It should be noted however that two of my subjects expressed discomfort with being filmed directly, so for their interviews the camera was turned away and employed essentially as a backup audio recording.

During the interview process, I endeavored to create a good rapport with the subjects by actively listening, paying attention, not interrupting and allowing time for follow-up questions as suggested by Kvale & Brinckman (2017, p 160). On reflection, the degree to which follow-up questions were answered and in turn often lead to further conversation probably affected the discrepancy in interview length. The most important factor when it came to interview length, however, was if the subject had read the interview questions in advance or not. One of the research subjects wanted to view the questions in advance, therefore I made the questions available to all of them. However, not everyone made use of the opportunity to read the questions in advance. This also affected the length of the interviews, but I do not think it affected the quality. The interviews always

started with a restatement of the purpose of the study as well as their rights as a participant and signing of a consent form (Kvale & Brinckman, 2017, p 160).

3.3.3 The interview guide

The interview guide (Attachment 1) has three main sections as well as an "intro" and an "outro" section. The introduction primarily deals with the field instructors' backgrounds and their field courses. The aim of this was to get some general information about their course, and to facilitate further reflection from the subjects during the interview process. The outro contained questions concerning their own experiences as students attending field courses, and furthermore was designed to give them opportunities to expand upon or to bring up any additional subject they felt needed more information or further clarification.

The main portion of the interview can be divided into three parts: Learning in the field, Aesthetic experience and specific learning outcomes.

Learning in the field: This section focused on the differences between teaching activities done at the University Campus compared to when one is in the field; as well as how they prepare themselves and how they instruct their students to prepare.

Aesthetic experience: This section focused on the unique emotional, sensory and aesthetic experiences the students might experience during their field courses, the value of these experiences, and the value of these experiences when it comes to learning biology.

Specific learning outcomes: This section focused on what exactly the student is supposed to learn during the field course, what they specifically do to learn this, how this is stated in their course descriptions, how students are assessed in the subject, and finally how their assessment practices cover their stated learning goals.

Prior to the interviews being conducted, the subjects received the outline for this project, which included a short introduction to any relevant background theory. This might have influenced their answers - on the other hand, it stands to reason that as biology educators, they could very well have already been familiar with these theories.

3.3.4 Directed qualitative content analysis

This study employs the analytical framework suggested by Assarroudi et al. (2018) as well as the principles outlined by Nilssen (2014). Nilssen tells us that even though the process of analyzing qualitatively can be broken into distinct segments, the lines between them are blurred. In her view, the

analysis starts immediately with the research question and every step of the way - be it interview, transcript, gathering literature, journaling, reflecting or coding - is part of the analysis. Furthermore, it is the shifting between the various parts that is essential for one's ability to be able to see analysis from all angles. A direct content analysis has its basis in a subject needing more research. Through a direct content analysis, the goal is to expand upon existing theory, to validate or to elucidate differences between what theory covers and what is discovered. It is a method expand upon concepts and theory. Therefore, existing research serve as a springboard to focus the research question (Assarroudi et al., 2018). Nvivo18 was the program used, and 'nodes' were employed to track answers and to code.

Through existing theory, the researcher arrives at key concepts or variables as initial coding categories. Through the interview a series of open questions followed up by closed questions for each one of these categories are posed, meaning that the initial pre-analysis coding is the interview question subjects themselves (Assarroudi, et al., 2018)

There are 'risks' associated with basing one's research on existing theory. For instance, one risks guiding the subjects and having a high probability of ending up supporting research that already exists. Through the questions the subjects could also feel a need to answer in a way that pleases the researcher, which in turn would lead the results to be a misleading representation of reality (Assarroudi, et al., 2018; Silverman, 2007).

3.3.5 Transcription

The transcription was done over several months and the researcher moved between transcribing and analyzing, as described in Nilssen (2016). The raw transcript consists of around 50 – 60 000 words. The citations presented in this thesis are (usually) translated and occasionally shortened to be made more concise. See Table 1 for examples.

Table 1: Examples of raw transcripts compared to concise, translated and cited transcripts

Raw transcript	Concise, translated and cited transcript
M: Ja, man kan jo ta en analogi "Man blir ikke en god kokk av å lese kokebok" sant? Eller på skolekjøkket, du kan jo ikke lage mat av å ha sett på bilder av mat. (latter) man må få det inn i fingrene og få prøvd ting.	"You can't become a great chef *just* by reading cookbooks" - Marianne
S: Ja fordi du skjønner det går litt på måten de spør på. Så prøver jeg og den som er med som assistent å være litt aktiv og stoppe de underveis og spørre dem "Hva er dette? Hvor kommer det fra?" Hvis du ser at det er en hasselnøtt og ingen har oppdaget det så kan du ta den frem og spørre dem "Hva ligner dette på?" Siden de ikke er mer enn femten så er de fire grupper så er det helt overkommelig å stå ved siden av de eller så litt i nærheten og høre hva de snakker om. Å korrigere om du oppdager at de tar helt feil eller spørre de mer for å fortsette den debatten eller diskusjonen de er i. så det tror jeg fungerer ganske greit for vi har ganske god tid. Så det er ikke stress fordi man kan jo fort sende de ut i felt også jobber de en hel dag også kommer de tilbake også har de ikke gjort noen ting av det de skal, men jeg tror ikke det er tilfelle... Nå hørtes det veldig sånn skrytete ut, men det er jo en fordel at vi er på et begrenset område og alle er der og hvis noen finner noe veldig spesielt så kan alle komme bort og se på det	It depends on the way they are asking questions… It's completely possible to stand nearby and just listen in on the students discussions. Then correcting them if they are completely wrong or asking them more to extend their discussion – Laila 24:31 – 26:16

The transcripts and recordings were sent to every participant for 'member checking' as described in section 3.1.2. This resulted in no one giving corrections or expanding on their points. In accordance with the methodology previously described, I take this to mean that all participants validate and confirm the accuracy of the transcripts.

4.0.0 Findings

In this segment I will give a brief summary of the major findings of this study, before they are explored and discussed in greater detail.

Everyone employs Vygotskian observations of engagement with scientific culture, but not everyone designated time to reflect on the tasks performed or utilized the pedagogical techniques they were employing to full effect. The professors highlighted the "unofficial" benefits of the field; such as its impact on group dynamics, how it can foster self-confidence and an increased sense of mastery.

Most felt that the final exam covers the field course, and that students wouldn't be able to perform as well on the exam if they had not attended the field course. They reported this despite there being no direct measurement of practical skill mastery associated with the exam. This was explained as a consequence of fears that that if the field course and the practical tasks done there were assigned a fixed value in terms of summative assessment, i.e. by influencing the final grade, this would be a distraction from learning, as well as potentially detrimental to student motivation. In short, they didn't want to "grade everything".

Most had trouble applying the theory of constructive alignment in practice. Many struggled with formulating effective Intended Learning Outcomes (ILO): How to state them, that the ILOs would make the course rigid or that the students would even read them.

Field courses was also highlighted as a unique resource and opportunity in the sense that it can lead to one discovering one's (future) specialty. One of the professors reported actively employing the field as an "idea incubator" for MA-level theses and beyond.

4.1.1 Assessment in field, the fields value to the course

"*You can't become a great chef just by reading cookbooks*" - Marianne 4:58-5:02

When asked how they go about assessing that their students are mastering the tasks they have been assigned in the field, the majority reported using methods which can accurately be described as Vygotskian. They observe the students' engagement with scientific culture and language; gauge whether they are performing scientific tasks correctly and consider if they are using the appropriate scientific terminology in any given context, as well as observe how students reflect on their own observations. This is what I mean by the term "Vygotskian methods": observing student engagement with scientific culture and the performance of scientific tasks. All teachers must do this to some extent but viewing this through a Vygotskian lens seems particularly apt in this situation, as it contains a high degree of intersection between (scientific) culture, pedagogy and psychology.

> It depends on the way they are asking questions… It's completely possible to stand nearby and just listen in on the student's discussions. Then correcting them if they are completely wrong or asking them more to extend their discussion – Laila 24:31 – 26:16

The professors actively listened in on student discussions to verify that they employ terminology and definitions correctly in practice. They noted the questions they asked each other and of the professors, and observed the nature of the questions being asked, particularly whether the students asked surface-level questions such as "what is this?" and "How do I do this?", or if they were asking questions that reflected a deeper level of understanding. Such questions would try to relate what they were doing to the relevant theory, and more clearly connect their reflections around the relevant theory and its employment in practice. The professors report doing this to be able to "have a feel" for if the students were performing the given tasks correctly and, apparently or not, connecting practices with theory. This distinction is important for any instructor, as it forms part of self-verification system that gives some indication of the degree to which what they are trying to communicate is being received as intended and reflected upon by students. Some of the respondents employed reflective diaries in their teaching, where the students answered general questions about the day's events or more specific, short reflective questions. Both allow students to express themselves in writing and in private, as well as offering some learning benefits through a meta-cognitive consolidation process (Scott et al., 2007). These diaries allowed the professors who read them to gain new insight into the thoughts of students who were less openly expressive than others:

> The diaries, I think, help. When you read them, you get a feeling [of students understanding]. I mean, some students are really quiet, but then they have these whole structures in their heads of thoughts and it's amazing – Karoline 1:01:36-1:01:46

They asked the students to answer in their diaries questions such as "What was the most important thing you realized/learned during your first data sampling, and why do you believe it is important?", "How do you feel about working in a group project? What parts did you particularly like or dislike?", or "What would you change if you had a chance to do the project planning and data collection over again?" In other words, the respondents were observing student activity (through writing) as the students actively engaged with scientific culture and language.

> In the field it becomes important that we have regular meetings and summarize what we have done that day… [Understanding] doesn't just appear at the end [of the course], it is a continuous process – Rebecca 42:47-43:10

Some of the respondents also employed debriefings or dialogue meetings where they asked their students various things, such as “what have we done today? What was interesting? What were the main results?”, “What was the most surprising thing?”, These questions are part of a meta-cognitive tactic to spur on the student’s reflections on the day’s tasks, to in turn make them consolidate knowledge and learn more effectively. It is practical application of the reflections on learning that Vygotsky found so important.

It is notable that when describing the debrief meetings and the reflective diaries, some of the respondents who employed them appeared to have had these well-founded pedagogical practices externally propagated to them, while their own understanding of the meta-cognitive learning potential of these methods remained incomplete. For instance, Henrik would routinely provide his students with several usefully conceived reflective questions and allow students the necessary time to engage in diary-keeping - but he would only actually read the answers himself if he coincidentally found the time to do so. It remains an effective practice if students are allowed an opportunity to get to reflect on their work, but he is cutting himself off from some of the potential benefits of the practice by not employing this source of feedback, and insight into potential student misconceptions. It should also be pointed out that only the professors interviewed at UNIS explicitly stated that they employed these methods, though it is possible that the professors interviewed at UiB did and it simply didn’t arise during the interview when asked about how they assess student learning in the field.

The respondents who employed debriefs expressed that after becoming familiar with the practice and being out in the field, the students overcame their innate shyness and got used to sharing their thoughts and experiences in front of onlookers. Every respondent pointed to such “unofficial” considerations and consequences of the field trip as being vital or important. They argue that the experience of a field course experience, whether good or bad, creates a distinct group identity or a sense of camaraderie among students, and that. the field serves as a “social glue” between students. This, in turn, translates to a more tight-knit group or class, akin to what one you would expect to find at a summer camp or in a high school class than what is typical for the social structures within higher education. Though highlighted by the respondents as an important aspect of the field, it is not expressed in the course descriptions (though there are references to “working as a team” in some courses).

Many respondents also noted that being presented with the opportunity to employ the varied methods of the field - and (ideally) succeeding - created a feeling of mastery, ownership and greater self-confidence. Another such “unofficial” aspect of the field is how the typical hierarchies of higher education are temporarily suspended. For a while, everyone is equal, lowering the threshold for social interaction. In such a setting, it is easier for students to talk to and get to know the professors (and

vice versa), in a way that would be hard to reproduce in a lecture setting. Or as my respondent Marianne put it:

> They can ask the person next to them, they can easily ask us. Because on board everyone is equal, even though I am the boss, everyone is equal (14:31,7-14:53,0)

When asked about testing the student's mastery of practical field skills, most respondents reported that they considered the "theory behind the skill"[10] to be the critical component, and that in their ultimate assessment (the exam) the student would not be able to perform well without having previously employed these skills in practice. Taking samples, performing skills and doing practical work quickly reveals challenging and sometimes unthought-of aspects of methodology, theory and sampling. It offers students greater insight into how studies that suffer methodological shortcomings might still constitute reasonably "good work" when considered pragmatically, as practical impossibilities might lead to methodological shortcomings. One of my respondents, Rebecca, made use of this as part of her teaching regimen: Every year the students reviewed the field reports of last year's students. Typically, the new students would then viciously (and gleefully) critique r last year's reports for several reasons, typically for being "sloppy" and "without methodological rigor". Then, they would go out into the field and try to replicate the study themselves, and almost certainly discover that the practical portion of a study is at the mercy of a wide range of happenstances that can interfere with methodology and the results. Therefore, there is a truth to this: Without attempting to do it themselves, they would likely not have as measured a perspective on their exams.

But what about assessing practical skills? When asked about this, most respondents said that even if they wanted to assess practical skills, they were uncertain of exactly how to do it, that it didn't strike them as meaningful use of time, or that they felt it was covered by a written work such as the field report, research poster or exam. Some suggested that maybe a lower-tier course could cover the actual mastery of practical skills.

Several respondents expressed concern about connecting the field component of the course to a component of the final grade. Their concern was that they didn't want to "grade everything". What this point of view neglects, is that even if there is an evaluation of the field through a report or exam, there are still aspects of field work that never directly plays into the final course assessment (i.e., the course grade). Based on the statements from the respondents, there is a distinct but hitherto unaddressed tension between the frequently expressed view that doing field work is considered a necessary, baseline part of educating biologists, and the observation that the practical application of

[10] For example: The theory behind taking an objective sample, not being actually able to take one. Why you're taking a snow sample and not how to do it (and safely).

skills the students are expected to glean from field work is, in fact, wholly exempt from their assessment.

A common worry or counter-argument to this, is that the students were already too "anxious" about grades, so that having parts of field courses assessed and thereby translate into grades would create a stressful atmosphere that risked ruining the joy, motivation and questions that can be generated in the field. This view has the following implications: The instructors believed that students being assessed by grading in their field courses would be detrimental to student motivation. Their reasoning was that knowing that one would be graded interrupts and negates the desired learning potential of practical work. These are two fascinating implications which the thesis will shortly return to, as part of the discussion of the results and what they suggest (see 5.0.0.).

4.1.2 Troubleshooting constructive alignment

> Everything has to be assessed, but how then, do I assess it? – Karoline 59:55-1:00

Constructive alignment[11] describes the alignment between intended learning outcomes (ILO), performing the aligned learning tasks and then testing the students in performance of those tasks. All my respondents expressed various issues with being able to implement the theory in practice.

With regards to the intended learning outcomes, the respondents faced varying kinds of problems. Karoline worried that by being (too) specific with her course description, she would restrict herself to narrowly adhering to the description, and in turn be rendered unable to dynamically adjust the course, whether to manage unforeseen necessities or to expand on "great ideas" by guest lecturers. Adding an explicit degree of flexibility to one's course description - with provisions such as an estimated amount of time and explaining the reasoning behind such a "flexible description" might offer a solution. In such a scenario, one could change and adapt the course in response to great ideas, if the attending students are adequately informed of any prospective changes.

Laila expressed that she had "inherited" her course description (i.e. it was originally developed by someone else) and had herself never been properly instructed in how to write course descriptions ("good ones", in her words). There should, as a matter of routine, be time set aside for professors to familiarize themselves with how to write good ILOs and how to phrase them in a way to be both be concise and informative to students, but still allow flexibility for instructors (presumably practices vary between institutions).

[11] For more see section 2.1.2

Karoline worried that the specific expression of skills would make the students expect some manner of formal training in that skills. If she, for instance, added “learn to drive a snow scooter” to the course ILOs, the students might expect several weeks dedicated to learning to drive a snow scooter and achieving a license. This is an example as it directly relates to the motivation behind this project - that is, investigating exactly these kinds of discrepancies between ILOs and actual field activities.

Two of my respondents, Philip & Marianne, expressed doubt concerning one of the fundamental assumptions of the theory, namely that students actually read the intended learning outcomes. Even when Marianne actively presented the ILOs to the students, she felt that the students would forget the ILOs and that every time she showed them it was like the first time. We will return to this problem shortly, as part of the wider discussion of my findings.

Phillip still believed, however, that a good course description and accurate ILOs were a great resource to him, personally - even though he believed that students didn’t read them, it allowed him to plan his course with a great deal of attention to detail.

It seemed at times like pedagogical theories had been introduced to some of the professors so sporadically that in describing their own teaching practices, they presented a smorgasbord of theories, including theoretically contradictory statements such as this: «That you can combine the active learning with the more traditional passive learning” (Phillip 08:10 - 08:19)

4.1.3 The field leads to questions, questions that need answering

The field and the experiences that one can have there have, for most of my respondents, had an explicit effect on them and their careers. As Laila says:

When I started studying, I planned on studying genetics, but then I went on a field courses in botany and it absolutely changed my mind. If it had just been a dry lecture, I never would have found this interest (44:00-44:10)

Rebecca suggest that students may have similar experiences: they might start studying biology because of their interest in one particular area, such as large mammals, but through varied experience in field (and lab) work they discover potential interests and passions for aspects of biology that run the risk of being perceived as “unexciting”, such as lichens and micro-variants of flora and fauna (27:00 - 29:20)

This notion, of students discovering their academic passion and interests through field courses experiences, was taken to novel territories by one of my respondents, Henrik:

The field can work as an idea generator, if the students know enough theory in advance. If they come out into the field and search for systems which can work for research questions. There is a feedback between what you see in these complex ecological settings. So many questions are generated and some of them are so interesting that the students feel a certain personal relationship to the question, and this can go on to become a master thesis and so on. (00:24,6-00:32,5)

What Henrik proposes here is that professors exploit the natural predilection of students toward asking questions about things that happen to strike their interests while in the field. This can be achieved by making it an explicit goal for educators in the field, or by giving students concrete tasks that will hopefully generate questions. This could then be followed by getting the students to write down their questions, reflecting on them, and by encouraging students to consider how these questions could be molded into actual research questions and hypotheses. Henrik is thus able to give his students the tools they need to articulate research questions and hypotheses that form the foundations of their master thesis or PhD dissertation.

5.0.0 Discussion

Here I will briefly outline the order of discussing the findings. We will first go through how the field can give rise to authentic questions and inspire interest (5.1.0), difficulties with constructive alignment (5.2.0), Assessment in the field and the fields value to the course (5.3.0), direct vs indirect assessment of practical skills (5.3.1), "I don't want to grade everything" how not grading is also a choice (5.3.2) and finally "how do I assess that" which has some practical suggestions for assessment (5.4.0).

5.1.0 Questions and the field.

The field as an idea generator is a sentiment echoed by Silverman (2007), particularly in his description of how to induce naturally occurring ideas for further qualitative work. The path described by Silverman is one of trying to discover the wonderful meanings and questions hidden in our everyday world. By close observation - and meticulous note-taking - of seemingly mundane activities, one can discover hidden categories, truisms and other established norms and habits. In the same way, the field can - according to Henrik - work as a similar idea generator in the sense that students with their heads full of theories experience the field, see mundane (or extraordinary) nature, and suddenly - when trying to make sense of their own observation and the theory - it can generate original questions. These questions can be wholly original, potentially deep and provide the student with a feeling of ownership and mastery. This observation has fascinating implications: If students routinely generate potentially original and valuable questions in the field, this should be considered, integrated into the field course and processed in to the course. Perhaps by facilitating work towards the ideas or planning for and allowing students to do authentic research. This makes a potential argument with the power to undercut criticisms that question the value of field courses.

Another valuable argument in favor of the field is found in how several professors themselves discovered their passion through the field. This suggests that field courses have the potential to spark genuine excitement about the learning process, which can in turn potentially lead to the foundation of career dedicated to answering the questions inspired from that first spark. At least to an extent, this could be considered an advantage to all the sciences that employ field courses. This is a sentiment reflected in Harland et al. (2015)

5.2.0 Constructive alignment, pedagogical theories and their applications

A general problem some of my subjects had can be summed up as not knowing how to translate constructive alignment well enough into practice, or not being familiar enough with pedagogical theories in general. For instance, Phillip's combination of active "and traditional passive

learning" appear at first glance as common phraseology used to describe his perspective on the differences between being in the field and giving a lecture. This particular wording, however, betrays a certain unfamiliarity with important concepts within pedagogy, which would have been immediately apparent (and probably considered unacceptable) within the professor's own fields of expertise. Given that every participant pointed to factors such as correct terminology, proper definitions and language as important indicators of understanding of their own fields of expertise, I believe a more conscious approach to practical engagement with constructive alignment would be greatly beneficial for these instructors in effectively communicating theory. A contrast can be observed between these less conscious approaches and others who employ constructive alignment (and other pedagogical tools) more deliberately during their field courses. Their practices rely on various methods - including, but not limited to dialogue meetings, mentorship programs, writing a diary of what went right and wrong during a day of teaching in the field and discussing it with a mentor or in a dialogue meeting; or even participating in a qualitative interview about assessing in the field. Any of these would present opportunities to reflect and engage with pedagogical theory, including constructive alignment.

The difficulties with their ILOs, as described by Karoline, Laila, Phillip and Marianne - though quite varied in their own right, hence why I will shortly address them individually - still face some of the same generic difficulties. I believe that part of the problem is caused by the relative "distance" between conceptualizing and writing ILOs, students studying them, and finally the actual course and exam. Changing ILOs remains a slow process, due to administrative restrictions. These restrictions are not necessarily a bad thing, course content needs to be somewhat stable over time for a variety of reasons[12]. But it does create a sense of distance between the creation of the ILOs and their creators. I believe this great distance between work and reward is one stumbling block for constructive alignment. As well, as distance I believe based especially on what Laila described that many professors end up viewing the ILOs as an administrative burden placed upon them and not as a resource to their course and to them. In addition, as the students are, most likely, not educated in pedagogical theory. The value of the ILOs must be brought to their attention, they must through the course be reintroduced and to get feedback on them students should be explicitly questioned on ILOs and course alignment when trying to get course feedback.

Karoline had two main worries about constructive alignment, both will be addressed here. One was, as briefly touched upon previously, that she did not want to "paint herself into a corner" with the course description: Essentially, her worry was that overly specific ILOs would make her course inflexible, a "dead thing" robbed of any dynamism. She also expressed concern that every

[12] To ensure that the university doesn't lose coverage of a subject, keeping courses stable enough that one year can be compared with another, allowing for enough time that previous changes get truly implemented. There is generally a one-year delay between finalizing changes to an official course description and the actual implementation of the changes.

statement in the course description would be perceived by students as a "promise", that she would then constantly be held to, potentially distracting from other goals and activities. Her example was that if she stated they were to drive snow scooters they would expect a prolonged course resulting in an official license[13]. She is quite right; the course description and its ILOs *are* in fact akin to a promise. They are statements of intent detailing what the students are expected to learn and should be treated as such. On the other hand, the ILOs should be statements of the main goals of the course, not every conceivable thing they should learn. An important aspect in this regard is the educator's ability to accurately describe the level of mastery expected from one's stated learning outcomes. Constructive alignment places a lot of emphasis on using the right verb to indicate level of mastery, as this relates directly to what the intent of the course is. Consider the notable difference between the level of knowledge indicated by relatively subtle changes to the wording of a potential ILO: Should they be able to *describe* Svalbard's fauna - or should they be able to accurately *classify* the fauna of Svalbard? Simply put, it should be clear what the most important goals are from the verbs, and the most important ILOs additionally should be understood at a higher level than the other ILOs (Biggs & Tang, 2011). If there isn't enough time to cover everything considered "important", that could be an important indicator suggesting that the course is simply trying to cover too much. Though exceptionally difficult, eliminating excessive coverage is an essential goal. Students simply can't learn - at least not with the desired depth of understanding and knowledge - a thousand granular goals. Biggs and Tang compare this to the writing process: Just like writers, educators must also be prepared to "kill their darlings" (Biggs & Tangs, 2011). This, of course, requires a great deal of trial and error, experience and reflection, but it is also one of the reasons - as Phillip pointed out - why careful deliberation concerning one's ILOs is a great resource for the educator in order to create a robust and detailed plan for the course.

Marianne and Phillip both believed that their students did not properly read their course description and its ILOs. However, both also expressed belief that students are "obsessed" about what will be featured during the exam, and that their students constantly hound them for answers. One potential way of counteracting this would be to accurately express what the students are asking for, i.e. detailing in the course description what, exactly, the exam will entail. On the other hand, writing a good course description is bound to be hard or at least frustrating if you're already convinced that the students don't read the descriptions, and don't much care about them. This is precisely why alignment is important: There needs to be accurate ILOs so that when the students ask what the exam will be like, they can confidently be pointed in the direction of the ILOs, which clearly state what, exactly, the students are expected to learn and at what level of mastery. As previously noted, this requires the

[13] The snow scooters are simply a practical mode of transport and it should be made clear that they are not a course goal in any way.

use of rigorously descriptive verbs - and then the exam (or exams) itself should focus on the students performing, demonstrating or describing - whichever is best suited to the ILOs - those goals.

5.3.0 Assessment in the field, the value of field courses.

The Vygotskian methods employed by the professors is a form of formative feedback for the educator that, as Black (1993) put it, provides feedback to both educator and student that can lead to change and improve student learning. This means that the teacher is offered insight into student perspectives, and is thus able to address misconceptions, and reveals to the teacher if the students understanding is surface-level or deep. On the other, hand these methods must be consciously employed by the teachers. They need to be acutely aware that student enthusiasm and verbal expression isn't synonymous with mastery. Especially when one factors in known cultural and gender differences (Biggs & Tang, 2011; Neill, Cotner, Driessen & Ballen, 2019). The field teachers need to be aware that if they choose to not employ field diaries (or equivalent method of cognitive consolidation) they risk being unable to accurately assess the level of student knowledge, and there might exist misconceptions that go uncorrected. Essentially, they need to be aware that if they do not read their students field diaries, they are limiting their own ability to give formative feedback. There needs to be a reflective perspective that not employing these methods to their fullest isn't falling back on some sort of "default", typical or standard level of education - instead, it is an active choice to cut the feedback loop between student and teacher. By embracing that the notion that there exists an active choice, the educator is enabled allot their time accordingly.

That special connection, the distinct feeling of social cohesion and easy access to professors that field courses causes, has a clear and positive impact on students (and their educator). It was a positive aspect of field courses mentioned by every respondent and it is a common mention in the literature on field courses (Boyle et al., 2007; Eaton, 1998; Orion & Hofstein, 1994; Rickinson et al., 2004, Hole, 2018). Creating a good social environment helps to create a positive learning environment and can potentially aid in combating negative aspects of student life such as loneliness and isolation[14] (Manger, 2013; Harland et al., 2015). But can this effect be viewed as wholly unique for the field course? Isn't it fully possible that students taking a BA in economics would experience the same closeness to each other and their educator if they went on a week-long cabin trip with no express goal beyond strengthening social bonds within the group? Harland et al. (2015) argues no, specifically they argue that an economics class could potential see the same social cohesion if they have a joint goal

[14] For more work on student loneliness see for example Hayley et al. (2017) an interesting possibility for study could be to look into if freshman courses with field courses has less lonely students and if so if this lowered loneliness sticks over their time in academia. If so it would be a powerful argument for field courses, but also (potentially) for team building trips for all courses.

beyond team-building. They describe achieving this social cohesion even with field courses that was mostly indoors and never even left campus. The important part and difference become that the team-building is incidental to the joint learning goals of the field course. Thereby, the social cohesion argument for field courses stands and opens the door for other scholarly disciplines to experiment with the format.

5.3.1 Direct vs indirect assessment of practical skills

Several of the respondents argued that it wouldn't be possible to do well on the exam if the students hadn't worked on their practical skills in the field. I previously mentioned[15] how taking samples and doing practical work in general tends to reveal challenges in methodology, theory and sampling. In that sense the respondents are quite right. On the other hand, being able to write or talk (as with reports or written and oral exams) about doing a practical skill does not necessarily mean one would be able to *do* a practical skill. These tests do not directly measure mastery of practical skills. Instead, what is used is an indirect method of assessing these skills (Abrahamsen & Reiss. 2015), yet these skills are often listed as course goals. It is in a sense striking that these practical skills - though described by respondents as extremely important and worth the expense, offering unique opportunities that cannot be replicated in the lecture hall or laboratory - simultaneously cannot be directly evaluated, or even be assigned a clear value that students understand, for instance by being tied to the course grade. This is even more striking in the case of courses where data from the field course is actually used as part of the exam. The data and its gathering is described and this process is assessed based on how well it is described, even if the data is poor. The student only needs to describe what went wrong and reflect on what they could do better - in this scenario to doing the practical work well has no intrinsic value[16]. Even if the implied benefits of doing practical work such as self-confidence, ownership, interest and mastery are all important for personal growth, the problem remains that students fundamentally view courses through their assessments (Biggs & tang, 2011, p. 197-198). Meaning then, that if an activity is not assessed, it has little or no value. Or stated differently: The student who tries to master all the ILOs is naive, while the cynical student who devotes their efforts towards working on what is actually assessed will thrive and likely be rewarded. This in turn points forward to the next point raised by my respondents.

[15] See segment 4.1.1

[16] Note: I am not arguing that students who collected poor data in the above example be penalised or denied their exam, I am merely pointing out that in regards to the grade the student achieves, collecting data well (or not) has no actual bearing on the way their performance is assessed.

5.3.2 "I don't want to grade everything"

Karoline made the point that they didn't want to grade everything, that the students were already anxious about their exam(s), and so they didn't wish to create a stressful atmosphere by constantly focusing on assessment. She felt that the students knowing that all their work would be graded, could be detrimental to motivation and learning potential[17]. The literature however, points to another problem with such a perspective. If students view subjects primarily through their assessment, as Biggs & Tang (2011,p. 197) suggest, and students as a result focus on learning what they think they will be assessed in, it stands to reason that a stressful, frustrating or anxiety-ridden atmosphere could just as well result from students being forced to waste their time doing activities they perceive as having little value. Or, as one of the respondents put it, that students have to "hound" their professors for answers about what is likely to be featured during the exam instead of simply having it clearly stated could easily be a source of frustration for both students and educators? This student backwash (or cynicism) needs to be directed towards the course's alignment, with a well aligned course the students should be simply reintroduced to the ILOs. In addition, trying to remove the stress factor of assessment from the field by bundling the entire assessment towards the end of the course[18], means the final assessment in the course counts for an even larger part of the entire course grade. This practice, as previously noted, has been shown to disfavor women in biology courses (Ballen, Salehi, & Cotner, 2017; Cotner & Ballen, 2017) and is most likely generalizable to more fields. In other words, if the assessment of practical skills in the field had an influence on the final grade by dividing up the grading into multiple assessments it would necessitate that the final assessment would count for less and penalize less.

In the next section methods of assessment is suggested as well as alternatives to combining the assessment of practical skills to grading.

5.4.0 How do I assess that? Or practical suggestions for field assessments

One frequently occurring difficulty in trying to align the field courses with constructive alignment theory is how to assess practical skills? First one must reflect on which skills are the most important, and on what level students should be able to perform these skills. The most literal and straight-forward way of assessing practical skills this would be to have the students perform the task, and then grading their performance. As field trips are often placed quite early in the course - and often these demonstrations of skills can only ever be performed in the field - this method strikes me as

[17] Biggs & Tang (2011) suggest several practical ways to cure this. First and foremost by creating a good work environment by trusting the students by not giving deadlines, and not constantly employing "do or die" tests, exams and hand ins.
[18] Instead of having multiple assessments throughout the course.

problematic in that it may not offer enough time for reflection. It is, however, not impossible or unheard of, both within and outside of academia[19]. A typical exam in many floristics courses, for instance, is to correctly identify a certain amount of plants. The ILO is to be able to classify plants - so during the course they go out into the field to gather and identify flora. The exam, then, takes the form of a practical exam where assessment is based on the students being able to identify provided flora. Such a course constitutes an example of constructive alignment being consciously employed to align theory and practice. This might however not be applicable to all ILOs and courses. In those cases, other venues of achieving a similar level of alignment need to be examined. One method could potentially be to be to give students 'skill checks' where they have to demonstrate performing the skill, this could be pass/fail and with a certain amount of skills passed the student could get a license or certificate. This could potentially be a boon for future job hunting. This then would rely on us convincing the students that these certificates could be a great addition to their CVs, that they would have value along with their grades.

One method might be a combination of the students filming themselves doing the practical skill and writing a reflection note or field diary entry about the skill. They would reflect on how it felt, what went well, what was difficult, and which theoretical perspectives were relevant and needed to be considered during the process. They could then make a narrow selection of these "videographic reports" and expand on their reflections, before finally turning this in to be assessed in the manner of a coursework portfolio[20], effectively incorporating field courses. field courses. This would have the benefit of clearly delineating the formative and the summative parts of the assessment for the students, as only the final portfolio would be graded. Another, potentially even more effective method, could be having the students plan and execute their own field work. This would necessitate some experience with field work on the part of the students but would in turn offer a unique opportunity to assess students based on their ability to perform field work from the planning stage all the way through to analysis of their findings. This suggestion, however, does present significant logistical and practical challenges, especially in distant and harsher locales like Svalbard. Finally, I encourage educators who oversee field courses to experiment with different teaching methods, and furthermore to perform action research on different methods of assessments in the field - and to not to fall back on ingrained practices purely out of habit or tradition.

[19] To give a radical example this is what I had to do in the military. There one would evaluate practical skills with practical tests. For instance after being instructed in and practicing first aid, the cadets are in groups of three brought into a room with three "wounded soldiers" with different wounds, they are randomly assigned one and have to perform correct first aid for that particular soldiers injury, while being observed by an officer

[20] Portfolios are pedagogically effective (Biggs & Tang, 2011, p. 256-260) and are a popular assessment practice in universities already

6.0.0 Conclusion

Marianne, one of my respondents, stated that "You can't become a great chef *just* by reading cookbooks" - and she is obviously right: to become a great chef, you must cook. Keeping hold of Marianne's culinary analogy for just a while longer, her statement and the resulting questions can all be boiled down to a single question: Why do so many of these courses test only for how well the students can *describe* cooking - and not how well they cook?

The intent of this thesis is not to wag a finger at my respondents. They are doing difficult, important work at the highest level, but to challenge some of the truisms and sacred cows of the education practices found in higher education. Every perceived "default" in the way assessments are handled - putting off assessment until the end of a course, so that it is not "in the way" or a distraction, and making a single assessment make up one hundred percent of a course grade - are all choices. Even if they are occasionally treated as self-evident and immovable, they remain choices - choices that have been shown to have several pitfalls, from the efficacy of learning to the disadvantaging of population groups. This is not to say that a written exam is never the most sensible choice, I am simply making the point that there are informed decisions to be made about methods of assessments.

Additionally, I contend that if the view of field courses expressed by my respondents - that the skills and experiences they offer are important - is echoed by the wider community, assessment practices should reflect this. Courses involving field work should, in that case, take care to implement assessments that more clearly indicate the actual value of the practical skills the students are expected to learn in the field. From the perspective of the students, that means practical skills ought to be a part of the overall assessment of these courses, due to how students typically view and prioritize course elements, emphasizing the elements to be assessed. Instead of lamenting cynical students trying to pre-empt exams, the exams, the course and the ILOs need to be properly aligned - including the field course element. This study indicates that, at least the practical elements, of field courses are currently only assessed in the most tangential of ways, and often at the individual discretion of the educator leading the field course. A properly aligned course would integrate practical skills into the assessment, whether that assessment takes the form of a final exam, or any of the alternatives suggested previously (or something else entirely).

The field is best viewed as a Vygotskian learning space, where culture, science and education intersect. Observing students, their interaction and usage of scientific language, artifacts and performance, is an accurate gauge of their learning - as long as one's observations include a written or similar component that allows one access to the student's thoughts and mental structures. This can typically be combined with giving students opportunities for reflections with such methods as reflective meetings and reflective diaries.

I believe that this thesis, on the basis of both the theories outlined and the results of the study conducted, makes the case for a more conscious approach to learning activities during field courses, and for students to be assessed not only based on how they reflect on their practical work and present knowledge in writing - but also on their ability to perform critical skills in the field. Or, returning to Marianne's metaphor one last time, that the students ought to be assessed not only based on well they recite the cookbook and identify ingredients, but on how they actually cook when they are allowed to operate the kitchen.

7.0.0 Literature list

Abrahams, I., & Reiss, M. J. (2015). The assessment of practical skills. *School Science Review*, *357*.

Abrahams, I., Reiss, M. J., & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education*, *49*(2), 209–251. https://doi.org/10.1080/03057267.2013.858496

Assarroudi, A., Heshmati Nabavi, F., Reza Armat, M., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, *23*(1), 42–55. https://doi.org/10.1177/1744987117741667

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0186419

Bell, B. (2010). Classroom Assessment of Science Learning. In N. G. Abell, Sandra K. lederman (Ed.), *Handbook of research on science education* (pp. 965–1006). Routledge.

Biggs, J. B. (John B., & Tang, C. S. (2011). *Teaching for quality learning at university : what the student does* (4th edition). McGraw-Hill/Society for Research into Higher Education/Open University Press. Retrieved from https://books.google.no/books/about/Teaching_For_Quality_Learning_At_Univers.html?id=XhjRBrDAESkC&redir_esc=y

Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, *21*(1), 49–97. https://doi.org/10.1080/03057269308560014

Bogner, F. X. (1998). The Influence of Short-Term Outdoor Ecology Education on Long-Term Variables of Environmental Perspective. *The Journal of Environmental Education*, *29*(4), 17–29. https://doi.org/10.1080/00958969809599124

Boyle, A., Maguire, S., Martin, A., Milsom, C., Nash, R., Rawlinson, S., … Conchie, S. (2007). field courses is Good: the Student Perception and the Affective Domain. *Journal of Geography in Higher Education*, *31*(2), 299–317. https://doi.org/10.1080/03098260601063628

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable? https://doi.org/10.1371/journal.pone.0189610

Dillon, J. (2012). Science, the Environment and Education Beyond the Classroom. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (2nd ed., pp. 1081–1122).

Dillon, J., Rickinson, M., Teamey, K., Morris, M., Choi, M. Y., Sanders, D., & Benefield, P. (2006). *The value of outdoor learning: evidence from research in the UK and elsewhere*. *School Science Review* (Vol. 87).

Donnelly, J. F., & Jenkins, E. W. (Edgar W. (2001). *Science education : policy, professionalism, and change*. Paul Chapman.

Donovan, A. L. (2019). Antoine Lavoisier. Retrieved October 9, 2019, from https://www.britannica.com/biography/Antoine-Lavoisier

Eaton, D. (1998). *Cognitive And Affective Learning In Outdoor Education*. University of Toronoto. Retrieved from https://tspace.library.utoronto.ca/bitstream/1807/12600/1/NQ41587.pdf

Falk, J. H., & Dierking, L. D. (1997). School Field Trips: Assessing Their Long-Term Impact. *Curator: The Museum Journal*, *40*(3), 211–218. https://doi.org/10.1111/j.2151-6952.1997.tb01304.x

Fusco, D., & Barton, A. C. (2001). Representing student achievements in science. *Journal of Research in Science Teaching*, *38*(3), 337–354. https://doi.org/10.1002/1098-2736(200103)38:3<337::AID-TEA1009>3.0.CO;2-0

Gipps, C. (1998). Student assessment and learning for a changing society. *Prospects*, *28*(1), 31–44. https://doi.org/10.1007/bf02737778

Harland, T., Spronken-Smith, R., Dickinson, K., & Pickering, N. (2006). Out of the ordinary: Recapturing the liberal traditions of a university education through field courses. Teaching in Higher Education, 11(1), 93–106. https://doi.org/10.1080/13562510500400222

Hattie, J. (2008). *Visible learning : a synthesis of over 800 meta-analyses relating to achievement*. Retrieved from https://books.google.no/books/about/Visible_Learning.html?id=lh7SZNCabGQC&redir_esc=y

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review of Educational Research*, *66*(2), 99–136. https://doi.org/10.3102/00346543066002099

Hattie, J., & Jaeger, R. (1998). Assessment and classroom learning: A deductive approach. *International Journal of Phytoremediation*, *21*(1), 111–122. https://doi.org/10.1080/0969595980050107

Hayley, A. C., Downey, L. A., Stough, C., Sivertsen, B., Knapstad, M., & Øverland, S. (2017). Social and emotional loneliness and self-reported difficulty initiating and maintaining sleep (DIMS) in a sample of Norwegian university students. *Scandinavian Journal of Psychology*, *58*(1), 91–99. https://doi.org/10.1111/sjop.12343

Hole, T. N. (2018). Working and Learning in a Field Excursion. *CBE—Life Sciences Education*, *17*(2), ar24. https://doi.org/10.1187/cbe.17-08-0185

Kent, M., Gilbertson, D. D., & Hunt, C. O. (1997). field courses in geography teaching: A critical review of the literature and approaches. *Journal of Geography in Higher Education*, *21*(3), 313–332. https://doi.org/10.1080/03098269708725439

Kolb, A. Y., & Kolb, D. A. (2005). Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education, 193–212.

Kunnskapsdepartementet (2016). *Kultur for kvalitet i høyere utdanning* (Meld. St. no 16). Oslo: Kunnskapsdepartementet [Eng. "Quality Culture in Higher Education"]. URL: https://www.regjeringen.no/en/dokumenter/meld.-st.-16-20162017/id2536007/

Lonely students more at risk of mental health problems. (2017). *Nursing Standard*, *31*(42), 17–17. https://doi.org/10.7748/ns.31.42.17.s19

Lyotard, J.-F., Bennington, G., & Massumi, B. (1984). *The Postmodern Condition: A Report on Knowledge*. *Poetics Today* (Vol. 5). https://doi.org/10.2307/1772278

Manger, T. (2013). Motivasjon for skulearbeid. In R. J. Krumsvik & R. Säljö (Eds.), *Praktiskpedagogisk utdanning* (pp. 145–170). Bergen: Fagbokforlaget.

Neill, C., Cotner, S., Driessen, M., & Ballen, C. J. (2019). Structured learning environments are required to promote equitable participation. *Chemistry Education Research and Practice*, *20*(1), 197–203. https://doi.org/10.1039/c8rp00169c

Olitsky, S., & Milne, C. (2012). Understanding Engagement in Science Education: The Psychological and the Social. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education part one* (Second, pp. 19–34). Springer.

Orb, A., Eisenhauer, L., & Wynaden, D. (2000). Ethics in Qualitative Research. *Journal of Nursing Scholarship*, *33:1*(2000), 93–96. https://doi.org/10.4324/9781315232157

Orion, N., & Hofstein, A. (1994). *Factors that Influence Learning during a Scientific Field Trip in a Natural Environment*. *JOURNAL OF RESEARCH IN SCIENCE TEACHING* (Vol. 31). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.897.3959&rep=rep1&type=pdf

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. towards a wider conceptual field. *International Journal of Phytoremediation*, *21*(1), 85–102. https://doi.org/10.1080/0969595980050105

Rickinson, M., Dillon, J., Teamey, K., Morris, M., Choi, M. Y., Sanders, D., & Benefield, P. (2004). *A review of Research on Outdoor Learning*. London. Retrieved from https://www.field-studies-council.org/media/268859/2004_a_review_of_research_on_outdoor_learning.pdf

Scott, P., Asoko, H., & Leach, J. (2007). Student Conceptions and Conceptual Learning in Science. In S. K. Abell & N. G. Lederman (Eds.), *Research on science education* (pp. 31–56). New York: Routledge.

SEER (State Education and Environment Roundtable. (2000). The effects of environment-based education on student achievement. Retrieved October 3, 2019, from http://www.seer.org/pages/research/CSAP2000.pdf

Silverman, D. (2007). *A very short, fairly interesting and reasonably cheap book about qualitative research*. Retrieved from https://books.google.no/books?id=hYcQAgAAQBAJ&printsec=frontcover&dq=david+silverman&hl=no&sa=X&ved=0ahUKEwjekdKt0evkAhWSwsQBHU-uD44Q6AEISzAE#v=onepage&q=david silverman&f=false

Wertsch, J. V. (1991). *Voices of the mind : a sociocultural approach to mediated action*. Harvard University Press. Retrieved from https://books.google.no/books/about/Voices_of_the_Mind.html?id=c52ufHz1uQ4C&redir_esc=y

8.0.0 Addendum 1 Interview guide

Interview guide

Research question:

How do field instructors assess that learning goals have been achieved?

My intent with this section is to put the interviewee at rest and get them reflecting about their own

experiences and their teaching

1. Background
    a. How long have you been teaching field courses?
    b. Tell me about the field course that you have been teaching run?
    c. What are some of the elements of the course?
    d. How many students & teachers does the course usually have?

(Depending on the answers, I'll return to these answers later and ask them how they apply what they experienced to their own teaching)

The intent with this section is to highlight what (if any) are the differences between uni. learning and in the field

2. Learning in the field

Learning in the field compared to learning in uni.

a) What do you think are the most important differences between learning activities in a university setting and learning activities in a field setting?

b)How do you prepare your students for these differences?

c)How do you prepare yourself for these differences?

d) Can you give an example of a learning activity that the students can't perform at the university that they have to learn in the field? (University setting Vs. Field Setting)

e) [Follow up] Why & how?

The intent of this section is to look into the experiences of students in the field, the instructors percieved value of of these, how these can be planned for and experiences with such practices both positive and constructive

3. Aesthetic experience

field courses have the potential to give students unique sensory, emotional and aesthetic experiences.

a) During your field courses(s) what are some aesthetic, emotional, sensory experiences your students might encounter?

b) Are these experiences valuable? (If only emotional. How are they valuable for learning biology?)

c) Why?

d) For what?

e) Is there any way you can plan for this?

f) Can you give an example of an experience where you felt you really succeeded [At giving your students such an experience]?

g) Can you give an example of an experience where you felt it didn't work and why you think that was?

This segment focuses on the spesifics, the nitty gritty of what students are supposed to learn and how they go about learning this.

4. Specific Learning Outcomes

a. What are some learning outcomes you want your students to achieve during their field courses? [Get a precise answer]

b. I'm curious, can you tell me about any tasks or problems that students are given to achieve this?

c. How do you (during the field course) assess that these goals have been achieved?

i. What do you think of this assessment?

ii. What could some alternatives be?

e. How are these outcomes related to the goals stated in your course description?

f. How are the students ultimately assessed on these learning outcomes (exam, project etc)?

iii. Do you feel that this final assessment gauges students mastery of the field course learning goals?

iv. (if some degree of "no") How could you test for this?

v. Earlier you mentioned the learning outcome [precise answer from earlier] how does the assessment address these?

vi. Should the assessment address this?

vii. Why/why not?

This section focuses on the instructors own experiences and if there is anything the subject feels that they haven't been able to say during the interview process

Outro

Have you experienced field courses as a student:

viii. From your experience as a student what is a positive element you remember?

Are there any negative or frustrating elements you remember from your field courses as a student?

Are there any valuable learning outcomes you remember from your time as a student in the field?

Did you feel that you were assessed in this?

a) Is there anything we have touched upon during this interview you would like to expand upon?

b) Is there anything else you have thought of during this interview you would like to bring up?

Notes:

Section 4: note for the methods part that I will include the professors stated course goals from the course description, this is unorthodox, but potentially valuable. If necessary I will give them the their course descriptions