University of Cyprus

DEPARTMENT OF EDUCATIONAL SCIENCES

“ON THE FLY” TEACHER-STUDENT INTERACTIONS AS A MEANS TO FORMATIVE ASSESSMENT IN SCIENCE EDUCATION

DOCTOR OF PHILOSOPHY DISSERTATION

MICHALIS LIVITZIIS

2019

DEPARTMENT OF EDUCATIONAL SCIENCES

“ON THE FLY” TEACHER-STUDENT INTERACTIONS AS A MEANS TO FORMATIVE ASSESSMENT IN SCIENCE EDUCATION

MICHALIS LIVITZIIS

A dissertation submitted to the University of Cyprus in partial
fulfillment of the requirements for the degree of Doctor of Philosophy

June 2019

University of Cyprus

ISBN 978-9925-553-26-6

VALIDATION PAGE

PhD Candidate: Michalis Livitziis

Dissertation Title: “On the fly” teacher-student interactions as a means to formative assessment in science education.

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Department of Education and was approved on the…………………. by the members of the Examination Committee.

Research Supervisor:____

Constantinos P. Constantinou, Professor,
Department of Educational Sciences, University of Cyprus

Committee Chairperson: ____

Charalambos Charalambous, Assistant Professor,
Department of Educational Sciences, University of Cyprus

Member: ____

Zacharias Zacharia, Professor,
Department of Educational Sciences, University of Cyprus

Member: ____

Theodora Kyratsi, Associate Professor,
Department of Mechanical and Manufacturing Engineering, University of Cyprus

Member: ____

Dimitrios Koliopoulos, Professor,
Department of Educational Science and Early Childhood Education, University of Patras

Declaration of Doctoral Candidate

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

……………………………………............

Michalis Livitziis

ΣΥΝΟΨΗ

Η διαμορφωτική αξιολόγηση αναγνωρίζεται ως βασικός παράγοντας για τη μάθηση και δίνεται μεγάλη προσοχή στην εκπαιδευτική έρευνα τις τελευταίες δεκαετίες. Ωστόσο, η άτυπη-διαμορφωτική αξιολόγηση και συγκεκριμένα η μέθοδος αξιολόγησης αλληλεπιδράσεων "on the fly" που συμβαίνει αυθόρμητα κατά τη διάρκεια ενός μαθήματος όταν προκύπτουν "μαθησιακές στιγμές" στην τάξη, παραμένει αρκετά ανεξερεύνητο πεδίο. Λίγες έρευνες έχουν πραγματοποιηθεί σχετικά με τον τρόπο αλληλεπίδρασης του εκπαιδευτικού και των μαθητών μεταξύ τους για την παροχή χρήσιμων πληροφοριών και παραγωγικών ανατροφοδοτήσεων που μπορούν να διευκολύνουν τη μάθηση. Σκοπός της παρούσας μελέτης είναι να εντοπίσει μοτίβα που προκύπτουν κατά τις αλληλεπιδράσεις "on the fly" και να αποκαλύψει τις δυνατότητες και τις προκλήσεις που προκύπτουν κατά την υλοποίησή τους. Για το σκοπό αυτό πραγματοποιήθηκαν εφαρμογές σε μαθήματα Φυσικής σε σχολεία δευτεροβάθμιας εκπαίδευσης από εκπαιδευτικούς που συμμετείχαν σε ομάδες εργασίας όπου επιμορφώθηκαν για διαμορφωτική αξιολόγηση και ειδικότερα για τη μέθοδο της αλληλεπιδράσεων "on the fly". Τα μαθήματα βιντεοσκοπήθηκαν και τα δεδομένα από τις βιντεοσκοπήσεις των συγκεκριμένων επεισοδίων που περιλάμβαναν αλληλεπιδράσεις "on the fly" εντοπίστηκαν και απομαγνητοφωνήθηκαν. Οι απομαγνητοφωνήσεις κωδικοποιήθηκαν χρησιμοποιώντας το σύστημα κωδικοποίησης ESRU και τις υποκατηγορίες που αναπτύξαμε σχετικά με τους διαφορετικούς τρόπους που χρησιμοποιούν οι εκπαιδευτικοί για να κατανοήσουν τους μαθητές, τους τρόπους που ανταποκρίνονται οι μαθητές, πώς οι εκπαιδευτικοί αναγνωρίζουν τις απαντήσεις των μαθητών και πώς χρησιμοποιούνται οι αναδυόμενες πληροφορίες. Έχουμε εντοπίσει μοτίβα στην αλληλουχία που εμφανίζονται αυτές οι υποκατηγορίες στο διάλογο αλλά και συσχετίσεις μεταξύ ορισμένων από αυτά τα μοτίβα με την ολοκλήρωση κύκλων ESRU, οι οποίες συνδέονται με πρακτικές που μπορούν να οδηγήσουν στην παραγωγική χρήση των αναδυόμενων πληροφοριών καθώς και άλλων που δυσχεραίνουν αυτή την προσπάθεια. Περαιτέρω ανάλυση αυτών των πρακτικών μαζί με μακροσκοπική ανάλυση σε επίπεδο επεισοδίου του διαλόγου αποκαλύπτει προκλήσεις και δυνατότητες που προκύπτουν όταν χρησιμοποιούνται αλληλεπιδράσεις "on the fly" ως μέσο διαμορφωτικής αξιολόγησης. Επιπρόσθετα, πραγματοποιήθηκαν συνεντεύξεις με τους συμμετέχοντες εκπαιδευτικούς προκειμένου να εκφράσουν τη δική τους αντίληψη για τον τρόπο με τον οποίο εφάρμοσαν αυτή τη μέθοδο διαμορφωτικής αξιολόγησης κατά τη διάρκεια αυτών των μαθημάτων. Τα δεδομένα από τις βιντεοσκοπήσεις κωδικοποιήθηκαν επίσης με τη χρήση εννοιολογικών χαρτών που

αναπτύχθηκαν από τον ερευνητή για να απεικονίσουν την εννοιολογική πληρότητα, το βάθος και τη συνοχή κάθε επεισοδίου του διαλόγου. Πιο συγκεκριμένα, αυτοί οι χάρτες απεικονίζουν την ποικιλία των εννοιών που προκύπτουν στη συζήτηση, αν αυτές προκύπτουν από τους μαθητές ή τον εκπαιδευτικό, σε ποιο βαθμό τυγχάνουν επεξεργασίας, τη μεταξύ τους συνοχή καθώς και τις βασικές αλλά ελλείπουσες έννοιες. Στη συνέχεια, εξετάσαμε αν οι προαναφερόμενες παράμετροι που συνθέτουν τους εννοιολογικούς χάρτες συσχετίζονται με το βαθμό συμπλήρωσης κύκλων ESRU σε κάθε επεισόδιο.

Μέσω αυτής της ανάλυσης, επιδιώκουμε να εμπλουτίσουμε την έρευνα σχετικά με την κωδικοποίηση και την ανάλυση της άτυπης διαμορφωτικής αξιολόγησης και ειδικότερα των διαλόγων στην τάξη όπου εφαρμόζεται η μέθοδος των αλληλεπιδράσεων "on the fly", στο πλαίσιο της μάθησης Φυσικών Επιστημών μέσω διερώτησης. Επιπλέον, τα αποτελέσματα μπορούν να είναι χρήσιμα τόσο για τους εκπαιδευτικούς όσο και για τους εκπαιδευτές εκπαιδευτικών που ενδιαφέρονται να βελτιώσουν τις στρατηγικές που χρησιμοποιούνται για άτυπη διαμορφωτική αξιολόγηση κατά τη διδακτική πρακτική στις Φυσικές Επιστήμες.

ABSTRACT

This study focuses on what we call “on the fly interaction”, which occurs spontaneously during the course of a lesson when “teachable moments” arise in the classroom and teachers have to make inferences on a moment-by-moment basis (Heritage, 2007). The aim is to identify and document what can facilitate, or impede, the effectiveness of this formative assessment method, the corresponding challenges that teachers encounter in this context and possible ways to address them so as to effectively promote students’ learning.

For this purpose, we organized teacher – researcher working groups. Teachers received professional development on formative assessment with special emphasis on “on the fly interaction”. They participated in the collaborative design of teaching learning materials for upper secondary school Physics promoting the competence of designing and implementing empirical investigations. Formative assessment activities were integrated in this sequence with the explicit purpose to evaluate students’ facility to demonstrate specific components of the competence, as well as their conceptual understanding on relevant aspects of the thematic content in each case. Teachers were prepared for enacting these teaching-learning sequences in upper secondary classrooms.

The enacted lessons were videotaped and video data were transcribed and coded in an effort to describe the various kinds of interactions that take place during assessment on the fly. We achieved this by using the ESRU (Elicit-Student’s response-Recognition-Use) coding scheme, which identifies the efforts of the teacher to elicit students’ ideas, the students’ responses, the teacher’s recognition of these responses and the efforts of the teacher to use these responses for promoting students’ learning. In addition, teachers were interviewed to gain insights into their own perception of how they implemented formative assessment during these lessons, the effectiveness of on the fly interactions, the challenges they perceived during the enactments and also the opportunities they could recognize from its use. In addition, for triangulation purposes, concept maps were developed in an effort to describe the flow of concepts as they emerged during classroom discourse, but also to depict the conceptual completeness, the depth and coherence of the dialogue.

We have identified patterns in the sequence of the dialogue regarding the different ways teachers utilize to elicit students’ understanding, the ways students respond, how the teachers recognize

students' responses and how the emerged information is used and correlations between some of these patterns and the completion of ESRU cycles. These patterns provide useful information regarding practices that can lead to productive use of the emerged information and others that hinder this effort. We have identified and categorized the practices that can lead to productive use of the emerged information and the practices that seem to hinder this effort. More analysis of these practices along with a macroscopic view at the level of episode of dialogue reveals affordances and challenges that arise when "on the fly" interactions are employed as a means for formative assessment.

Through this analysis, we believe we have enriched research regarding coding and analyzing informal formative assessment and particularly classroom dialogues where the method of "on the fly" interactions is applied, in the context of inquiry-based learning Science. In addition, the results can be useful both for teachers and teachers' trainers that are interested on improving the strategies that are employed for informal-formative assessment during teaching practice in Science.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στην οικογένειά μου και ιδιαίτερα τη σύζυγό μου Κυριακούλα Γεωργίου για την υπομονή και τη συμπαράστασή της κατά τη διάρκεια της εκπόνησης της διδακτορικής διατριβής μου. Επίσης στους γονείς μουπου με στήριξαν και με ενθάρρυναν σε όλα τα χρόνια των σπουδών μου.

Θα ήθελα ακόμα να ευχαριστήσω θερμά τους συναδέλφους με τους οποίους συνεργαστήκαμε στο πλαίσιο της έρευνας αυτής, για τις ιδέες και την ανατροφοδότησή τους αλλά και την βοήθεια που προσέφεραν και σε πρακτικά ζητήματα.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τον επιβλέπων μου καθηγητή Κωνσταντίνο Π. Κωνσταντίνου για την ευκαιρία που μου έδωσε να συνεργαστώ μαζί του και την καθοδήγηση του σε αυτή την πορεία.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The main objective of science education is to help students develop knowledge structures, reasoning skills and problem-solving skills needed to follow (even contribute) to the scientific discourse (Schauble, Klopfer, & Raghavan, 1991), but also to be able to discuss and take decisions on important socio-scientific issues. Formative assessment activities in classrooms can help to achieve such goals but also, can provide information about progress toward these goals (Black, 2010). There has been consistently strong evidence that improving formative assessment can raise standards of students' performance (Lewis, 2008).

Formative assessment can be (a) formal or planned, which focuses on explicitly obtaining information about student learning using specific tools; and (b) informal or interactive formative assessment, which takes place spontaneously in the learning environment, as part of the student–teacher or student-student interaction (Bell & Cowie, 2001). Some approaches to formal or planned formative assessment methods are written feedback (Higgins, Hartley & Skelton, 2002), self and peer assessment (Topping, 2003) and structured assessment dialogue (Christensen, 2004). On the other hand, informal assessment is usually employed spontaneously during the classroom dialogue (Yorke, 2003), either during whole class discussion or during team work.

The term informal formative assessment refers to the assessment that does not employ formal data collection procedures or tools. Perhaps the most prevalent approach to data collection within this paradigm involves the information being exchanged between the students and/or the teacher in classroom discussions. Classroom discourse is a rich data source that encompasses assessment data that could be of great value to both the teacher and the students in terms of offering feedback and guidance about next steps during teaching/learning (Bell and Cowie, 2001). Informal data collection can also take place in settings extending beyond classroom discussions. For instance, self and peer assessment activities can be also realized in an informal way without using specific tools.

Formative assessment has received substantial attention in the published research literature, which has provided significant inputs about its affordances and constraints within which it seems to function effectively (Black and William, 1998; Duschl & Gitomer, 1997). The focus of this research has been primarily placed on formal formative assessment with significantly less attention being paid to informal formative assessment despite it is probably the most common in use during the teaching practice according to teachers (Ruiz-Primo & Furtak, 2006). This creates a need for further research into this latter area to supplement our understanding of the potential of informal formative assessment and how to best go about realizing this potential. This study seeks to make a contribution towards this direction.

This study focuses on “on the fly” interaction as means of formative assessment that occurs during the course of a lesson when “teachable moments” arise in the classroom and teachers have to make inferences on a moment-by-moment basis (Heritage, 2007). On the fly interactions take place unexpectedly as an immediate response, from the teacher or from a student, to something that was overheard or observed (Shavelson et al., 2008). This spontaneous dialogue offers valuable opportunities for teachers to probe students’ thinking and therefore collect valuable real-time information on where students are in terms of their learning. This rich information that arises during the classroom discourse can help teachers to take decisions about their next steps in learning instruction.

On the fly interactions are important because they allow fine-tuning of instruction by providing immediate information about students’ understanding to be used as feedback to modify the teaching and learning activities in which students are engaged. However, there has not been extensive study regarding relations between the various ways teachers employ to elicit students’ understanding, how students respond and the use of the emerged information. Therefore, the factors that facilitate or hinder the emergence of information regarding students’ understanding and the productive use of it in a way that promotes learning, remain unexplored.

There is a number of challenges that teachers meet when applying such an assessment method. Duschl and Gitomer (1997) described some of the challenges to successful implementation of the assessment conversation. Mastering informal assessment strategies is extremely complex, introducing significant challenges to the assumptions and methods underlying the current practice of most science teachers (Heritage, 2007). On the other hand, Coffey et al (2011) claim that, while researchers and teacher educators seem to believe that strategies are what teachers need first or most to help them engage in formative assessment, there is little focus on the disciplinary substance of what teachers and students assess. This study focuses on “on the fly interactions” since there has been little effort on mapping the characteristics of these interactions (Duschl, 2010), on identifying affordances and challenges that emerge when employed and on how could teachers be enhanced to apply this method in a productive way.

The study reports on an empirical investigation of the potential of interactions on the fly as a means of formative assessment. This investigation was situated in a series of teaching sessions intended to promote students' ability to design investigations, with appropriate control of variables (Kyriazi, 2004), in a range of physics topics. Designing and implementing investigations is a key competence in learning science, since students are encouraged to explore phenomena and engage with their own ideas and those of their peers (Wenning, 2005) by employing strategies resembling those used in authentic scientific practice. This process includes a considerable degree of unpredictability for both teacher and the students. Therefore, opportunities are created for the teacher to interfere "on the fly" in an effort to support students on designing how to test their claims, to make sense of emergent findings and communicate their conclusions or, to redesign their investigation if needed.

1.1. Contribution of this research

The aim of this study is to explore how teachers use on the fly interactions as a means to collect evidence for students' current state of learning and guide their actions towards facilitate learning. We are especially interested in how the structure and characteristics of these interactions come to shape formative assessment practices in the classroom. Through this analysis, we aim to supplement the available body of research-based knowledge regarding coding and analysing informal formative assessment in classroom discourse. In addition, we identify and document specific challenges and opportunities that emerge when this method is implemented exploring how teacher-student on the fly interactions promote or impede learning in situ. In addition, we discuss implications about teachers' professional development on formative assessment and particularly on informal formative assessment with emphasis on "on the fly interactions".

2. LITERATURE REVIEW

2.1. Formative assessment

Assessment is a substantive process for both teachers and students, an integral part of the teaching-learning process. It promotes educators' development in both aspects of qualification and learning (Tillema, 2010). Scriven, in 1967, made the initial distinction between the two types of assessment in the curriculum evaluation framework, proposing that summative evaluation is concerned with providing information on the value of the educational system, while formative assessment's main role is to facilitate program improvement (in Bennett, 2011 p.6). Bloom refined the definition of formative assessment relating it for the first time to the progress of students' learning. Bloom, in 1969 stated:

> *"By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process"* (in Black, 2010 p.359).

Formative assessment might best be conceived of as neither a test nor a process, but some thoughtful integration of process and purposefully designed methodology or instrumentation (Bennett, 2011). Black & Wiliam (2009), suggest that practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, in order to make decisions about the next steps in instruction.

In order to support students' learning, it is essential to monitor the progress of student understandings, and learning more broadly, during actual teaching and learning. The purposes of formative assessment also include providing feedback to students in relation to the learning goals, giving legitimacy to the students' scientifically acceptable ideas, supporting long- or short-term learning objectives, and finding out whether an activity or task was effective. Taking action to enhance learning is an integral part of formative assessment. The overlap between the action inherent in formative assessment and teaching has been acknowledged frequently in teaching practices, for example, when suggesting further questions, suggesting further activities, questioning of students' ideas, explaining the science, or giving feedback to the students with respect to scientifically acceptable or unacceptable ideas.

Student actions are more concrete. During formative assessment, students identify information about what the teacher aims to provide, they filter it and decide what is useful to them. They interpret what they need to do, they act on those ideas, and then from whatever the teacher does or from whatever other students do, they gather more information and so on. Student self-assessment has also been acknowledged as an important part of formative assessment (Bennet, 2011).

In an effort to provide framing for formative assessment, Wiliam and Thompson (2017) drew on Ramaprasad's (1983) three key processes when applied to learning and teaching: i) establishing where the learners are in their learning; ii) establishing where they are going and, iii) establishing what needs to be done to get them there. In the next session we are going to describe in more detail what characterizes formative assessment.

2.1.1. Characteristics of formative assessment

In this section we are going to present the most important characteristics of formative assessment that distinguish it from other approaches and make it effective in science teaching and learning: the character of a dynamic and iterative process incorporated in instruction, the scaffolding and the feedback that is provided to students.

According to the Assessment Reform Group (1999), one of the three basic principles of formative assessment is that it is fundamentally connected to instruction. Formative assessment is incorporated into the instruction, as opposed to summative assessment, which occurs after instruction. Apart that, formative assessment needs to take the form of a continuous cyclic process, which informs ongoing teaching (Harlen, 2010; Bell & Cowie, 2001), in an iterative manner (Bell & Cowie, 2002; Nelson, Robison, Bell & Bradshaw, 2009), in the sense that it facilitates informed adaptation of the learning process. In other words, it is meaningful for learning when it is a dynamic process, not merely a superficial exercise measured by a test for which students are dutifully prepared. The cycle is more productive when it involves the gathering of evidence concerning learning, the interpretation of that evidence to identify where students are along a path of achieving the goals of the work, the identification of the next steps to be taken and decisions on how to take them in order to close the gap between what has been achieved and what is aimed for (Harlen, 2010).

Teacher instruction is seen as a very important theme concerning the required provisions needed to reap the benefits of formative assessment. Therefore, scaffolding strategies are considered as a key element and there is a prevailing need for further research on how to integrate scaffolding with formative assessment. Shepard (2005) notes that scaffolding refers to teachers’ support that is provided to the learner during problem solving of any context and it can be achieved through the use of reminders, hints, and encouragement, thus ensuring effective completion of a task. As has been shown, students benefit from ‘‘scaffolds’’ that embed instructional guidance in ongoing investigations (Linn, Bell, & Davis, 2004; Quintana et al., 2004; Reiser, 2004), particularly when feedback can concentrate on where the learner is and where they need to go, bridging the discrepancies in between (Tillema, 2010).

Airasian (2001) states that the purpose of formative assessment is to improve student learning through the collection, synthesis and analysis of information in a way that it can be used for instructional scaffolding. Chin and Teou (2009) conducted a study to investigate how scaffolding tools could be used in the context of formative assessment, to stimulate talk and argumentation among small groups of students, as part of peer-assessment and self-assessment; and to provide diagnostic feedback to the teacher about students’ misconceptions so that education can lead to conceptual change. The sample of their research was based on two primary classes of levels 5 and 6 and the results revealed the importance of the teacher’s responsibility to design and use scaffolding structures to support ‘assessment conversations’ and to guide whole-class discussions so as to elicit information about students’ understanding. Shepard (2005) also notes that scaffolding as part of formative assessment is a collaborative process involving both the educator and the student on how to improve performance on teaching and learning respectively.

An integral element of effective formative assessment is feedback from the teacher, teaching assistants, peers and oneself (Nelson, Robison, Bell & Bradshaw, 2009). The manner in which feedback is given to the student is crucial regarding the possibility to enhance student motivation. In particular, according to research, providing feedback through comments had a positive impact on students’ achievements and interest in undertaking further work, in comparison with giving only grades or grades with comments. The feedback must be given focusing on each student’s individual work, mentioning positive aspects and providing guidance to improve it (Harlen, 2010).

Furthermore, feedback must present educationally useful information and this information must be used to advance learning and instruction (Hickey, 2012).

2.1.2. Benefits of using formative assessment as part of the teaching and learning process

Science instruction improves when teachers are provided with curricula and instructional strategies that allow for frequent and ongoing assessment opportunities (Duschl, 2011). The strongest benefit of assessment for learning is the insight it provides into students' performance and its potential for subsequent use in improving it (Tillema, 2010). Numerous individual studies and meta-analyses report a significant positive effect on student achievement as an outcome of making formative assessment an integral part of the teaching and learning process (Marshall et al, 2009). Contingent feedback and follow-up instruction that include explanations and worked examples have been shown to promote student achievement (Pashler et al., 2007). Black and William (1998) assert that "*student-involved formative assessment raises student achievement as reflected in summative assessment.*" As Nelson, Robison, Bell and Bradshaw (2009) also report, final exam scores improved when formative assessment was included in the course design. Particularly for low-ability students, formative assessment can be a crucial strategy for improving student learning (Black, 1998, p. 25).

There might also be possible relations between conceptions of assessment and academic performance that could be interpreted by self-regulation theory, as suggested by Brown and Hirschfeld (2007). Students who perceive assessment as a constructive force for personal responsibility may gain higher grades, than those who seek to blame schools or teachers for poor assessment results and those who do not take assessment seriously, or who ignore it. On the other hand, formative assessment permits a student to reflect on progress in a safe environment without the stress and anxiety associated with someone passing judgment that also prematurely impacts on his or her grade in the course (Nelson, Robison, Bell, Bradshaw, 2009).

Formative assessment is also seen as a critical component in teaching in order to achieve conceptual development (Bell, 1995). A central part of this teaching is dialogue (in contrast to monologues) with students to clarify their existing ideas and to help them construct meaning consistent with the scientifically accepted ideas. Sharing learning intentions and success criteria, or negotiating them with the students might necessitate teachers finding out about a student's understanding at the start of a topic and beginning with ‘big’ questions to set the scene and strengthen engagement (Black, 2001; Harlen, 2006b).

Therefore, giving feedback to students with the aim to change or improve their existing conceptions in order to converge to the scientifically accepted ones and helping them to modify their thinking accordingly is a part of formative assessment that contributes to conceptual development. Black and Wiliam (1998), in their review of classroom assessment, state, “*The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, amongst the largest ever reported for educational interventions*” (p. 61).

Formative assessment, as a method of assessment, along with many other benefits, fosters an educational environment of student-centered pedagogy. With its implementation in teaching, a classroom becomes the core of ongoing discourse that helps students improve through their efforts to meet goals, expectations, and objectives based on appropriate teacher mentoring. Hence, formative assessment is underlying student-centered education (Chung, Shel and Kaiser, 2006). Therefore, consistent use of formative assessment has been shown to improve student motivation, both intrinsically and extrinsically (Brookhart, 2007; Black & William 1998, Cauley and McMillan 2010). An explanation of this might be that it helps indicate to students whether learning is relevant, whether it is possible and whether it is worth the effort or not.

Finally, formative assessment is the component of teaching in which teachers can find out about the effectiveness of the learning activities they are providing. It can be viewed as the process by which teachers gather assessment information about the students’ learning and then respond to promote further learning. Shavelson (2008) notes that often teachers proceeded through a unit until the end, before they would try to find out what had been gained and why they were teaching the activities they did. That is why embedded assessments should be used to signal a unit’s goal structure and provide complementary direction to teachers.

2.1.3. Limitations to the use of formative assessment

Given the potential benefits of formative assessments, it is worth pointing out that there are some limitations to both educational research and teaching practice with respect to the effective use of formative assessment in the science classroom.

First, much of the potentially formative information from assessments can go unused or used in ways that don't improve learning and educational improvement (Hickey, Ingram-Goble & Jameson, 2009).

Most studies of formative assessment only indicate that students who receive feedback on problems learn to solve those problems better than those who do not get feedback. They don't provide evidence of broader learning outcomes or of the potential to transfer feedback information to other contexts. Due to the fact that students are given feedback solving a particular problem, their ability to solve variants of that problem may or may not indicate knowledge of the basic concepts (Messick, 1994). If formative feedback encourages learners to memorize excerpt conceptual definitions and isolated skills in order to succeed on summative assessments, they are presumably left with fragile knowledge that is unlikely to transfer to subsequent learning or testing contexts (Hickey, 2012).

A situative perspective suggests that without evidence of broader learning outcomes and/or productive curricular transformation, many studies of formative assessment only demonstrate that students who receive feedback on problems learn to solve those problems better than students who do not get feedback (Dunn & Mulvenon, 2009).

Much of the literature on formative assessment conceptualizes it as an activity essentially rooted in pedagogical knowledge (e.g., Black and Wiliam 1998c) – i.e., as simply a process of good teaching interaction. Such a conceptualization does not always include reasonably deep cognitive-domain understanding and knowledge of measurement fundamentals. Bennett (2011) has emphasized that any subset of these three competencies is unlikely to work.

2.1.4. Formative assessment methods

According to Bennett (2011), the five key strategies for formative assessment are: a) Sharing Learning Expectations (i.e., clarifying and sharing learning intentions and criteria for success), b) Questioning (i.e., engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning), c) Feedback, d) Self-Assessment (i.e., activating students as the owners of their own learning), and e) Peer Assessment (i.e., activating students as instructional resources for one another). The sources of formative assessment information include students' written, practical, simulation or oral work, the teachers' observations of the students working (especially useful in practical activities and in active learning situations) and the students' reflections on their work (e.g., through reflective diaries, presentations or group reports). Processing formative assessment information can take the form of teachers reading student-written work in their books, posters, charts or notes; teachers listening to students' speech, including their existing ideas, questions and concerns, and the new understandings they were developing; teachers (preferably collaborating) to analyze the ideas and reasoning reflected in students' work or self-reports with a view to (a) interpreting ongoing progress as well as barriers to learning, and (b) making decisions on supplementary scaffolding to support and guide subsequent learning.

Written feedback

One formative assessment method which is often used in teaching practice is written feedback, where the teacher provides written commentary on the work of students (Higgins, Hartley & Skelton, 2002) who are then asked to read the comments and somehow act upon them (Nicol, 2010). In such cases, "feedback has the capacity to turn each item of assessed work into an instrument for the further development of each student's learning" (Hyland, 2000; p. 234). The major advantage of this method is that the teacher is able to provide differentiated feedback for each student, considering each student's commitment, needs and abilities.

On some occasions, written feedback can be provided using work done for summative assessment, despite the fact that summative and formative assessments are designed for different purposes (Black 2001). For example, students can spend time looking at their responses to test papers they have taken in previous lessons and receive teacher guidance in contrasting them to the mark-scheme for each question, also interjecting and questioning the reasoning behind the 'official' answers.

Self and peer assessment

Self- and peer-assessment are key processes for encouraging students to act more autonomously, able to identify their own learning needs, and develop their own next steps. Involving students in self- and peer-assessment can strengthen the level of active engagement in learning and encourage students to take ownership of their own learning and intellectual progress (Harlen, 2010).

Students' self-assessment might include self-monitoring and teacher checking on progress, self-diagnosis and recognition of learning needs, as well as self-reflection on good learning practices.

Involving students in their own assessment gives them direct feedback without waiting to receive it from the teacher (Harlen, 2010). The reflection activities provide opportunities for sense-making and scientific discourse, as well as reflection and self-assessment that support metacognitive self-regulation. Consistent formative assessment that is built into lesson plans and contributes in explicit ways to the summative assessment allows students to set learning goals and monitor their own progress helping them to identify concrete ways on how they can improve. Through developing an increasing awareness of their role in their learning, students ‘become more scientific in their enquiries’ (Lindsay and Clarke, 2001). Students become more self-critical and proactive learners. Students take responsibility for their learning and direct their activities towards their own learning goal because they set it, rather than have it be externally imposed. Students monitor their own learning, progress, and can identify the areas in which they feel confident, and those that they need to develop further. The ability to direct their own learning benefits both them and the classroom as a whole and is, as Harlen (2007) points out, ‘an essential outcome of education’. Furthermore, learners engaged in self-assessment have been found to become more interested in their work and more able to interpret the relevance of what they are doing (Tillema, 2010).

Peer assessment is a particularly useful form of interaction among students (Harlen, 2010) that can maximize shared engagement. When taking part in peer-assessment activities, students are on an equal footing rather than participating in a ‘novice and expert’ situation. Harrison and Harlen (2006) state that students perceive themselves as partners in the teaching-learning process. Many learning theorists emphasize that lasting learning comes from social interaction and co-construction of ideas (Vygotsky, 1986). Students collaboration allows this to happen and group work can also encourage critical thinking (Black & Harrison, 2001) Teachers gain insight into students’ understanding as a result of seeing them reflect on their peers’ assessments. Peer-assessment can help bring misconceptions to the fore, according to Black and Harrison (p47), and increase students’ willingness to present work more clearly. Metacognition, formative assessment, and reflective practice become meaningfully intertwined when individual responses are united with small and large group discussions. A common example of this is the think-pair-share learning strategy (Marshall et al, 2009). Students need to engage in paired-assessment against specific criteria, ideally agreed amongst themselves (Harrison, 2006; Lindsay, 2001). Creating an environment in which students can learn from each other is also important because students learn from each other all the time through observation, trying out their ideas in interaction with their peers or set test questions for their peers.

Structured assessment dialogue

“Structured classroom dialogue" refers to formative feedback that is conducted based on oral activities. A structured dialogue is a disciplined form of dialogue, where participants agree to follow a framework or facilitation, enables groups to address complex problems that are shared by the community (Christensen, 2004). Both open and structured classroom discussions depart from authentic questions or issues that are deemed as relevant by the community (Stewart et al., 1995; Christensen, 2004; Black et al., 2003). Students are asked to express their ideas and opinions on the issue; this serves as a basis for assessment of their understanding. Black and Harrison, 2004, emphasize that it is not trivial to frame good questions that have the potential of stimulating classroom discussion. They therefore advise teachers to devote time and effort in carefully designing the topic of discussion.

Since students should be able to express their ideas and thoughts, it is important to have a "supportive climate" in the classroom, where all students are encouraged to participate freely and are aware of and follow the rules of communication (Stewart et al., 1995; Christensen, 2004; Black et al., 2004). The most important advantages of this approach are that students take ownership of their own learning and that potentially a higher level thinking is achieved through respectful interaction.

An example for such a dialogue is the Socratic Seminar (Adler, 1982; Polite & Adams, 1997; Pihlgren, 2007) where the teacher is the facilitator, not the leader, posing an authentic text or a question and providing scaffolding only where needed. Within the context of the discussion, students listen closely to the comments of others, think critically for themselves and articulate their own thoughts or views and their responses to the thoughts or views of others.

Informal formative assessment

Informal formative assessments are assessments that take place as part of the classroom discourse, but which are not specifically stipulated in the design of the teaching-learning sequence. These include instantaneous feedback as the student takes part in a learning activity and the teacher comments on drafts of student-produced material for inclusion in portfolios (Yorke, 2003). Duschl

(2003) has adopted the term *assessment conversation* to refer to these daily instructional dialogues that embed assessment into an activity already occurring in the classroom. Assessment conversation is a specially formatted instructional dialog that embeds assessment into the activity structure of the classroom. The intent of an assessment conversation is to engage students in the consideration of a diversity of ideas or representations produced by class members and then to employ evidence and age appropriate adaptations of scientific ways of knowing to foster a dialog about what does and does not fit with the emerging thematic structure of the lesson (Duschl & Gitomer, 1997).

Classroom dialogue can provide both teacher and students with ample opportunities to explore, exchange, evaluate and challenge each other's ideas (Mercer & Dawes, 2014). Teachers use questions to probe students' ideas and use the answers to scaffold students' thinking and plan next steps in learning (Gillies & Nichols, 2015). When applying informal formative assessment, the teacher tries to identify learning problems while the learning process is in progress, in order to be able to address and correct misconceptions at an early stage, either by prompting students' thinking or by changing the teaching sequence, introducing unscheduled activities. The terms proactive and reactive are also used to indicate the notion of responsiveness inherent in formative assessment. That is, the teacher could be proactive in deliberately seeking formative assessment information from students or reactive, when they undertake formative assessment in response to information that emerges spontaneously from the students and that the teacher uses in order to promote their learning.

Planned-for-interaction formative assessment is deliberate, according to Shavelson (2008). A teacher plans for and crafts ways to find the gap between what students know and what they need to know. For example, while developing a lesson plan, a teacher may prepare a set of "central questions" that get at the heart of the learning goals for that day's lesson. These questions may be general ("Why do things sink or float?") or more specific ("What is the relationship between mass and volume in floating objects?"). At the right moment during class, the teacher poses these questions, and through a discussion she or he can learn what students know, what evidence they have to back up their knowledge, and what different ideas need to be discussed. This contrasts with typical classroom recitation where teachers use simple questions to "keep the show going". However, despite the existence of planned questions, it is inevitable that during the discussion there will be need for more questions and consequently the dialogue will shift to on the fly interaction (Figure 1).

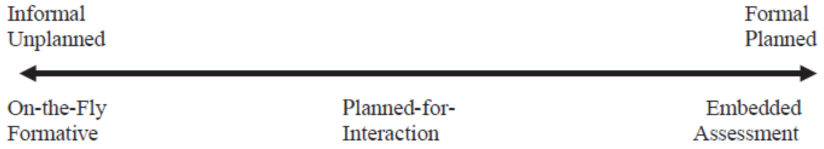

Figure 1: Spectra from informal and unplanned interaction to formal and planned

2.2. On the fly interaction

This study focuses on “on the fly” interaction as an assessment method (Heritage, 2007). Whole class discussions are usually rich with “on the fly” interactions between teacher and students. This type of interaction includes a lot of information that can be used formatively during the classroom discourse. Usually in the case of on the fly interaction, no written record of the information is gathered. Unplanned formative assessments arose from the students’ responses, which often could not be predicted and planned for in advance.

Van Zee and Minstrell (1997) proposed reflective classroom discourse as an alternative to the more conventional classroom discourse which is heavily led by the teacher in an authoritative way. In reflective classroom discourse, students are given opportunities to articulate and express their own ideas and understandings, and to pose questions. Typically in this form of interaction, teachers and students engage in sustained questioning exchanges that support students in articulating and making sense of their ideas and understandings. Students interact with other students in a variety of ways to exchange, challenge and make sense of each other’s ideas.

Classroom dialogue can take place between a teacher and a student, a group of students or the whole class. It can involve informal and spontaneous dialogue, or more formal and planned dialogue (Shavelson et al., 2008). In the context of inquiry lessons, the spontaneous dialogues take place when students are actively engaged in activities such as planning and conducting investigations or discussing their experimental findings. Shavelson et al. (2008) place informal assessment dialogue at one end of a continuum of classroom assessment practices that are used for formative purposes.

On the fly interactions for assessment enable the teacher to identify learning problems during the presentation of information, while there is an opportunity to recognize and correct misconceptions (Bell & Cowie, 2001). To do so, the teacher must find questions to ask that will help students externalise their understanding or confusion. When students provide feedback on their ongoing learning, they are giving the instructor an opportunity to highlight concepts that require additional explanation (McConnell, Steer, Owens, 2003). Sharing with students the multiple ways they have presented or represented scientific evidence or ideas makes it possible to provide feedback on the quality of evidence and ideas put forth by class members (Duschl, 2010). Effective formative assessment provides ‘‘short term feedback so that obstacles can be identified and tackled’’ (Black, 1998, p. 25).

On the fly interactions provide ample opportunities for real time feedback between the teacher, groups of students and individual students, whenever the teacher questions elicit information from students (feedback from student to teacher) and the teacher then uses this information to support students in moving ahead in their learning (feedback from teacher to student) (Ruiz-Primo & Furtak, 2007). It provides teachers with a significant amount of evidence of student learning, that is collected in real time as students engage with inquiry activities. This creates opportunities for teachers to make informed decisions (in real-time) regarding next steps to support students in their learning (Harrison, 2015). This method could be useful when students are engaged (Bell & Cowie, 2001) in an open-ended inquiry (Hickey & Filsecker, 2012; Marshall, Horton & Smart, 2009), which includes more constructivist and sociocultural instruction (Hickey & Filsecker, 2012). The teacher's role is to pose questions and facilitate discussion that results in a consensus view acceptable to the classroom (Duschl, 1997).

Questioning is an important part of assessment conversations and is essential to the success of both formative assessment and science learning. It is one of the key strategies discussed in the literature, with questions asked by both teachers and students. Formative assessment has impact on individual understanding only if the problems are challenging for the students (Hickey, 2011). Different types of questions include: effective questions, open questions, questions for finding out misconceptions and questions as part of feedback to prompt further learning. Examples from the literature on the teacher’s role in design based learning projects refer to formulating prompting questions (Etkina et al. 2010; Linge and Parsons 2006), providing formative feedback (Lyons and Brader 2004;

Maase and High 2008), supporting students in their approach to problem-solving tasks and aiding students in exploring alternatives iteratively (Chang, Yeh Liao, and Chang 2008; Geber et al. 2010).

Close-ended questions enable teachers to check 'if' students know or understand something, while open-ended questions enable teachers to probe 'what' students know and understand (Torrance & Pryor, 2001). Teachers' questions convey to students what the teacher considers to be productive lines of thought or highlight aspects that require further consideration. The type of questions used and the ways in which teachers do this have two important consequences. One is the impact it has on student autonomy, in terms of the degree of freedom to explore and test ideas and decide what to do next. The other is the impact it has on the teacher, as the quality of evidence that is gathered is one of the factors that influences the way in which the teacher assesses and responds to developing student understanding (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p. 41).

Classroom portfolio assessment was the central component of the educational model that Duschl et al (2010) applied. The portfolio served as a repository of students' ideas and findings, which became the basis for classroom discourse and activity. The class worked through a series of iterative cycles in which some form of exploration was conducted, either through demonstration or investigation, and students represent their understanding in some form (e.g., written, oral. graphical, or design product). Once students represent their understanding, the model they used, called for an assessment conversation, which has three general stages:

- The first step is to receive student ideas. This requires that students be allowed to represent their understanding as fully as possible.
- The teacher to recognize the ideas in the classroom in relation to unit or lesson goals.
- Once the diversity is public, the teacher can use the diversity of ideas as a basis for achieving a consensus view in the classroom.

Ruiz-Primo & Furtak (2006) have characterized these assessment conversations as consisting of four elements:

- the teacher elicits student thinking,
- the student provides a response,
- the teacher recognizes somehow the student's response,

- the teacher then uses the information collected to support student learning.

These components are aligned not only to the formative assessment components described (i.e., gathering, interpreting, and acting), but also to the moves (soliciting, responding, reacting) used to describe classroom discourse (Bellack, Kliebard, Hyman, & Smith, cited in Carlsen, 1991). Within ‘‘use,’’ a teacher can provide students with specific information on actions they may take to reach learning goals: ask another question that challenges or redirects the students’ thinking; model communication; promote the exploration and contrast of students’ ideas; make connections between new ideas and familiar ones; recognize a student’s contribution with respect to the topic under discussion; or increase the difficulty of the task at hand (Ruiz-Primo & Furtak, 2007).

2.2.1. Advantages of *on the fly interaction*

Ongoing assessment like “on the fly interaction” allows fine-tuning of instruction since it provides information to be used as feedback to modify the teaching and learning activities in which students are engaged. Using formative assessment helps facilitate more informed, intentional instructional practice (Marshall et al, 2009). It is also the component of teaching in which teachers find out about the effectiveness of the learning activities they are providing and allows them to improve or adjust instruction in order to meet students’ needs. This is in contrast to initiation–response–evaluation (IRE) sequences that involve the teacher initiating a query, the student responding, and the teacher evaluating the student’s contribution (Lemke, 1990), since assessment conversations permit teachers to gather information about the status of students’ conceptions, mental models, strategies, language use, or communication skills with the purpose to use it for guiding their instruction.

Feedback as dialogue means that the student not only receives initial feedback information, but also has the opportunity to engage the teacher in discussion about that feedback. Some researchers maintain that teacher–student dialogue is essential in order to increase the effectiveness of feedback in higher education (Laurillard, 2013). Freeman and Lewis (1998) argue that the teacher ‘should try to stimulate a response and a continuing dialogue—whether this be on the topics that formed the basis of the assignment or aspects of students’ performance or the feedback itself’ (p. 51). Discussions with the teacher help students to develop their understanding of expectations and standards, to check out and correct misunderstandings and to get an immediate response to difficulties.

Dialogue between teachers, students and peers allows for the co-construction of ideas by, for example, a teacher posing questions to elicit what students think. Further discussion elaborates on previous answers helping to advance construction of conceptual knowledge. When teachers paraphrase a student's response, this allows the opportunity to co-construct a response with the teacher and peers (Chin, 2001). Ruiz-Primo and Furtak (2007) showed that the teacher whose whole-class conversations were more consistent with the ESRU cycle described above had students with higher performance on embedded assessments. This result cannot be generalized because of their small sample (3 teachers).

In their review on the educational functions of dialogue, Mercer and Howe (2012) provide evidence that classroom dialogue can develop reasoning and improve academic performance. Not every type of classroom dialogue achieves this, and for it to support learning, students and the teacher must engage in a genuine dialogue, where different ideas are heard and followed up on (Alexander, 2006, p. 28).

In addition, such use of formative assessment can transform a traditional passive lecture into an active learning experience, as it requires that students provide feedback on their ongoing learning (McConnell, Steer, Owens, 2003). The extent to which classroom discourse resembles conversation is in fact an excellent criterion for judging both its instructional quality and the extent of substantive student engagement (Nystrand & Gamoran, 1991). Nystrand & Gamoran also note that, the dialogue values thinking as something students need to do regularly in class in the course of learning.

Classrooms that demand more, intellectually, from students could benefit students and teachers in the following ways:

- They provide more opportunities for students to engage deeply with fundamental concepts and processes based on their own current thinking and, in some cases, are linked to bigger achievement gains in problem solving and conceptual understanding (Carpenter et al., 1989;

Jacobs & Ambrose, 2008; Stein et al, 1996; Stein & Lane, 1996; Wood, 1999; Wood et al., 2006; Vygotsky, 1978).

- They are aligned with practices of mathematicians and scientists, namely the practices of argumentation and the zigzag process of proof and refutation (Ernest, 1991; Goos, 2004; Kline, 1980; Lakatos, 1976; Lampert, 1990; Yackel, 2002).
- They can lead to increased teacher knowledge both of student thinking and the mathematical content itself so she (or he) can better plan and target instruction to students needs (Fennema et al., 1996; Franke et al., 2001; Wiliam, 2007; Wiliam et al., 2004).
- They enable the creation of classroom environments in which students are empowered as doers of science with increased confidence, motivation, agency, persistence and positive identities (Boaler, 2002; Boaler & Greeno, 2000; Empson, 2003; Gresalfi et al., in press; Hiebert et al., 1997; Pierson, 2008a; Stipek et al., 1998).

In Pearson's study (2008), correlational analyses and hierarchical linear modeling (HLM) revealed a statistically significant positive relationship between the responsiveness of teachers' follow-up and student learning of rate and proportionality. ***Responsiveness*** reflects the extent to which teachers "take up" students' thinking and focus on student ideas in their moment-to-moment interactions. This result supports the idea that classroom discourse and normative interaction patterns can guide and influence student learning in ways that improve achievement. There is tremendous and often unrealized power in the ways teachers talk with their students.

2.2.2. Requirements and challenges of applying informal assessment methods

There are a number of challenges that teacher meet when trying to apply such an assessment method. Gitomer and Duschl (1995) described some of the challenges to successful implementation of the assessment conversation. Mastering informal assessment strategies is complex and also introduces significant challenges to the assumptions and methods underlying the current practice of many science teachers.

To use formative assessment effectively in the classroom, teachers need skills and four basic elements of knowledge that are critical: 1) domain knowledge, 2) pedagogical content knowledge, 3) knowledge of students' current understanding and abilities, and 4) knowledge of how formative assessment works (Heritage, 2007).

It is very important that both teachers and students have a clear idea of the goals of their work. The clarity of goals is an essential requirement for formative assessment (Harlen, 2010). With domain knowledge, teachers are able to define a learning progression of subgoals toward the desired learning that will act as the framework to guide assessment and instruction. Without a strong base of domain knowledge there is a danger that teachers' analyses will focus on the surface aspects of learning at the expense of deeper levels of understanding. An inaccurate analysis of the students' learning status will lead to wrong decisions about what the next instructional steps should be.

Conceptual dilemmas in engaging in constructivism can be seen in the tensions teachers feel in focusing on cognitive versus sociocultural forms of formative assessment. For example, it may be easier to assess learning as right or wrong relative to a learning objective – a cognitive approach – instead of mapping student work onto a range of student ideas, prior understandings, and linguistic understandings – a sociocultural approach; however, the information gleaned from the first example provides only limited insight to inform instruction whereas the latter example may provide more meaningful information on where students are at (Otero, 2008).

Coffey et al. (2011) analysed four publications by prominent researchers in assessment and science education to depict that, there is a disciplinary substance of what teachers and students assess, while researchers and teacher educators seem to believe that strategies are what teachers need first or most, to help them engage in formative assessment. Strategies are the means for teachers to be aware and to understand the students' understandings and progress. As they note, formative assessment should be understood as a matter of attention to disciplinary substance, and in this sense it should be inherent throughout classroom activity, not restricted to specifically designate ‘‘assessment activities’’. They state that it is not sufficient to consider only the teacher's actions but also, to appreciate the quality of a teacher's awareness, it is essential to consider disciplinary substance: What is happening in the class, and of that, what does the teacher notice and consider? Teacher noticing involves two main processes, attending to specific events within classroom interactions and then making sense of these events within the instructional setting, in order to broaden the range of responses a teacher has in their repertoire (Sherin, Jacobs, & Phillip, 2011).

Noticing, understanding, and acting on student responses are linked: Understanding student responses requires detailed noticing of literacy and metacognitive behaviors. Acting on this information engages the teacher and student in interactive work so that student understanding or confusion becomes more apparent. Interactive, adaptive teaching, supported by accurate and comprehensive noticing, allows teachers to engage in a generative cycle supporting development of pedagogical expertise (Gibson & Ross, 2016).

Noticing encompasses the ways in which teachers are able, during instruction, to observe important details in students' responses, and interpret this information accurately and comprehensively to adapt instruction in the moment (Gibson and Ross, 2016). The discipline of noticing is an established field of research that theorises how teachers learn to develop their practice, so they become more sensitive to the implications of student actions and the possibilities for acting creatively on the fly. Mason (2002) outlines this as a systematic process whereby noticing and its development involves consideration of what is deemed significant in a classroom situation, to then make sense of and reason about what is observed by making a connection with and drawing on relevant knowledge. There is a substantial body of literature on teacher noticing, which focuses on teacher sense-making and use of evidence of student thinking as it unfolds (Jacobs, Lamb, & Philipp, 2010). Noticing can be considered as a responsive act that invites action, which suggests an inclusive, dynamic and purposeful response to evidence of student ideas and interests. The value of Mason's formulation is his focus on noticing as central to being better prepared to choose to respond in the moment creatively rather than react out of habit (Mason, 2002).

On the other hand, Cazden (2001) claims that it is hard to find discussions where ideas are explored instead of evaluated, where teachers talk less than students and where students address each other directly. A reason for that is the fact that, in a typical class, the teacher has the authority of who will speak and when. Another is the way of asking and the fast pace of questioning that a teacher follows (Dillon, 1983). Although teachers' discursive moves can fulfill a wide variety of functions, responsive are those which validate students' contributions by taking up and extending them or inviting students to do so themselves (Jacobs & Empson, 2016).

Revoicing can position students in intellectual roles consistent with practices in the discipline; it can reformulate ideas through clarification, introducing new terminology, or highlighting a specific aspect; and it can provide an opportunity to explicate one's reasoning (O'Connor &

Michaels, 1993). Revoicing is not synonymous with high levels of responsiveness; it is a linguistic construct that can be categorized with respect to responsiveness. In other words, each revoicing move exists somewhere on a continuum from low to high responsiveness (as does any other linguistic construct). For example, rebroadcasting (an echoing move) is a form of revoicing with low responsiveness, whereas truly revoicing a student idea is a highly responsive move. Highly responsive follow-up moves afford teachers opportunities to gather evidence of student understanding in order to adjust instruction. In other words, responsiveness enables teachers to work within students' ZPDs by assessing for learning as opposed to the assessment of learning (Wiliam, 2007).

Successful work within one's ZPD requires a more capable other continually responding and reacting to the learner as they move toward new possibilities and deeper understanding. The teacher must be willing to adjust and change planned activities based on current student needs; this type of interaction exemplifies responsiveness at its best. Finally, responsiveness, by definition, elicits student ideas and thinking. This is good both for the speaker whose ideas are drawn out and for the larger math community who is exposed to these ideas. By encouraging students to verbalize their thinking and provide explanations of their ideas, discursive moves with high levels of responsiveness (and intellectual work as well) can support coherence and clarity in thinking, help the speaker plan and regulate a course of action, encourage the organization and integration of new ideas into prior experience, and expose errors in reasoning (Chi, Lewis, Reimann, & Glaser, 1989; Vygotsky 1986). Eliciting and reflecting back the speaker's thoughts externalizes them, providing an opportunity for reflection and increased metacognitive awareness.

As developed by Robin Alexander (2006), dialogic teaching harnesses the power of talk to stimulate and extend students' thinking and improve their learning and understanding. It helps the teacher to diagnose students' needs more precisely, frame their learning tasks and assess their progress. It empowers the student for lifelong learning and active citizenship. Dialogic teaching is not just any talk. It is as distinct from the question-answer and listen-tell routines of traditional teaching as it is from the casual conversation of informal discussion. It requires:

- Interactions, which encourage students to think, and to think in different ways
- Questions, which invite much more than simple recall

- Answers, which are justified, followed up and built upon rather than merely received
- Feedback, which informs and leads thinking forward as well as encourages contributions which are extended rather than fragmented exchanges, which chain together into coherent and deepening lines of enquiry
- Discussion and argumentation, which probe and challenge rather than unquestioningly accept
- Professional engagement with subject matter, which liberates classroom discourse from the safe and conventional
- Classroom organisation, climate and relationships, which make all this possible.

Duschl et al. (2010) identified conceptual conflicts that emerged from the display of student work and that teachers had difficulties on how to (a) recognize and (b) attend to and manage in a science classroom. Perhaps even more important is the need for the students to be given opportunities to develop the habits of mind used to examine and evaluate knowledge claims.

To effectively adapt instruction to student learning, teachers' pedagogical content knowledge must include familiarity with multiple models of teaching for student achievement in a specific domain and knowledge of which model of teaching is appropriate for what purpose. The gap between current status and learning goals will differ from student to student, so teachers will need differentiated instructional strategies and knowledge of how to use them in the classroom. Teachers are not used to using student information to guide and revise instructional decision-making (Duschl, 1997) and thus they have difficulties in adapting their teaching plan accordingly.

The professional knowledge and experiences of the teachers are seen as important in attending to some sources of information (rather than others), in interpreting the elicited information, and in taking action. This professional knowledge and experience includes the teachers’ knowledge and experiences of the topic, of the students as learners, and is enriched from having taught the unit of work before (Bell and Cowie, 2001).

Bell and Cowie (2000) investigated 11 teachers’ developing assessment practices over the two years of a project, while teaching to take into account students’ thinking. Formative assessment tended to be informal, with no written record of the information gathered. The information was used in the teaching and learning in the classroom for building up a picture of the student learning by the teacher. Interviewing the teachers, they said that they had to manage the degree of responsiveness when doing formative assessment. They were aware that they had to manage the behavior and learning of the whole class as well as that of individuals. They also had to manage attending to the students investigating their own interests and ideas, and to the students learning what was listed in the curriculum. In both these situations, responding to one aspect meant that they could not respond to the other. They could not always be as responsive to a situation as they wished or were able to. The nature of these dilemmas was evident in the discussions on the teacher development days, on the tensions between formatively assessing the class or an individual; between formatively assessing the science or the personal and social development; between formatively assessing the science in the curriculum and the science outside the curriculum; and between the different purposes for eliciting and taking action.

A significant challenge posed to teachers is that assessments need to occur on multiple fronts. Students need feedback on their developing understanding of the core science concepts, the characteristics of the emerging science explanation, the reasoning they employ when considering evidence and relating it to explanations, and the ways in which they choose to represent and report scientific information and knowledge claims (Duschl, 2010). What are often missing are the processes of science that address argumentation and the social dynamics of the classroom that stress the management and assessment of information and ideas. For the inability of teachers to engage students in meaning-making and reasoning has much to do with confusion surrounding how to manage the flow of information, knowledge claims, and ideas produced by students.

Teachers need to be flexible enough to adjust their lesson plans appropriately during the instruction based on the assessment, so as to meet students' learning needs (Lewis, 2008; Black & Wiliam, 1998). Researchers in the domain of adaptive teaching hold that, classrooms are highly dynamic, unpredictable and constantly changing environments that require more than established routines and procedures. Teachers need to be thoughtful decision-makers who think on their feet to appropriately adjust instruction to circumstances (Corno, 2008). Adaptive teaching requires informal, responsive, ongoing assessments wrapped into teaching to devise activities to overcome impediments to learning. Teachers assess and revise instruction in a continual evaluation process that includes direct observation of students' responses in the moment. Adaptive teachers have a mindset for adaptive teaching and tend to view learner variation as an opportunity, rather than an obstacle to be overcome.

Teachers' view of science and their concomitant view of teaching science are dominated by tasks and activities rather than conceptual structures and scientific reasoning. Thus, steps of the assessment conversation that focus on activity (e.g., drawing and presenting an explanation) are more readily mastered than those that focus on conceptual structures and reasoning (e.g., relating evidence and applying criteria to student explanations) (Duschl, 2010).

Among these challenges is being sensitive to students’ present level of understanding so that instruction can be continuously modified while learning is still taking place. This means that continuous assessment of students’ understanding to improve teaching and learning is required (Ruiz-Pinto & Furtak, 2006). Students' previous learning includes: 1) their level of knowledge in a specific content area, 2) their understanding of concepts in the content area (i.e., the degree to which they can make generalizations through a process of excerption from a number of discrete examples), 3) the level of their skills specific to the content area (i.e., the capacity or competence to perform a task), 4) the attitudes the students are developing (e.g., the value the students place on the subject, the interest they display, and their levels of initiative and self-reliance), and 5) their level of language proficiency (Heritage, 2007).

Students’ conceptions of science assessment and their conceptions of science learning need to be taken into consideration. Lee et al. (2008) suggests that students’ conceptions of science assessment may be interrelated with their own conceptions of science learning and in turn influence their approaches to science learning. Students with conceptions of science assessment as “improving learning,” “problem solving,” and “critical judgment”, that could be categorized as formative conceptions of science assessment, tend to view science assessment as a way of processing higher order science thinking or problem-solving skills (Lee, 2013). These students may value the feedback as improving their learning and learning ability by identifying whether they have integrated and refined the scientific knowledge and then extended it to other situations (Lee, 2013).

Science teachers belong to a community of people who already speak the language of science. However, students, at least for a long time, do not. Teachers use that language to make sense of each topic in a particular way. Students use their own language to express a view of the subject that can be very different. This is the reason why communicating science can be so difficult. Teachers need to learn to see science teaching as a social process and to bring students, at least partially, into this community of people who talk science (Lemke, 1990). Newmann (1992) makes a plea, on similar grounds, for assessment in social studies to focus on discourse, defined by him as language produced by the student with the intention of giving narrative, argument, explanation or analysis. The plea is based on an argument that current methods, in which students are constrained to use the language of others, undermine the constructive use of discourse and so trivialize social knowledge.

Teachers need to understand that the quality of the assessment is an important concern. Gotwals and Birmingham (2015), using a comparative case study with multiple data sources at two instances in time (Yin 2009), found that most teacher candidates conceptualized students' ideas in a dichotomous fashion: either correct or a misconception. There were very few instances of identifying student ideas that, while not what the teacher candidate was anticipating in connection to their question could be considered productive or resourceful.

The challenges of teaching and of managing a classroom learning environment are significantly altered when one is asked to receive, recognize, and then use student-generated information for conducting assessments on a frequent basis. A challenge to the successful implementation of assessment-driven instruction is the need to initiate and sustain a learning environment that emerges out of students' personal efforts, products, and ideas (Duschl, 2010).

First, the classroom needs to be a place where all students feel that they are respected and valued and that they have an important role to play. Second, teachers must have the skills to build a community of learners, characterized by a recognition and appreciation of individual differences. Classroom norms of listening respectfully to one another, responding positively and constructively, and appreciating the different skill levels among peers will enable all students to feel safe in the learning environment and to learn with and from one another (Heritage, 2007).

The ways the formative assessment information is elicited, interpreted, and acted on is influenced by the learning situations used (whole class, small groups, or individuals); by the learning activities chosen (e.g., brainstorms, investigations, watching a video, and library projects); the teacher's knowledge of the students; the professional knowledge and skills of the teacher; the topic of the lesson and the teacher's purposes for the lesson. These characteristics suggest that formative assessment may be seen as a sociocultural and discursive activity (Bell, 2000; Bell & Cowie, 2001).

Teachers are not always consciously aware of doing formative assessment, and in particular unplanned or interactive formative assessment. However, it seems that those teachers thinking about formative assessment, become more aware of their professional knowledge and skills and more able to use these in the formative assessment process in the classroom (Harlen, 2010).

2.3. Classroom discourse analysis

In this section, we will present efforts identified in literature to code, analyse and describe informal formative assessment that unfolds during classroom dialogue, as part of classroom discourse.

Classroom talk in general has been established as a legitimate object of study (Edwards & Westgate, 1994). Research in the process–product tradition focused on the conceptual level of teacher questions (Redfield & Rousseau, 1981) and other teacher ‘‘behaviors’’ that might be correlated with measures of student learning (Brophy & Good, 1986). Studies have placed teacher–student discourse in context by examining authority structures, the responsiveness of the teacher-to-student contributions, and patterns in classroom talk (Cazden, 2001; Edwards &Mercer, 1987; Lemke, 1990; Scott, 1998).

Student-teacher interaction can be named as discussion and not a question-answer sequence, when there is a considerably long exchange of information among the teacher and at least 3 students. Discussion displays regular uptake as long as participants listen and respond accordingly to each other (Nystrand & Gamoran, 1991).

One of the patterns of student–teacher interaction that has become the subject of extensive discussion has been alternately described as IRE (Initiation, Response, Evaluation) or IRF (Initiation, Response, Feedback) (Sinclair and Coulthard, 1975). In this sequence, the teacher initiates a query, a student responds, and the teacher provides a form of evaluation or generic feedback to the student (Cazden, 2001). The IRE and IRF sequences are characterized by the teacher often asking ‘‘inauthentic questions’’ (Nystrand & Gamoran, 1991) in which the answer is already known by the teacher, sometimes for the sake of making the classroom conversation appear more like a dialogue than a monologue. Teaching practices that constrain students within IRE/F patterns have been criticized because they involve students in ‘‘procedural’’ rather than ‘‘authentic’’ engagement (Nystrand & Gamoran, 1991). Nevertheless, these researchers classified teachers’ evaluation to students responses as “high level evaluation”, when it consists of teacher’s certification of response and incorporation of the response either by employing an elaboration or a follow up question.

Analyses of classroom dialogue using conversational analysis have revealed that the most common interactional exchange between a teacher and a student in a classroom follows the Initiation-Response-Feedback (IRF) pattern (Mercer & Dawes, 2014). In this pattern, the teacher frequently asks a close-ended question (initiation) to which the student provides a typically brief response, which is then evaluated by the teacher (providing evaluative feedback to the student). Torrance and Pryor (2001) associate the IRF pattern of ‘dialogue’ with convergent formative assessment practices where the teacher is checking ‘if’ students know or understand something, rather than exploring ‘what’ they know and understand. From the perspective of formative assessment, this type of “dialogue” is of little use to both student and the teacher because it leaves the student with no useful feedback on how to move forward and the teacher with a very limited understanding of what the student’s learning needs are (Black & William, 1998).

Regarding initiation, Dillon (1983) identified four alternative prompt categories for teacher to manage the classroom dialogue, apart from the usual teacher questions:

- Declarative statements
- Reflective restatements
- Invitations to elaborate
- Silence

On the other hand, student's responses to questioning will depend on a number of factors. Whether the student believes that his/her abilities are incremental or fixed will have a strong influence on how the student sees a question - as an opportunity to learn or as a threat to self-esteem (Dweck, 1986). Other observational studies of children’s talk in groups (Fisher, 1992; Dawes, Fisher & Mercer, 1992; Mercer, 1994; 1995) also created a typology, by which they described children’s talk as being more or less like three archetypical forms: Disputational, Cumulative and Exploratory.

Some alternative students’ responses have been classified by Malcolm (1979) as:

- Empty bidding – followed by silence.
- Declined replying – after a direct elicitation.
- Deferred replying – after a longer than normal pause.
- Shadowed replying – in the shadow of the next speaker.
- Unsolicited replying – without having been nominated.

However, a range of discursive moves that expert teachers make within a coherent framework of broader purposes that can be productively used to support teacher learning have been identified, like the ECR technique (Elicit – Confront – Resolve) (Cazden, 2001). The effective orchestration of these moves involves a balance between the exercise of authority by the teacher to introduce and establish scientific knowledge at the same time as allowing room for students to explore the meaning of these often new and challenging ideas, in their own language and terms (Tytler and Aranda, 2015).

Another kind of discursive moves identified by researchers is that teachers often elaborate and reformulate the contributions made to classroom dialogue by students (for example in response to a teacher’s questions) as a way of clarifying what has been said for the benefit of others and also to make connections between the content of children’s utterances and the technical terminology of the curriculum (Mercer, 2004). Mercer compared classrooms in which teachers were asked to use a specially-designed program of discourse strategies and activities with classrooms in which no interventions were made. Through a sociocultural discourse analysis she was able to examine how teachers and students use language to introduce new information, orientate to each other’s perspectives and understandings and pursue joint plans of action. A list of techniques the teachers use as identified from Mercer (1995) was:

To elicit knowledge from learners:

- Direct elicitations
- Cued elicitations

To respond to what learners say:

- Confirmations
- Repetitions
- Elaborations
- Reformulations

To describe significant aspects of shared experiences

- ‘We’ statements
- Literal recaps
- Reconstructive recaps

Other researchers noticed a shift on the discursive moves the teacher employed during the lesson. More specifically, Roychoudhury and Roth (1996) studied classroom dialogue focusing on one teacher and a group of students following a set of open-inquiry laboratory classes (part of an introductory physics course for high school juniors). Their findings show that teacher-student dialogue changed across the investigation stage. During the planning and data interpretation stages, the teacher used questioning and comments mainly to support students’ conceptual understanding of subject matter. In the data collection stage, the teacher acted as a “research advisor” using questioning and comments to provide guidance on experimental procedures.

Ruiz-Primo & Furtak (2006) developed an ESRU model (teacher’s Elicitation, Student’s response, teacher’s Recognition of the response, teacher Uses the information collected) for coding and describing teacher – student interaction during informal assessment. This model can be a useful way of distinguishing those teacher–student interactions that go beyond the generic description of ‘‘feedback’’; it is the final step of using information about students’ learning that distinguishes ESRU cycles from IRE/F sequences. “Using” implies more than providing evaluation or generic forms of feedback to learners, but rather involves helping students move toward learning goals.

In a study of middle school mathematics Pierson (2008) examined teachers’ responsiveness to student ideas. She defined ‘‘responsiveness’’ as ‘‘the extent to which teachers ‘take up’ students’ thinking and focus on student ideas in their moment-to-moment interactions,’’ and in particular ‘‘High II’’ responsiveness in which the focus is on the students’ meaning and logic for the immediate purpose of understanding their reasoning on its own terms. She distinguished High II from High I responsiveness, in which the teacher worked to identify student ideas with the purpose of correcting them.

Tytler and Aranda (2015) analysed video sequences from five expert elementary teachers across three countries to develop a coding scheme for these teachers' 'discursive moves' to guide and respond to student inputs, that unpacks more completely the strategies they use to develop interactive discussion. The analysis showed varied patterns of knowledge transaction, with teacher discursive moves serving three broad purposes: to elicit and acknowledge student responses, to clarify and to extend student ideas. The patterns of talk were also related to the dialogic-authoritative distinction in analysis of talk, to show that this distinction is only clear for particular types of expert practice. The analysis of patterns of discursive moves made by these teachers representing expertise in disparate settings has identified a discursive richness that goes well beyond the classic simple IRE sequence described so often in the literature, in which student responses are narrowly channeled and there is little room given for elaboration of student ideas. The moves have been identified as falling within three broad categories of purpose—Eliciting/Acknowledging, Clarifying and Extending—which were all well represented in these expert teachers' practices, and provide shape to the conceptual intent of lessons.

Nieminen et al. (2015) adapted from Torrance and Pryor (2001) and Alexander (2006) a coding system by developing two categories to describe teachers' use of questions or follow up comments (TD- Divergent and TC-Convergent). Teacher Divergent talk refers to all instances in the dialogue where the teacher asks questions to probe and encourage students' thinking and to promote discussion. The questions used are open-ended (typically, how? and why?), and intended to promote higher order thinking. The focus of this type of talk is not only to unveil what the student is thinking but also create an environment where the thinking can be extended, so ultimately it is about opening up and keeping the discussion going. Teacher Convergent talk, on the other hand, refers to all instances in the dialogue where the teacher asks questions to check if students are on an intended path, and if they know something. The questions used in this context are typically close-ended questions that promote lower order thinking (mostly recall of factual knowledge). As the focus of this type of talk is to check if students are on an intended path, it is likely that this type of talk reveals repeating or reformulating the same questions until students give the answer the teacher is looking for.

The aforementioned studies indicate that there is an evolution on where relevant research focus is, moving from identifying various types of questioning by the teachers and students' responses, towards an effort to understand the impact of these types of questions on the dialogue and on the effectiveness of formative assessing students' understanding to promote their learning. We will next examine the use of concept mapping as a tool to describe the classroom dialogue.

2.4. Concept maps

Concept maps are often used in education as powerful tools to represent knowledge structures in all subject matter fields and for learners of any age (Novak & Gowin, 1984). Concept mapping was originally developed by Novak and the members of his research group as a means to representing conceptual frameworks and the interrelationships between concepts (Stewart et al., 1979, Novak and Gowin, 1984). Researchers have continued to develop and refine this representational technique for use in teaching, learning, research and assessment. Concept maps have been used for many instructional purposes, in many subjects, and with many levels of students. In some cases, students are asked to build their concept maps themselves (Regis et al., 1996), whereas, in other cases, teachers or researchers construct the concept maps when assessing students, for example by interviewing them (Nakhleh, 1994).

Concept maps have been used in the study of physics (Roth & Roycoudhury, 1994; Gangosa, 1996), chemistry (Markow & Lonning, 1998), ecology and environmental education (Heinze-Fry, 1997), biology (Coleman, 1998), astronomy (Zeilik et al., 1997), engineering (Moreira & Greca, 1996), geology (Gonzalez, 1993) and mathematics (Moreira & Motta, 1993).

The basic element of a concept map consists of concept words or phrases that are connected together with linking words or phrases to form complete thoughts called 'propositions' (in the format, concept → linking word → concept) (Nicoll, 2001). Concept maps are built by placing terms, which represent the concepts to be mapped, in structures called nodes. The nodes are then linked together into propositions to show how students connect or link the concepts.

While concept maps are a qualitative representation of students’ conceptual understanding, researchers have attempted to use a variety of scoring techniques on concept maps to be able to quantitate the trends among concept maps. There exists a wide variety of ways to generate and subsequently grade or assess concept maps (Stewart 1980, Moreira 1985, Raven 1985, Stuart 1985, Shavelson et al. 1993, Liu 1994). Liu (1994), for instance, proposed using item response theory, which takes into account the number of links, the number of hierarchies, the number of cross-links, and the number of examples when scoring concept maps. The presence of crosslinks in a map reflects the extent of knowledge integration.

The approaches to scoring concept maps generally combine an interest in the content validity or accuracy of the content displayed in the map with an interest in the elaborateness of the map as measured by counting various map components, such as concepts or links. Early scoring systems tended to place much emphasis on elaborateness. Novak and Gowin (1984) originally proposed a scoring system in which the number of valid propositions, levels of hierarchy, examples, and crosslinks are counted. Each of these counts is given a weight (for instance levels of hierarchy might be multiplied by 5, while number of valid propositions might be multiplied by 1), and then the weighted counts would be added to obtain a final score. More recent scoring systems show a trend towards more sophisticated ways to assess a concept map’s content validity with a relative de-emphasis on a count of map components (Ruiz-Primo et al., 1997, Rice et al., 1998), the use of what could be called an expert link matrix. This latter option consists of a process in which one or more experts on the given topic produce an exhaustive set of possible relationships between each pair of concepts in the allowed set. These possible relationships can then be categorized in various ways.

Finally, statistical analysis can be performed on the data. To evaluate the quality of student learning, Stoddart et al. (2000) calculated the proportion of scientifically accurate propositions relative to all propositions, and the proportion of higher order explanation propositions (how or why) relative to all propositions, in each concept map. Factual statements, often answering ‘what’ questions were defined as basic descriptions while explanations that describe function or purpose were defined as higher-order ones. For the complexity of the proposition, a structure criterion is used to assess the elaboration of an idea within a proposition, using two levels which are (i) simple, which is a proposition containing only one subject-object clause, and (ii) compound, a proposition containing one or more dependent clauses.

Ruiz-Primo & Shavelson (1997) developed a proposition for each link in the criterion map. The labeled links between pairs of concepts provided by the teachers, the expert and the researchers varied in the quality of their explication of the relationship. For example, the propositions used by the expert more completely explained the links between concept pairs than those used by the teachers. To account for the variation in the quality of the proposition, they developed a Propositions Inventory, which classified each proposition into one of five categories: Accurate Excellent, Accurate Good, Accurate Poor, "Don't Care" and Inaccurate. The scoring system, based on the criterion map and the Propositions Inventory, evaluated two components of the map: the propositions and the nodes. The accuracy of each proposition in a student's map was assessed on a five-level scale (from 0 for inaccurate to 4 for accurate excellent) according to the Propositions Inventory. The concepts used as nodes were noted, counted, and classified as contained/not contained in their list of 20 key concepts. Then, three map scores were formed: (1) a total proposition accuracy score - the total sum of the scores obtained on all propositions; (2) convergence score - the proportion of valid propositions in the student's map out of all possible propositions in the criterion map (i.e., the degree to which the student's map and the criterion map converge); (3) salience score, the proportion of valid propositions out of all the propositions in the student's map. Total proposition accuracy scores were based on an evaluation of the quality of propositions that students constructed.

As it is clear from the aforementioned studies, there are numerus ways that concept maps are developed and analysed in order to examine students’ understanding. However, no study was found in literature that has employed concept mapping for illustrating the classroom dialogue as it evolved during a lesson. In any case, we believe that the process of developing and analyzing concept maps could be adopted for the purpose of the present study; to depict the conceptual completeness, the depth and coherence of the classroom dialogue.

2.5. Conceptual completeness, depth and coherence of the classroom dialogue

Assuming that knowledge within a content domain is organized around central concepts, to be knowledgeable in the domain implies a highly integrated conceptual structure (Ruiz-Primo, 1997).

According to Ausubel's assimilation theory, (a) meaningful learning involves the assimilation of new concepts and propositions into existing cognitive structure, modifying those structures and (b), knowledge is organized hierarchically in cognitive structure and most new learning involves incorporation of concepts and propositions into existing hierarchies ((Ivie, 1998).

It has been shown that the ways in which scientific understandings develop tend to follow ‘common conceptual trajectories’ (Driver et al., 1994). An awareness of these trajectories allows certain developmental pathways to be anticipated by the teacher. From this, an awareness and respect of these trajectories during the evolution of classroom dialogue should promote meaningful learning. This kind of dialogue could make it possible for pre-instructional conceptual structures of the learners to be fundamentally restructured in order to allow understanding of the intended knowledge, that is, the acquisition of science concepts (Duit & Treagust, 2003).

In this study, we will refer to ***conceptual completeness*** of classroom dialogue to describe the degree in which, the essential concepts for discussing a scientific phenomenon were brought into the classroom talk when investigating the particular phenomenon.

There appears to be no study which found that a particular student's conception could be completely extinguished and then replaced by the science view. Indeed, most studies show that the old ideas stay alive in particular contexts. Usually the best that could be achieved was a 'peripheral conceptual change' (Chinn & Brewer, 1993) in that parts of the initial idea merge with parts of the new idea to form some sort of hybrid idea (Jung 1993). Consequently, it is of great importance that important concepts relating to a scientific phenomenon under discussion and their links to other concepts are revisited during the classroom dialogue. We will refer to ***conceptual depth*** of the dialogue in this study to express at what extend there where multiple references on connections between the related concepts during an episode of classroom dialogue.

Conceptual change involves not only change in specific beliefs and presuppositions but also requires the construction of theoretical frameworks with greater systematicity, coherence, and explanatory power (Vosniadou & Ioannides, 1998). Thus, it is of great importance that, during the classroom dialogue, related concepts are linked not only one to another but to as many relevant ones as possible, in a way that the discussion develops a meaningful and coherent network of concepts. We will use the term ***conceptual coherence*** of the dialogue to represent the extensiveness of linking concepts in a coherent network.

2.6. Scientific Inquiry

In this study, the method of "on the fly interactions" is implemented in the context of scientific inquiry. Scientific inquiry from the perspective of science education is the learning framework that consists of providing students opportunities to experience the natural phenomena, to develop their understanding about the mechanisms that underlie natural systems, to practise their abilities on designing and carrying out investigations, to improve their thinking strategies for analysing data and formulating conclusions, to communicate experimental results and to develop epistemological awareness about the nature of science along with positive attitudes towards learning with evidence. Scientific inquiry learning has been identified as an essential element of students' experience in science education (National Research Council, 2001).

In inquiry lessons, students sometimes engage in investigating disciplinary-orientated questions where they learn to seek evidence and use it to make decisions (Anderson, 2002). In open-ended and student-led inquiry, students are encouraged to explore phenomena and engage with their own ideas and those of their peers as they work collaboratively with other students throughout the inquiry process (Wenning, 2005). In open-ended and student-led inquiry, students engage in decision-making such as, how to approach a problem and make decisions on what to investigate, how to go about it, how to make sense of emergent findings, and adjust an investigation accordingly. The very nature of open-ended inquiry introduces a considerable degree of unpredictability for both teacher and the students. Therefore, opportunities need to be created to encourage students to reflect and reconsider their plans and actions as they proceed with an inquiry. These opportunities enable students to compare their thinking with others in their group. Therefore, teacher-student and student-student on the fly interactions (through questioning and answering, follow-up, and reflection) create opportunities to assess emergent learning and respond to students' learning needs, as the inquiry unfolds.

A key characteristic of effective informal formative assessment is to promote frequent assessment conversations that may allow teachers to 'listen' to inquiry (Duschl, 2003). Listening to inquiry should focus on helping students ''*examine how scientists have come to know what they believe to* *be scientific knowledge and why they believe this knowledge over other competing knowledge claims*'' (Duschl, 2003, p. 53). Therefore, informal formative assessment that facilitates inquiry should: (1) involve discussions in which students share their thinking, beliefs, ideas, and products; (2) allow teachers to acknowledge student participation; and (3) allow teachers to use students' participation as the springboard to develop questions and/or activities that can promote their learning (Ruiz-Primo & Furtak, 2007).

2.7. The need for this research

As has been described earlier, enacting an informal method of formative assessment like "on the fly interactions" is very challenging. Assessment specialists or curriculum designers cannot expect teachers to use formative assessment effectively without training. Capacity for student argumentation and teacher elicitation of student evidence needs to be strengthened through professional development, and curriculum designers and assessment developers should adopt a comprehensive theory of teacher feedback to students in order to facilitate this professional development (Ayala, 2008).

Thus, if we can better capture what effective informal formative assessment looks like in science classrooms, we can help other teachers improve their own practices through pre-service training and professional development. As the science education community continues to struggle to define what it means to be an instructionally responsive teacher in the context of scientific inquiry, being explicit about the differences between IRE/F and ESRU managing of classroom dialogue could help teachers to understand the differences between asking questions for the purpose of recitation and asking questions for the purpose of eliciting information with the aim of improving student learning. Pre- and in-service teachers alike may benefit from learning about the ESRU cycle as a way of thinking about classroom discussions as assessment conversations, or opportunities to understand the students' understanding and to move students toward learning goals either by prompting students thinking or by adapting their instruction towards their needs.

As the developers of the ESRU coding technique suggest, future research should build on the methodology in order to explore discussions that are taking place in the context of inquiry based learning (Ruiz-Primo and Furtak, 2006). Further interpretation of complete and incomplete ESRU cycles, in association with information about the quality of student responses, could supply valuable information about the characteristics of scientific inquiry taking place in classrooms. In addition, other educational scientists have identified the need for research to focus on how the disciplinary substance is addressed during the employment of informal assessment methods in science lessons (Hammer, 2011).

For these reasons, it is of great importance to investigate what facilitates effective use of on the fly interactions as a mean for formative assessment in science classrooms, what are the restrictions that appear and how teachers could be supported in order to improve their own practices.

2.7.1. Research Questions

This study aims to answer the following research questions:

1. What patterns can we identify in *on the fly interactions* between the teacher and the students when the teacher is interested in using the available information for formative assessment?
2. What factors facilitate or impede teachers' attempts to use *interactions on the fly* in order to guide students towards the learning goals?
    - What are the emergent factors that seem to afford productive use of the collected information?
    - What are the various types of missed opportunities encountered in the interactions on the fly and what are possible interpretations for why these opportunities were missed by the teacher?
3. Can we represent the conceptual completeness, the depth and coherence of the classroom dialogue when interaction on the fly is applied using concept maps?
4. How do identified patterns of *on the fly interactions* relate to the conceptual coherence and depth of the classroom dialogue?

3. METHODOLOGY

We addressed these questions through classroom-based research, in which we organized a learning environment focusing on the competence of designing and implementing investigations. Investigation is the process that places theoretical knowledge claims to the empirical test in order to test existing theories and hypotheses or to revise and develop new ones, by examining causal or other relations between variables that are of relevance to the phenomena under investigation (Kyriazi, 2004).

3.1. Teaching interventions

The teaching enactments were implemented during school year 2014-2015, in three different school contexts in order to have as much diverse input as possible from secondary education in Cyprus. The physics teachers involved in this study volunteered to participate with their intact classes in classroom implementation of formative assessment methods developed by the EU-funded project *Assess Inquiry in Science, Technology and Mathematics Education* (ASSIST-ME) (http://assistme.ku.dk/). The particular teaching-learning sequence that was used in this study was grounded in the competence of designing and implementing investigations.

The first implementation took place at a vocational school, in a class of twelve 15-16 years old students. The Physics teacher of this class holds a bachelor in Physics and a master in Science Education. She had more than ten years of experience in teaching physics, most of them in secondary education and in Vocational schools particularly. She often follows seminars for her professional development and she tries to apply in her teaching practice teaching innovations.

The first implementation lasted ten 45-minute teaching periods. The first part of this (six periods) was in the context of "freefall", that is the vertical motion of an object when the only significant force that is acted on the object is the gravitational force. In particular, the students designed and implemented investigations to examine variables that might influence the acceleration of a falling object, like the mass of the object, the height that it is released from, the shape of the object and its volume. The aim was to help students improve their skills on designing and implementing investigations with appropriate control of variables, but also to be able to interpret the measurements they would produce in a meaningful way. Also, the teacher applied their conclusions to the 2nd Newton's law to prove why acceleration is not influenced by the mass of the object. The second part of this implementation (four periods) was in the context of spring elongation when a force is applied on it. Students designed and applied investigations to examine whether and how variables like the length of the spring, the material which it is made of and the mass of an object hanged on it influence its elongation.

The second implementation took place in a Lyceum, in two classes of 20 and 21 15-16 year old students, respectively. The instruction in both classes was undertaken by the same physics teacher who had about 20 years of experience in teaching physics in secondary education, most of them in higher secondary education. This teacher also follows seminars for her professional development in regular base.

The second implementation lasted eleven 45-minute teaching periods and used the contexts of "freefall" and motion on an inclined surface. In particular, the students designed and implemented investigations to examine variables that might influence the acceleration of a falling object, like the mass of the object, the height that it is released from, the shape of the object and its volume. In addition, they designed and implemented investigations to examine variables that might influence the acceleration of an object moving on an inclined surface, like the mass of the object, its volume, the material of the surface and its inclination. The teacher also tried to introduce the Newton's 2nd law for motion in an effort to explain the motion of the objects in the aforementioned cases.

The third implementation was realized at a gymnasium, in two classes of 19 14-15yo students each, by a teacher with 8 years of teaching experience in physics, most of them in lower secondary schools. The teacher holds a bachelor in physics and a master in educational technology. She also follows seminars for her professional development quite often and she tries to apply teaching innovations in her teaching practice.

The third implementation lasted ten 45-minute teaching periods and used the context of "Newton's laws of motion". In particular, the students designed and implemented investigations to examine variables that might influence the acceleration of a falling object, like the mass of the object, the height that it is released from, the shape of the object and its volume. They also examined through experimentation the relation between action and reaction forces, for example by having model cars of different masses colliding with different velocities between them and measuring the forces that the one applies to the other.

All three teachers had previously participated in a series of three seminars related with formative assessment practices in general but also, with “on the fly interaction” in particular. These seminars included theoretical parts regarding formative assessment practices, but also practical activities, where teachers observed videos of other teachers applying the method of “on the fly interaction”. Then they had discussions about the opportunities and the challenges of this method and, how they could better apply it in their own teaching practice. In addition, during the implementations, the teachers had reflective discussions with the researcher who was observing their lessons, trying to identify moments that they successfully applied the method of “on the fly interaction” and instances that they could have facilitated in a more productive way.

“On the fly” formative assessment was applied during whole class discussions, where the teacher raised specific issues relevant to the experimental design, the experiment itself or interpretation of the results. To some extent, these discussions were planned, in the sense that the teacher identified beforehand issues that s/he should be aiming to bring into focus. However, the teacher entered into these discussions expecting to manage the dialogue according students’ needs.

3.2. Data collection

A total of 52 lessons were videotaped and those parts of the video that include interactions “on the fly” were identified and transcribed. For reasons of triangulation, teachers were interviewed after the implementations. Semi-structured interview protocols were used for the interview sessions, along with video recorded episodes of the corresponding classroom discussions.

3.3. Data analysis

Interactions between teachers and students were coded using the ESRU scheme (Ruiz-Pinto & Furtak, 2006). The analysis was applied in three levels (Tiberghien & Malkoun, 2009):

1. Microscopic: Line by line coding of each teacher or student contribution in the dialogue.
2. Mesoscopic: Characterization of each cycle as complete or incomplete ESRU cycle.
3. Macroscopic: The use of the emerging information is evaluated at the level of a whole episode.

Specifically, in an attempt to identify patterns in the interactions between the teacher and the students, the interactions were coded at microscopic level as either instances of E (elicit), S (students’ response), R (recognize) or U (use). In addition, we sub-categorized these instances to the various ways they occurred in class, for example the different ways a teacher might have used for eliciting information (Table 1).

Our subcategorization, which emerged after following an open-coding procedure, is quite different from the one Furtak and Ruiz-Primo (2006). For example their subcategories for the *Elicit* category were a quite extensive list of scientific processes that the teacher could ask for during the lesson, like “Teacher asks students to make predictions”, “Teacher asks students to provide observations” or “Teacher asks students to evaluate the quality of evidence”. These subcategories relate to the subject under discussion at the particular episode of the classroom dialogue; in other words, other subcategories are more likely to appear if the discussion relates to the design of an investigation and others if the discussion relates to the results of an investigation.

On the other hand, our subcategories were developed not to be related to the discussion of each episode but they rather focus on how the teacher and the students engage or respond during the dialogue. So, we have a subcategory “Teacher asks students an open-ended question” that allows us to identify instances where the teacher asks an open-ended question and this can be referring to a prediction, reporting data or providing an example to support or contradict a claim. This kind of categorization of teacher’s contribution to the dialogue is similar in purpose to TD- Divergent and TC-Convergent categorization of the coding system Nieminen et al. (2015) adapted from Torrance and Pryor (2001) and Alexander (2006). Our aforementioned subcategory falls for example in Teacher Divergent talk, which refers to all instances in the dialogue where the teacher asks questions to probe and encourage students’ thinking and to promote discussion.

Table 1: *Sub-categories of teacher – students’ interactions*

Categories	Subcategories	
Elicit	E1	Teacher poses a closed-ended question to elicit students` reasoning about a new (although interrelated) concept/idea/relation
	E2	Teacher asks students an open-ended question (i.e. to offer an example or report data)
	E3	Teacher seeks to sustain the cycle without using emergent information (i.e. repeats the previous question)

Student's Response	S1	Student suggests a concept/relation in response to question posed by the teacher
	S2	Student offers justification for his/her reasoning
	S3	Student provides an example or reports data
	S4	Student explicates an inference/poses a question about an aspect of the topic under discussion
	S5	Student provides a "yes/no" answer
Recognition	R1	Provision of affirmation
	R2	Teacher readily offers the right answer to a question posed by him/herself or by a student.
	R3	Provision of disconfirmation
	R4	The teacher acknowledges a contribution made by the students
Use	U1	Teacher suggests an activity that could help students resolve a specific (conceptual) issue
	U2	Teacher seeks to focus students' attention on something with the intent to facilitate the discussion (e.g. stated opinions/data/examples)
	U3	Teacher seeks to engage students in deeper reasoning on something (further analysis/explanation)
	U4	Teacher articulates the consensus from series of contributions that were exchanged
	U5	Asks the opinion of other students about an expressed idea

Another important differentiation of our subcategorization is that we developed subcategories for students' responses also. This allowed us to illustrate in detail how students contribute to the dialogue and to look for patterns between the various teacher's contributions and students' responses.

On a mesoscopic level of analysis, we tried to identify instances in which the cycle happened to break and elaborate on the different reasons underlying this. Our aim was to reveal possible trends or patterns between the ESRU subcategories we had developed, to identify the variation as to what has caused the ESRU cycles to break. We also looked for patterns that might reveal possible factors that might have facilitated the completion of ESRU cycles. We expected that this could provide evidence of challenges and intricacies associated with interactions on the fly as an unplanned formative assessment method, in order to address our second research question. To examine whether the completion of an ESRU cycle is influenced by the codes that appear in the cycle but also whether other codes tended to appear together, we ran cluster analysis through the codes that appeared in each cycle, both separately for each implementation and by mixing all the implementations together. For this purpose, we used each subcategory as a binary variable to indicate whether is appeared or not in each cycle. We also ran Φ correlation tests among subcategories that the cluster analysis revealed as connected in order to attain a more reliable conclusion regarding the strength of the relation between them.

For reasons of triangulation, teachers were interviewed after the implementations. Semi-structured interview protocols were used for the interview sessions, along with video recorded episodes of the corresponding classroom discussions. Specifically, the teachers watched pre-selected parts of the video recordings of their lessons, to elicit useful reflections on the part of each teacher-interviewee, about particular issues associated with aspects of the specific assessment method that emerged during the discussion.

After preliminary analysis of a number of dialogue episodes at macroscopic level, which is the part of the dialogue that focuses on a particular theme, we were able to realize that, in many cases, the teacher chose to utilize particular information that comes up during discussion, at the expense of alternative ideas and possible misconceptions that also seem to exist. Consequently, despite the fact that several ESRU cycles were realized, at the same time, instructionally valuable contributions from students tended to be either dismissed or used in a non-optimal manner. Throughout this study we used the term "missed opportunities" to refer to these instances of either not utilizing contributions from students or not doing so in a productive manner. We decided to identify and categorize these instances because, we believe that, this would provide more insights to the challenges that teachers meet when applying “on the fly interactions” as a method of formative assessment.

Research evidence suggests that completed ESRU cycles and iterations of complete ESRU cycles are indicative of productive interactions between the teacher and the students in the assessment dialogue framework (Ruiz-Pinto & Furtak, 2006). However, we believe that in order to be productive, on the fly interactions need to address disciplinary substance at the appropriate level (Coffey et al, 2011). Thus, we need an alternative way to illustrate the conceptual completeness, depth and coherence of each episode of the dialogue. This brings us to the third research question, whether we can represent the conceptual completeness, the depth and coherence of the classroom dialogue when interaction on the fly is applied using concept maps. More specifically these maps that that were developed for each episode of the classroom dialogue depict the variety of concepts that emerge in discussion, whether these emerge from the students or the teacher and how much these are interpreted and linked with each other.

Finally, in order to address the fourth research question, we examined whether the patterns found between the ESRU subcategories relate to the conceptual coherence and depth of the classroom dialogue as it is illustrated on the concept map of each dialogue. In order to do so, we had to quantify the conceptual coherence and depth as it was depicted on the concept maps. We did so in three dimensions:

- The rate of the essential concepts that were actually used in each episode of the dialogue.
- The rate of connections between concepts that were revisited in at least one more ESRU cycle than the one they appeared.
- The rate of concepts in each episode that were connected with each other forming loops on the conceptual maps.

To find the rate of the essential concepts that were actually used in each episode of the dialogue, we divided the number of these concepts with the number of concepts that would have been used if concepts we perceive as essential for students' understanding on the particular theme under discussion were also included. The rate of connections between concepts that were revisited in at least one more ESRU cycle than the one they appeared was found by dividing the number of links between concepts in the particular episode which were realised at least in two ESRU cycles with the whole number of links between concepts that were made in the episode. The rate of concepts in each episode that were connected with each other forming loops on the conceptual maps was found by dividing the number of concepts that were included in such loops at each episode by the whole number of concepts that were used in the episode.

Next, we used Pearson correlation to examine whether this value for the conceptual coherence and depth of the classroom dialogue is related with the rate of completion of ESRU cycles of each episode.

3.4. Validity and liability of the research

The inter-rater agreement was calculated for the application of the ESRU coding system described above. We developed a coding manual including code name, code description, and an example of code attribution for each code used. The coding manual was then used independently by three researchers to code the data. We used an extract from one implementation to calculate the inter-rater agreement for the E, R, and U codes, which were used to characterise teacher's actions. The S code (student responds to teacher question/comment) posed no ambiguity in attribution, and therefore was not considered in the calculation. We reached the following agreements (each value represents the agreement between each pair of independent researchers): 89.5, 84.2, and 73.7%. Disagreements in the coding were then discussed among all three researchers and consensus in attribution was reached raising the measured value of Cohen's kappa to k = 0.9.

The inter-rater agreement for the application of ESRU subcategories was also calculated in the same way. We reached the following agreements (each value represents the agreement between each pair of independent researchers): 65.4, 63.7, and 63.1%, which are considered good enough, taking into account the complexity of the dialogue and the reactions involved. After discussing the disagreements in the coding among all three researchers, the inter-rater agreement was increased and the measured value of Cohen's kappa was k = 0.8.

The transferability of the results of this research is limited by a number of factors. Firstly, despite we collected data from a large number of lessons (52) in three different school contexts (Gymnasium, Lyceum and Vocational school), the number of the teacher that participated is small. Secondly, the lessons we collected data from focused only in a small range of science curriculum that relates to mechanics (motion and forces).

Finally, the researcher collected all the essential legal licenses for collecting and keeping data from the Ministry of Education, the schools' headmasters, the participating teachers and their students, including parents approve for videotaping the lessons.

Researcher's biography

The researcher has got a bachelor degree on physics from the University of Patra in Greece and a master on opto-microelectronics from the University of Crete in Greece. He has been working as a researcher in Learning in Science Group of University of Cyprus for three years and then as special teaching staff at the department of Education of the University of Cyprus for six years. His research interests during these nine years of work at the University of Cyprus were related to learning in science by inquiry, the use of innovative/technological tools in teaching and learning science and formative assessment. The last two years works as a physics teacher in secondary education.

4. RESULTS

4.1. Research Question 1: Patterns of *on the fly interactions* between teacher and students.

In order to answer to the first research question regarding the patterns we can identify during “on the fly interactions” between the teacher and the students, we coded all the dialogues of on the fly interaction we identified in our video recordings, using the ESRU scheme and the subcategories we developed.

4.1.1. Examples of coding using ESRU categories and sub-categories

On the next tables we will present representative excerpts from all three implementations in order to depict how we used the ESRU categories and their subcategories to analyse the dialogues on a microscopic level. We will also explain through these examples, in which cases ESRU cycles were defined as complete and when as incomplete.

Table 2: *Excerpt from 1st implementation, Lesson 3, Episode 1.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	This came out as a straight line. What could we conclude when we get a straight line graph?	E	E1
2	Student 1	The mass of the object which is applied on the spring influences the elongation	S	S1
3	Teacher	I see. So the first conclusion is that the mass of the object influences the elongation.	R	R1
4	Teacher	Is there a pattern?	U	U2
5	Student 2	They are proportional.	S	S1
6	Teacher	There is a pattern, the 2 quantities are proportional.	R	R1
7	Teacher	T: What is the measurement which does not agree with this conclusion A?	U	U2
8	Student	…		
9	Teacher	I have a measurement, second measurement, third measurement fourth measurement, fifth measurement. Which measurement does not agree with this pattern?	E	E1
10	Student 3	Zero	S	S1
11	Teacher	The first. The first did not agree with this pattern. Something was wrong here. Good.	R	R1

The excerpt above (Table 2) is from the 1st implementation, Episode 1, where the teacher and the students discussed the results of their experiment, in which they investigated whether the mass of the object applied on the spring influences its elongation. The 4th and 5th column on Table 2 refer to the microscopic level of analysis where each contribution in the dialogue is coded using ESRU categories and subcategories respectively.

In this excerpt, on turn 1, the teacher asks about students' conclusions drew from the graph representing the experimental results. This is a question that aims to reveal students' understanding and thus is coded as "Elicit". In particular, the aim of the question is to reveal their understanding about the relation between the mass of the object and the elongation of the spring where it was hanged on and thus, is coded in the subcategory E1 (teacher poses a closed-ended question to elicit students` ideas about a new concept/idea/relation). Then, in turn 2, student 2 express his/her conclusion and this is coded as S (student response) and in particular in subcategory S1 (student suggests a concept/relation in response to question posed by the teacher) since the student suggests a relation between concepts (mass, elongation). Next, in turn 3, the teacher recognizes (R) student's response by repeating it in a way that shows that she agrees with him/her and thus, it is coded as subcategory R1 (provision of affirmation). Then in turn 4, the teacher tries to take students' thinking a step forward by asking them to look for a pattern on the relation they mentioned which is coded as U2 (teacher seeks to focus students' attention on something with the intent to facilitate the discussion).

In turns 5-7 the subcategories S1, R1 and U2 are repeated as well as E1, S1 and R1 in turns 9-11. However, there is a differentiation between them on a mesoscopic level of analysis. Turns 1-4 form a complete ESRU cycle because all ESRU categories appear there. Turns 5-7 also form a complete ESRU cycle despite Elicit category seems to be missing, because elicitation takes place from the end of the previous cycle (turn 4) where information is used to open a new cycle. However, turns 9-11 form an incomplete ESRU cycle since there is no Use of student's response apart from just confirming his/her answer as correct (turn 11).

Table 3: *Excerpt from 1st implementation, Lesson 1 Episode 2.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	So you have 1, 2, 3, 4. What do you notice; the elongation depends on the length of the spring?	E	E1
2	Student 5	Yes	S	S5
3	Teacher	Tell me	U	U3
4	Student 5	Yes because we see that with increasing elongation increases and …	S	S2
5	Teacher	Which quantity increases and which…	E	E1
6	Student 5	When length increases…	S	S3
7	Teacher	The length of the spring	R	R4
8	Student 5	Then the elongation increases too	S	S3
9	Teacher	The elongation.	R	R1
10	Teacher	And you had a measurement that …	E	E1
11	Student 1	S1: It didn't fit to the pattern	S	S1
12	Teacher	T: Why didn't fit?	(R)U	U3
13	Student 2	S2: Because we used a different type of spring with different material than the other and so …	S	S2
14	Teacher	T: Well, you think the reason was because you used a different kind of spring.	R	R4

During the 2nd episode of the same lesson more subcategories appeared (Table 3). In turn 2 a student responds with a single "yes" which is coded as S5 while the contributions of Student 5 (turn 4) are coded as S2 (student offers justification for his/her reasoning) because s/he tries to justify his/her conclusion or S3 (turns 6 and 8, student provides an example or reports data) when backing his argument with evidence from their measurements. These contributions were triggered by the teacher who encouraged (turn 3) the student to elaborate his/her previous answer (turn 2) and thus, her contribution was coded as U3 (teacher seeks to engage students in deeper reasoning on something (further analysis/explanation)). The same case appears in turn 12 with the teacher asking for explanation (U2) and the Student 2 explaining (S2) in turn 13. Another subcategory that appears in this excerpt is R4 (the teacher acknowledges a contribution made by the students) in turn 7 where the teacher repeats and student's contribution to make it clearer, but without confirming it as right neither disconfirming it as wrong.

On a mesoscopic level of analysis, we realize that in this excerpt there is an incomplete ESRU cycle that is also extended. This means that categories S and R, are repeated at least once more before the end of the cycle. This is the case of the cycle in turns 5-9 which consists of a pattern ESRSR. This indicates that neither complete nor incomplete cycles have always a straightforward pattern like ESRU and ESR or ES respectively.

Table 4**:** *Excerpt from 2nd implementation, Lesson 5, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What do you realize from the measurements we got?	E	E1
2	Student 1	Basically the measurements are the same.	S	S1
3	Student 2	As we increased the mass, it falls faster	S	S3
4	Teacher	9.67, 9.69, 9.66, if we remove the last decimal, how much would these measurements be?	(R)U	U2
5	Student 3	9.7	S	S1
6	Teacher	9.7 all of them. Right?	R	R1

Looking at the micro-level analysis of the next extract (Table 4), the teacher uses an open-ended question (turn 1) to invite students to share their views on their interpretation of experimental findings (E1). Two students respond to this, providing different interpretations for the experimental data. Student 1 interprets measurements as being approximately the same (turn 2), while the Student 2 thinks that the object falls faster as mass increases (turn 3). The teacher responds to this situation in turn 4 (use) by attempting to resolve this difference in interpretation recalling students of the data they have gathered and suggesting rounding them to the 1st decimal.

Looking at the meso-level of analysis, this extract represents one extended and complete ESRU cycle (turns 1-4) since the teacher responded to students' arguments and used their measurements to guide them to reach a conclusion. It is an extended cycle because it includes responses from two students, thus the pattern is actually ESSRU. However, the teacher didn't justify her suggestion to round the measurements to the 1st decimal, thus the 2nd cycle (turns 5-6) remains incomplete.

Table 5: *Excerpt from 2nd implementation, Lesson 5, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Student 1	Mass has an influence because they put 7gr and we put 13gr.	S	S4
2	Student 2	But mass doesn't have an influence.	S	S4
3	Student 1	It does influence.	S	S4
4	Teacher	So it does…	R	R4
5	Teacher	But here we found the same acceleration whether we put 4gr, 7gr or 14gr.	U	U2
6	Student 1	We added only 4gr, how would that make any difference?	S	S4
7	Teacher	What are you saying? Does mass influence?	U	U3
8	Student 2	If we had added 1 Kg it would still be the same, yes?	S	S1
9	Teacher	If we had added 1 Kg would it still be the same?	R	R4
10	Teacher	Why does it have to be the same?	E	E1
11	Student 3	Because the acceleration of gravity is the same.	S	S1
12	Teacher	That is what we wanted to prove. We know it, but we wanted to prove it.	R	R1

At a micro-level of analysis, turns 1-3 (Table 5) show that two students initiate a discussion around the effect of mass on the acceleration of an object during free fall and thus their contributions are coded as S4 (Student explicates an inference/poses a question about an aspect of the topic under discussion). In turns 6 and 8, students seem to use "rhetoric" questions to make statements about their position in the discussion, and in turn 11 one student makes a statement of fact. There are three instances in which the teacher recognises students' contributions. In turn 4 the teacher is reaffirming what student 1 said in turn 3. In turn 9 the teacher repeats the contribution of student 2 (in turn 8). In turn 12, the teacher reinforces the statement made by student 3 (turn 11). The teacher uses the information gathered from students' contributions (that some students seem to be confused about the effect of mass on acceleration) to help them make sense of their data. An example of this is on turn 5 where the teacher reminds the students of their findings. On turn 7, the teacher asks probing questions to get students to articulate their ideas encouraging them to exchange ideas and present different perspectives. The teacher uses an open-ended question (turn 10) to probe students understanding on the relation between mass of an object on free fall and its acceleration (Eliciting).

At a meso-level of analysis, we observe the absence of Eliciting instances in the first three ESRU cycles (turns 1-5, 6-7 and 8-9 respectively), and the absence of Using instance on the 4th ESRU cycle (turns 10-12) as well. The absence of Eliciting instances in the first two cycles is due to the discussion being initiated by two students as they are expressing a disagreement about a previous statement. Considering that students initiate the discussion in these cycles, we consider them complete because there is evidence that the teacher is responding to the students to move them towards the learning aims (coded as Use). We could also consider cycles 2 and 3 as a single extended cycle since in turn 7 the teacher invites the student to explain his opinion more explicitly. Nevertheless, this extended cycle remains incomplete, since at turn 10 the teacher moves to what is established scientific knowledge without further elaboration on student 1's objection.

Table 6: *Excerpt from 1st implementation, Lesson 1, Episode 1.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Did you find a relationship between the mass and the object's velocity?	E	E1
2	Student 1	No	S	S5
3	Teacher	What did you find?	E	E4
4	Student 1	We found the same velocity for every object… not the same, more or less the same.	S	S1
5	Teacher	The differences were very small.	R	R1
6	Teacher	But for the mass… How were you altering the mass? Tell me the values of the masses you used.	E	E2
7	Student 1	3.15g, 7.8g, 12.9g και 24g.	S	S3
8	Teacher	So how would you compare the masses?	(R)U	U2

In the excerpt above (Table 6), we observe two more Elicit subcategories. In turn 3, the teacher sustains the cycle without using any information (E4), since there was not provided any actually from the part of the student, encouraging the student to provide a more comprehensive response. In turn 6, the teacher asks students to report data, which falls in subcategory E4 (teacher asks students to offer an example or to report data).

Table 7: *Excerpt from 2nd implementation, Lesson 3, Episode 3.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category

1	Teacher	During this experiment, the weight was causing acceleration?	E	E1
2	Student 1	Yes	S	S5
3	Teacher	How do you know?	(R)U	U3
4	Student 2	Because it falls.	S	S2
5	Teacher	Because it falls, because you measured it!	R	R2

In the excerpt above (Table 7), we used the code R2 for the subcategory of Recognition which corresponds to cases where the teacher actually ignores student's response and provides the answer to her own question. In this case the teacher asks a question about the kind of motion an object has during free fall and wants students to refer to their measurements from the experiment they carried out beforehand. Instead, the student only mentions that weigh causes acceleration because the object is falling. Of course, this is insufficient, because there could be the case of an object falling without acceleration. So, the teacher decides to indicate that they should notice that from their measurements, without problematizing them whether the previous respond was correct. That is why, on mesoscopic analysis, we also consider this cycle (turns 4-5) as incomplete, since there is an indication for misunderstanding in student's respond that was not used. On the other hand, in turns 1-3 there is a complete cycle because the teacher was not satisfied with student's answer in turn 2 but encouraged him/her to explain him/herself.

Another example from the same implementation relates to the discussion about the results found in the experiment undertaken by the students in which they measured the acceleration of an object, when released from different heights:

Table 8: *Excerpt from 2nd implementation, Lesson 4, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What is your conclusion?	E	E1
2	Student 1	When the height increases, the acceleration drops.	S	S1
3	T	By how much does it drop? This 8.4 could be 8.5 so if we just ignore the decimal part of the number this would round up to 9. This is also the case for these two measurements (points to other 2 measurements).	R	R3
4	S2	So they are the same…	S	S1

In the above excerpt (Table 8), the code U (use of the emergent information), is not appeared, so on a mesoscopic level, it is an example of an incomplete cycle. The teacher recognises that students are confused by some variation at the calculation of acceleration using measurements of objects falling from different heights (turn 3). Yet, in this instance, the teacher avoids discussing experimental errors; how they might emerge and how they could influence measurements, as well as whether the variation has a specific trend or not. Rather, she skips the potentially useful discussion on this and, instead, disconfirms student's response (R3). Then, she seeks to illustrate for students how the observed variation in the measurements of acceleration is sufficiently small to be dismissed as insignificant. In particular, she suggests that putting aside the decimal parts of the numbers, the measurements seem to round up to the same whole number, which implies that height does not influence acceleration.

Table 9: *Excerpt from 2nd implementation, Lesson 4, Episode 2.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What does weight means?	E	E1
2	Student 1	Weight relates with gravity.	S	S1
3	Student 2	The force that Earth attracts objects.	S	S1
4	Teacher	It is the force that Earth attracts objects.	R	R1
5	Teacher	This is a force that exists due to gravity.	U	U4

The new subcategory that is presented on Table 9 is U4, which refers to contributions where the teacher articulates the consensus from series of contributions that were exchanged. In this case, students expressed various opinions regarding the concept of "weight" (turns 2 and 3) before the teacher used them in order to provide a more complete definition about the particular concept (turn 5).

Table 10: *Excerpt from 1st implementation, Lesson 3, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What did you find here? Your team didn't measure the elongation; what did you measure?	E	E1
2	Student 1	The whole length of the spring.	S	S1
3	Teacher	The whole length of the spring.	R	R4
4	Student 2	But you can still find the elongation.	S	S4
5	Teacher	Can you explain him what you mean?	(R)U	U5

During this excerpt (Table 10), a student states that their team measured each time the whole length of the spring instead of the net elongation (turn 2). Another student interferes noting that they can still calculate the elongation (turn 4) and the teacher encourages the second student to explain to the other team how to elaborate their measurements. This contribution of the teacher (turn 5) is coded as U5 (teacher asks the opinion of other students about an expressed idea).

Finally, in the excerpt in Table 11, a student initiates the discussion asking whether gravity can exist without atmosphere. The teacher first asks his opinion (turn 2) and then demonstrates an experiment in order to solve his query (turn 5). This action is coded with U1 which stands for the case where the teacher suggests or carries out an activity that could help students resolve a specific (conceptual) issue.

Table 11: *Excerpt from 3rd implementation, Lesson 4, Episode 3*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Student 1	Can I ask something? If atmosphere is lost from Earth, there will still be gravity?	S	S4
2	Teacher	What do you think?	U	U3
3	Student 1	There will be	S	S1
4	Student 2	There will be	S	S1
5	Teacher	Look: [Teacher takes a vacuum tube and removes the air using a pump. Then leaves an object to fall in the tube]. Alright? I removed the air but still the Earth attracts it downwards.	U	U1
6	Student 1	So there is no air now in there?	S	S4
7	Teacher	No,	R	R1
8	Teacher	Listen. [The teacher opens a valve and the air is heard getting back in.] Alright? So, even when there is no air, gravity still exists.	U	U1

4.1.1. Frequencies of ESRU Subcategories

1st implementation (Vocational School)

We will now present the frequencies for the appearance of each ESRU subcategory during each implementation, starting from the 1st implementation, which took place at a vocational school.

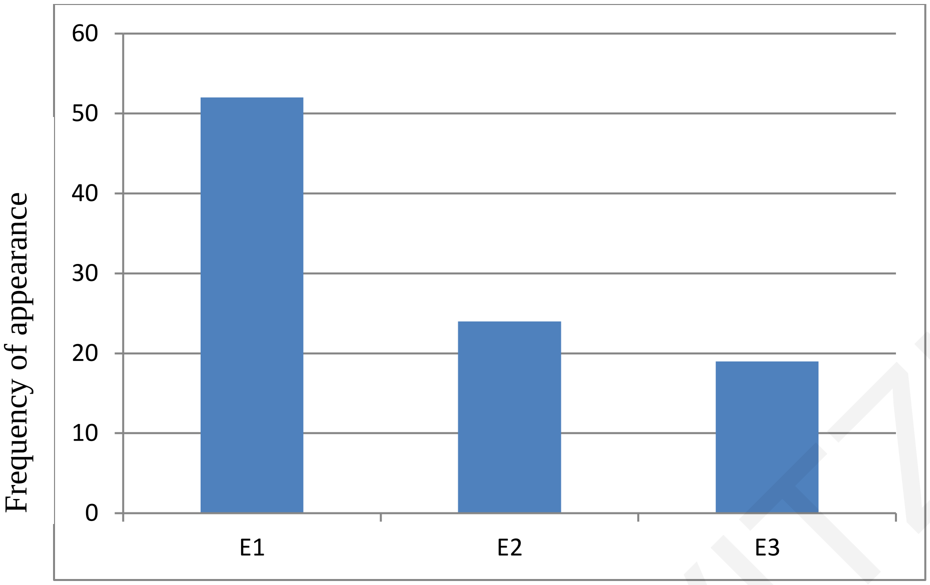

Figure 2: Frequency of appearance of Elicit subcategories in the 1st implementation.

Regarding the ways the teacher tried to elicit students' ideas, the E1 code (teacher poses a closed-ended question to elicit students` reasoning about a new (although interrelated) concept/idea/relation) appears most often (55%) whereas the E2 (teacher asks students an open-ended question, i.e. to offer an example or report data) appears in less cases (25%). In some cases (20%) the teacher seeks to sustain the cycle (E3) but without using any emergent information (i.e. repeats the previous question, asks for clarification or suggests a false concept/idea/relation and gets students to reflect on).

Students' responses varied in this implementation, with S1 (student suggests a concept/relation in response to question posed by the teacher) was appeared more frequently than the other subcategories (43%). The next were S3 (student provides an example or reports data) with 22% and S2 (Student offers justification for his/her reasoning) with 19%, indicating that students' contribution in the dialogue was quite rich. Less often students responded with a yes/no answer (S5, 15%). However, rarely they explicated an inference or posed a relevant question about an aspect of the topic under discussion (S4, 13%).

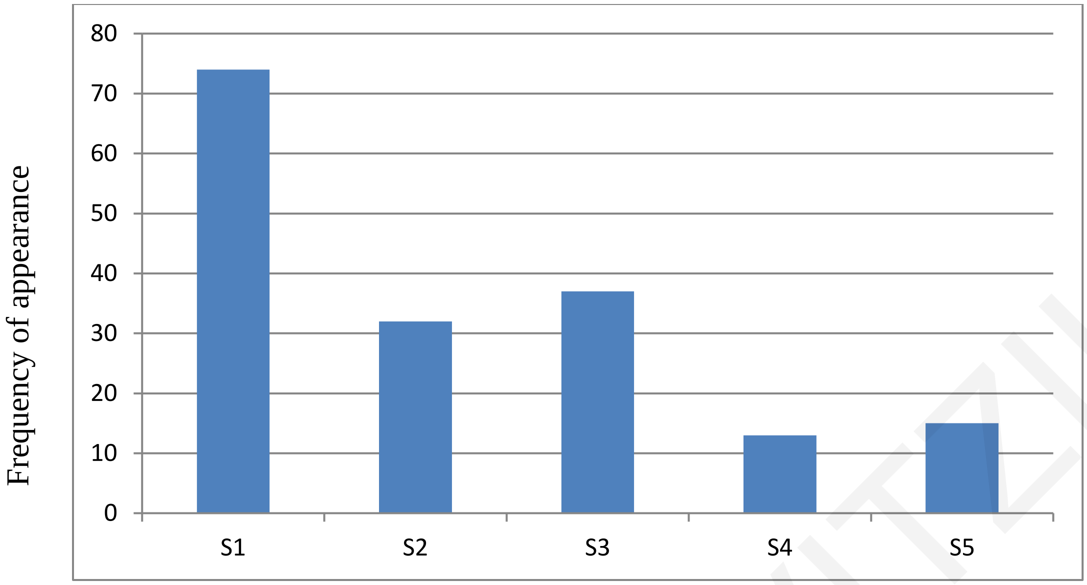

Figure 3: Frequency of appearance of Students' response subcategories in the 1st implementation.

The teacher recognized students' contributions mostly by acknowledging them in a neutral way (R4, 52%) or by providing affirmation (R1, 39%). Only in few cases she disconfirmed students' response (R3, 6%) or readily offered the right answer to a question posed by herself or by a student (R2, 3%). This allocation of manners might have allowed to sustain the discussion.

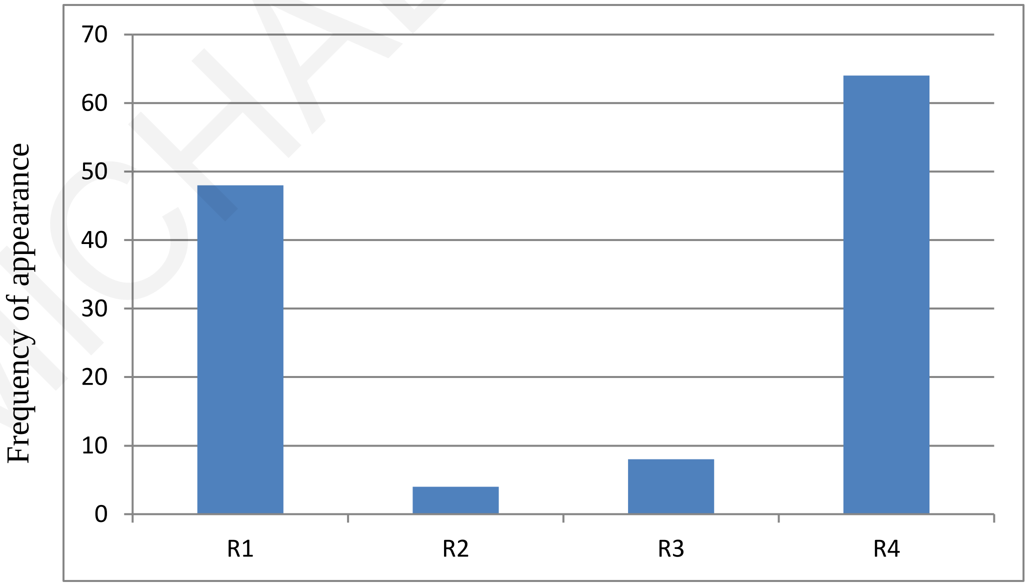

Figure 4: Frequency of appearance of Recognition subcategories in the 1st implementation.

The teacher in many cases used the emergent information by focusing students' attention on stated opinions/data/examples with the intent to facilitate the discussion (U2, 53%). In some cases tried to engage students in deeper reasoning/analysis on something (U3, 19%) where in others to articulate consensus from series of contributions that were exchanged (U4, 19%). In fewer cases she suggested or asked for an activity that could help students resolve a specific issue (U1, 4%) or asked the opinion of other students about an expressed idea (U7, 4%).

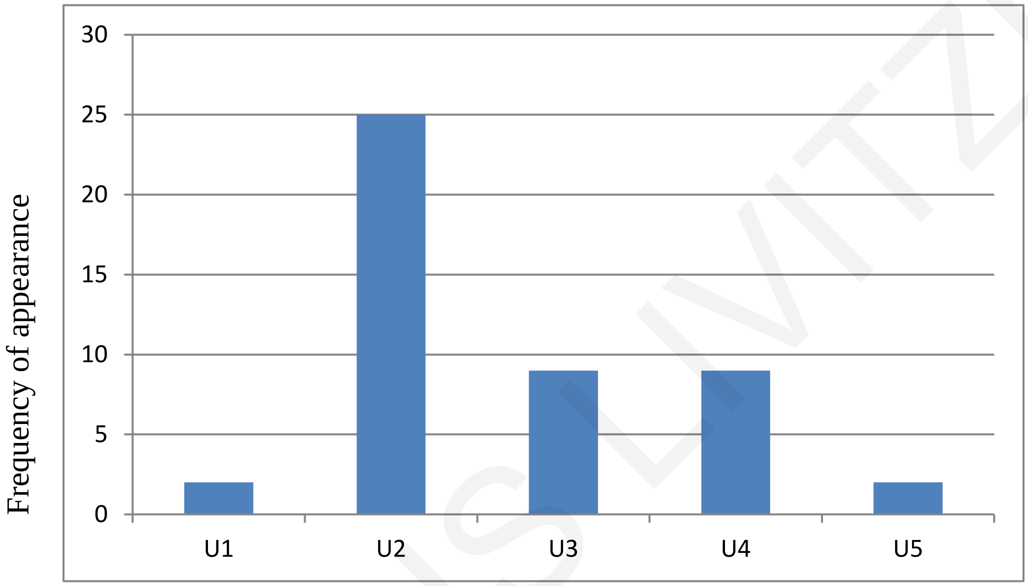

Figure 5: Frequency of appearance of Use subcategories in the 1st implementation.

2nd implementation (Lyceum)

Next, we present the frequencies in which each ESRU subcategory appeared during the implementation at the Lyceum. Analyzing the transcribed data of the 2nd implementation, it appears that code E1 (teacher poses a closed-ended question to elicit students` reasoning about a new concept/idea/relation) appears most of the times (65%) when teacher tries to elicit students ideas. In some cases she sustains the cycle without using information (i.e. repeating question) (E3, 32%) whereas only in very few cases asks students an open-ended question, i.e. to offer an example or report data (E2, 3%).

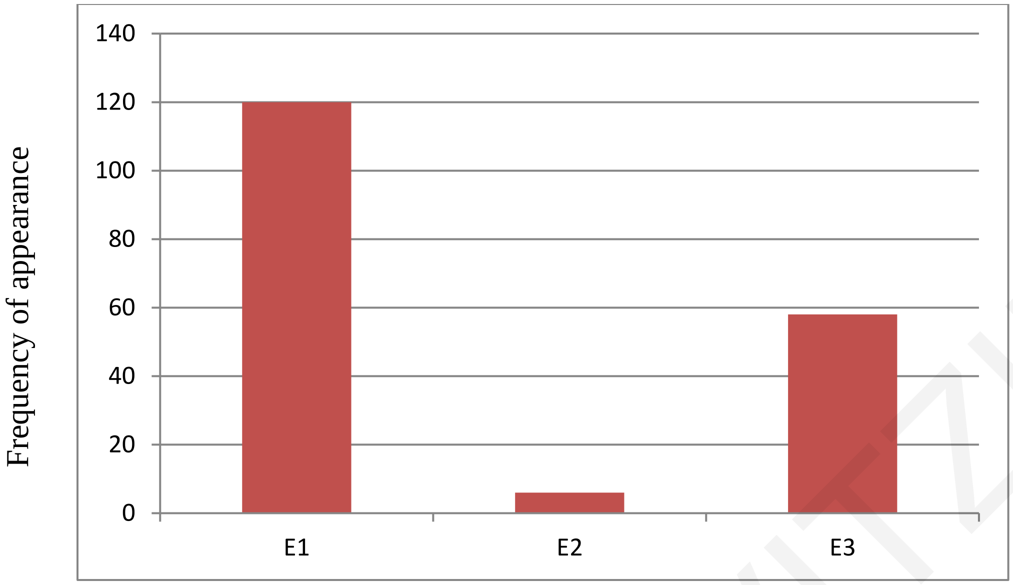

Figure 6: Frequency of appearance of Elicit subcategories in the 2nd implementation.

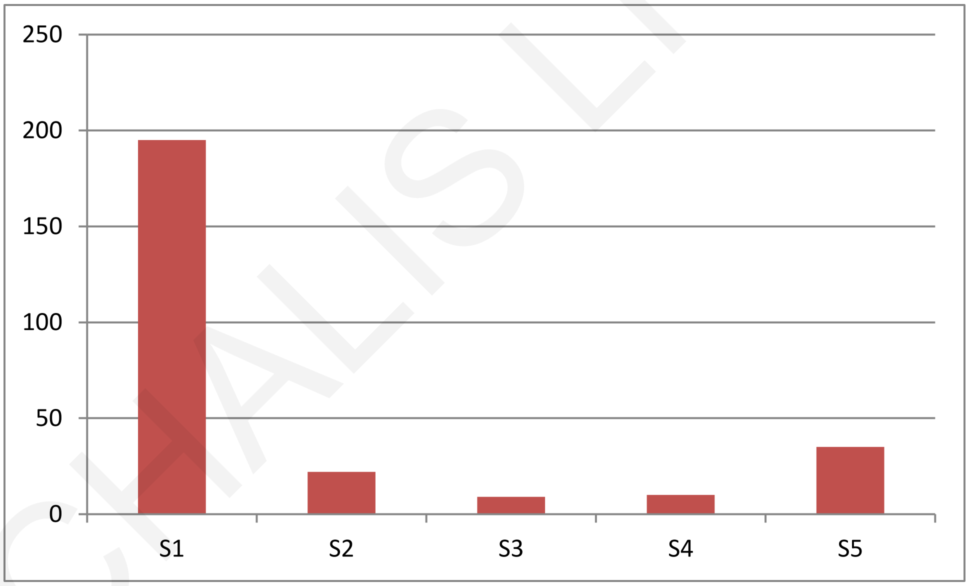

Figure 7: Frequency of appearance of Students' response subcategories in the 2nd implementation.

Students in this implementation mainly suggested a concept or relation in response to a question posed by the teacher (S1, 72%). In some cases students explicated an inference/posed a question about an aspect of the topic under discussion (S4, 19%) or they responded with a yes/no answer (S5, 13%), while appeared much more rarely. This indicates that the dialogue was quite poor since the other subcategories that relate with students offering justification for their reasoning (S2, 8%) and with providing examples or reporting data (S3, 4%), appeared only in rare occasions.

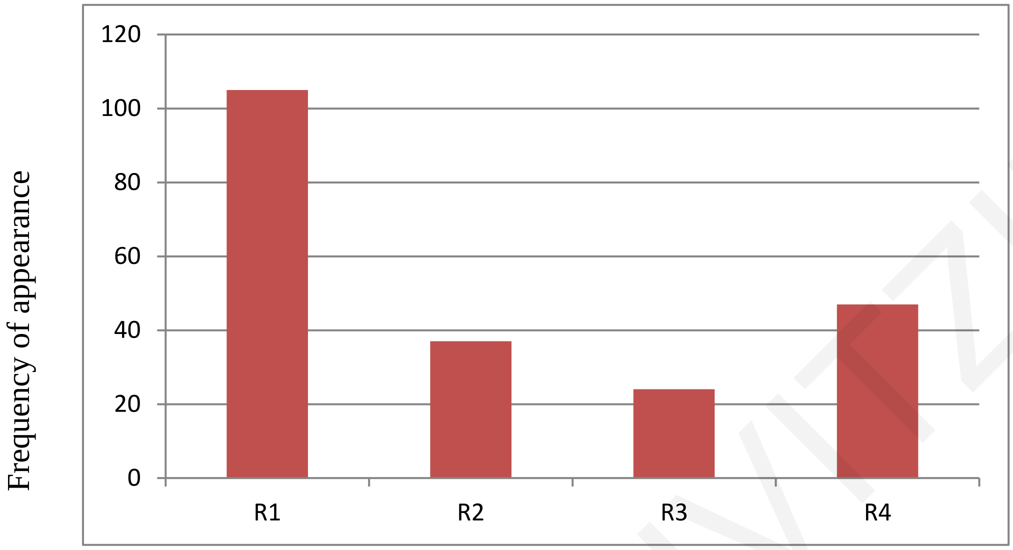

Figure 8: Frequency of appearance of Recognition subcategories in the 2nd implementation.

The teacher recognized students' contributions in most of the cases by providing affirmation (R1, 49%). In fewer cases she readily offered the right answer to a question posed by herself or by a student (R2, 17%). In some cases she just acknowledged students' responses in a neutral way (R4, 22%). Also, in few cases she disconfirmed students' response (R3, 11%). The way the teacher responded to students' contributions, might be accountable for impeding an open discussion in class, as it has been suggested already before.

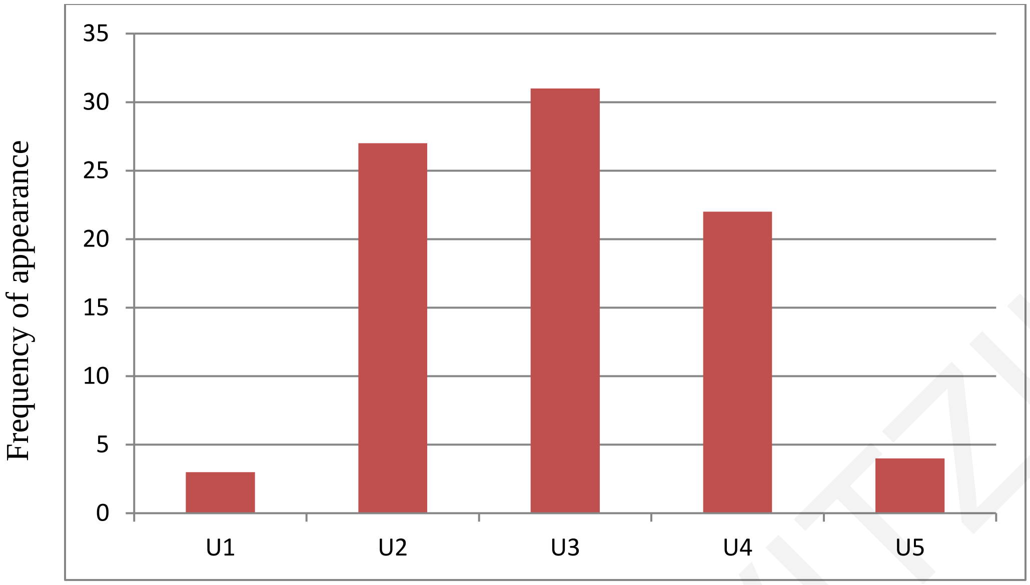

Figure 9: Frequency of appearance of Use subcategories in the 2nd implementation.

The teacher in many cases used the emergent information by focusing students' attention on stated opinions/data/examples with the intent to facilitate the discussion (U2, 31%). In a similar number of cases, she tried to engage students in deeper reasoning/analysis on something (U3, 36%) where in a bit fewer to articulate consensus from series of contributions that were exchanged (U4, 25%). In much fewer cases she suggested or asked for an activity that could help students resolve a specific issue (U1, 3%), or asked the opinion of other students about an expressed idea (U5, 5%).

3rd implementation (Gymnasium)

Next, we present the frequencies of appearance of ESRU subcategories during the implementation at the Gymnasium. From the analysis of the transcribed data of the 2nd implementation, revealed that the code E1 (teacher poses a closed-ended question to elicit students` reasoning about a new concept/idea/relation) appears most of the times (83%) when teacher tries to elicit students’ ideas. In some cases she sustains the cycle without using info (i.e. repeating question) (E3, 13%) whereas only in very few cases asks students an open-ended question, i.e. to offer an example or report data (E2, 4%).

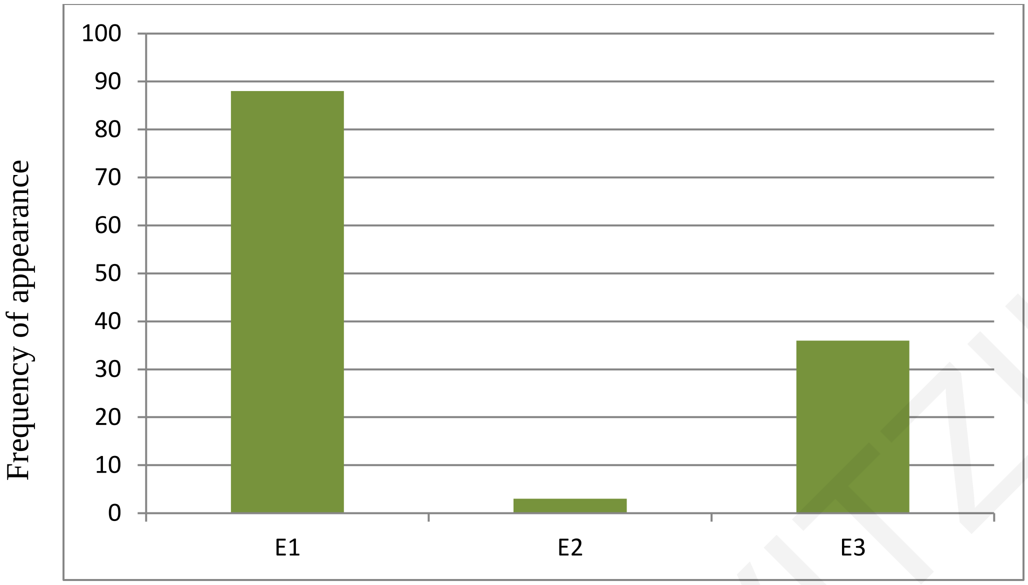

Figure 10: Frequency of appearance of Elicit subcategories in the 3rd implementation.

Students in this implementation mainly suggested a concept or relation in response to a question posed by the teacher (S1, 46%). In some cases students explicated an inference/posed a question about an aspect of the topic under discussion (S4, 12%) or they responded with a yes/no answer (S5, 28%). This indicates that the dialogue was quite poor since the other subcategories that relate with students offering justification for their reasoning (S2, 10%) and providing examples or reporting data (S3, 4%), appeared only on rare occasions.

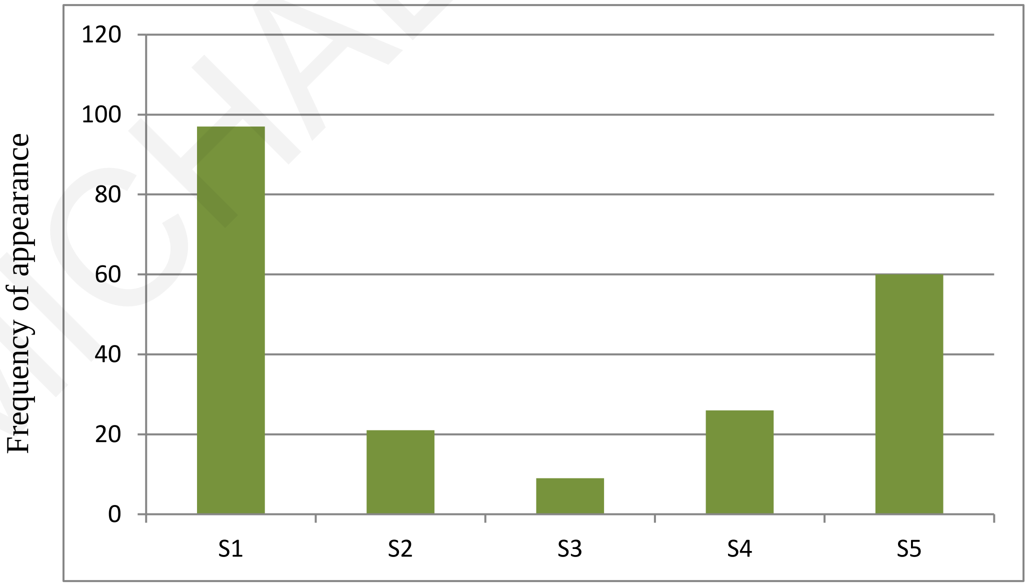

Figure 11: Frequency of appearance of students' response subcategories in the 3rd implementation.

The teacher recognized students' contributions in most of the cases by providing affirmation (R1, 53%) or by just acknowledging students' responses in a neutral way (R4, 33%). In some cases she readily offered the right answer to a question posed by herself or by a student (R2, 10%) while only rarely did she disconfirm students' response (R3, 4%).

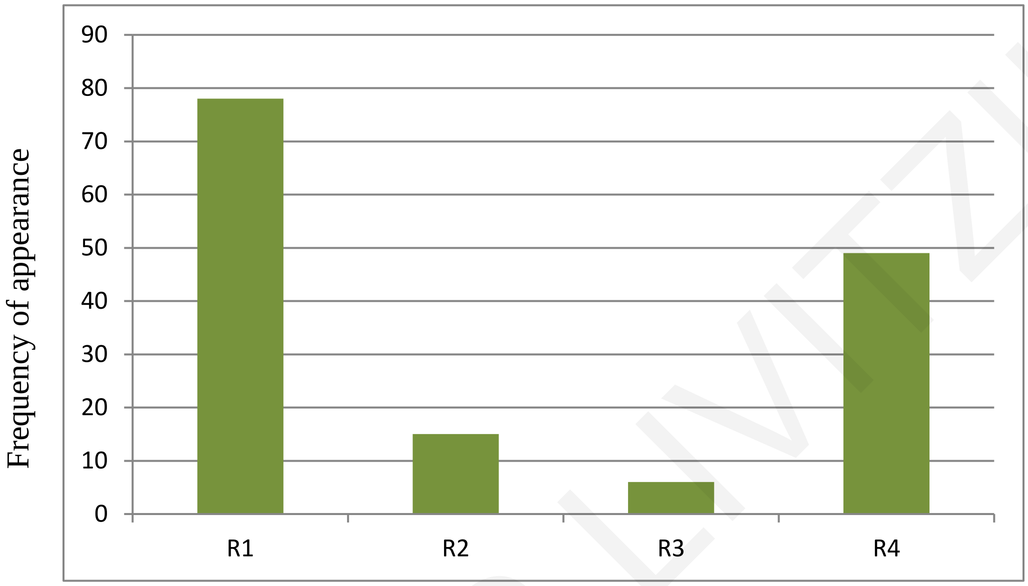

Figure 12: Frequency of appearance of Recognition subcategories in the 3rd implementation.

The teacher in many cases used the emergent information by focusing students' attention on stated opinions/data/examples with the intent to facilitate the discussion (U2, 28%). In a similar number of cases she tried to articulate consensus from a series of contributions that were exchanged (U4, 27%) where in somewhat fewer she tried to engage students in deeper reasoning/analysis on something (U3, 20%). In some cases she suggested or asked for an activity that could help students understand a specific issue (U1, 14%), or asked the opinion of other students about an expressed idea (U5, 11%).

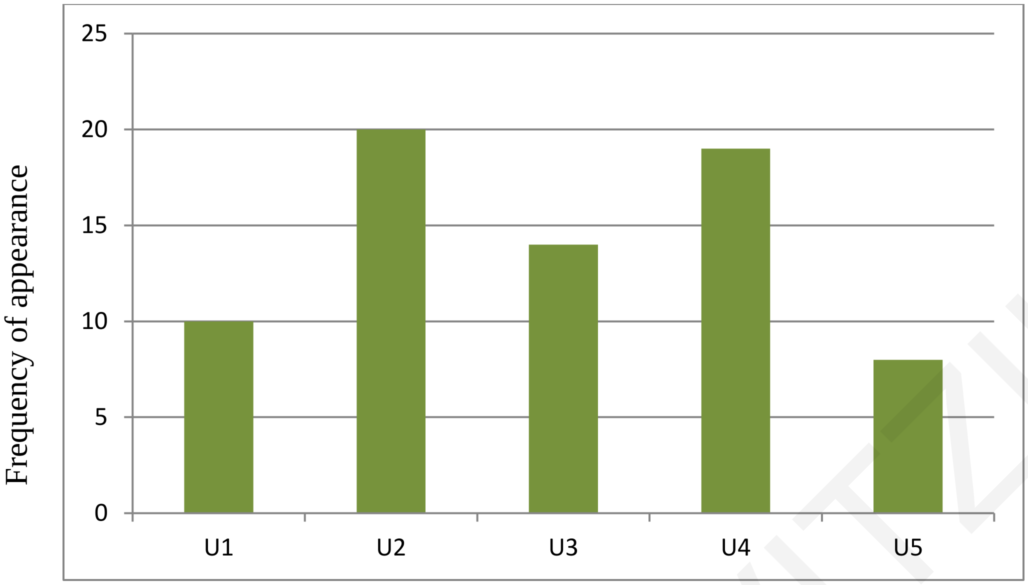

Figure 13: Frequency of appearance of Use subcategories in the 3rd implementation.

4.1.1. Clusters between ESRU subcategories

Looking at the frequencies with which the various subcategories appear in the dialogue from each implementation, there can be seen indications that there might be some relationships between the ways the teacher elicits students' ideas or recognises their contributions with the ways the students answer and the completion or not of ESRU cycles. However, it is not correct to extract conlusions comparing the implementations between them because they took place in different classes with different students and teachers. In order to examine whether the completeness of an ESRU cycle is influenced by the codes that appear in the cycle but also whether other codes of ESRU subcategories tended to appear together in the dialogue, we run cluster analysis on the codes that appear in each cycle of each implementation separetely.

1st Implementation (Vocational)

As it seems from the Dendrogram in Figure 14, R2 (teacher readily offers the right answer to a question posed by him/herself or by a student) along with R3 (teacher provides disconfirmation) combines with S4 (student explicates an inference/poses a question about an aspect of the topic under discussion). This might reflect the need of the teacher to respond with an immediate answer to students’ questions instead of using them for a productive dialogue. This is consistent with the further clustering on these by E3 (teacher sustains the cycle without using emergent information), which often results in short (S5, yes/no) answers from the part of students. However, the relation between S4 and R2 is not verified by Phi correlation (p>0.05), maybe because of the low appearance of these codes.

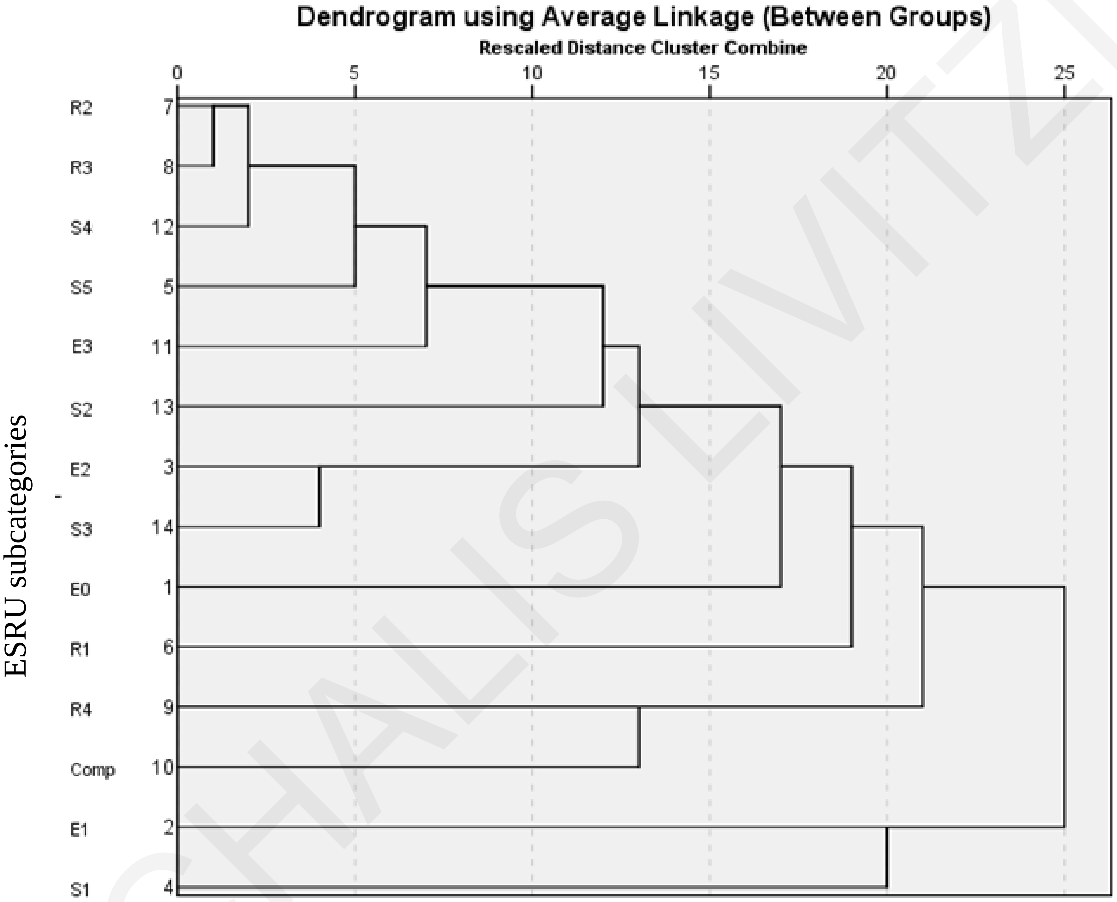

Figure 14: Tree diagram of subcategories and ESRU completeness clustering in the 1st implementation

Table 12 presents an example of such case, where a student has a query regarding the way another group carried out their experiment, asking twice (turns 1 and 3) how they handle the height they left the object to fall from. The teacher is providing an answer in both questions (turns 2 and 4).

Table **12**: *Excerpt from the 1st implementation, Lesson 2, Episode 1.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Student 1	Were they (the other group) changing the height?	S	S4
2	Teacher	They were changing the shape.	R	R2
3	Student 2	But was the height always stable?	S	S4
4	Teacher	Yes	R	R2

On the other hand, an interesting clustering is between E2 (teacher asks students an open-ended question) and S3 (students provide an example or report data) which indicates that when the teacher prompts more active and fruitful engagement of students in the discussion, they usually respond accordingly. This is verified from a Phi correlation between the two, where a quite strong and statistically significant correlation is revealed ($\varphi=0.6$, $p<0.01$).

The extract below (Table 13) is an example where the teacher asks the students to present their data in order to interpret them and come up with a conclusion. In turn 2 a student claims that the velocities they measured were much different between them for the same object falling under the same conditions. The teacher asks him to support it with data (turn 3) and the students responded appropriately (turns 3-5).

Table 13**:** *Excerpt from the 1st implementation, Lesson 1, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	How do you know that there are experimental errors?	U	U3
2	Student 1	Because the velocities were too different.	S	S2
3	Teacher	Tell us an example.	E	E2
4	Student 1	For example, the first measurement…	S	S3
5	Student 2	It was 3.21	S	S3
6	Student 1	The first measurement of the velocity of the 4-gr object was 3.49, then 3.27 and then 2.93.	S	S3
7	Teacher	Say the measurements again.	E	E2
8	Student 1	3.49, 3.27 and 2.93.	S	S3

Similarly, more specific questions to elicit students' ideas about a new concept/idea/relation (E1) usually result on shorter, unreasoned suggestions of a concept/relation by the students (S1). The Phi crosstab is weak but significant between the two items ($\varphi=0.2$, $p<0.01$). In Table 14 is an example where the teacher asks a specific question referring to a variable that doesn't influence the falling velocity of an object (turn 1) and a student responds suggesting one without providing any explanation of how s/he concluded that (turn 2).

Table 14**:** *Excerpt from the 1st implementation, Lesson 2, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What else doesn't influence the falling velocity?	E	E1
2	Student 1	The shape of the object	S	S1
3	Teacher	The shape of the object doesn't influence the falling velocity.	R	R1

Finally, the completeness of ESRU cycles (Comp) clusters together with neutral acknowledge of students' contributions by the teacher (R4). This is verified from a Phi correlation between the two items, where a medium but statistically significant correlation is revealed ($\varphi=0.4$, $p<0.01$). Below (Table 15) is an example of a part of the dialogue where the teacher accepts students' responds without confirming or disconfirming them (turns 3 and 4) and then uses them to problematize students (turn 6).

Table 15**:** *Excerpt from the 1st implementation, Lesson 2, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Which values are slightly different?	E	E1
2	Student 1	The 3rd one	S	S1
3	Teacher	This one…	R	R4
4	Student 2	And the last one	S	S1
5	Teacher	This one…	R	R4
6	Teacher	Which one do you think is the correct?	U	U2

We also ran the cluster analysis using as variables the various subcategories of Use instead of the variable "completeness of ESRU cycle" in order to identify possible relations between the Use subcategories and the rest subcategories of ESRU.

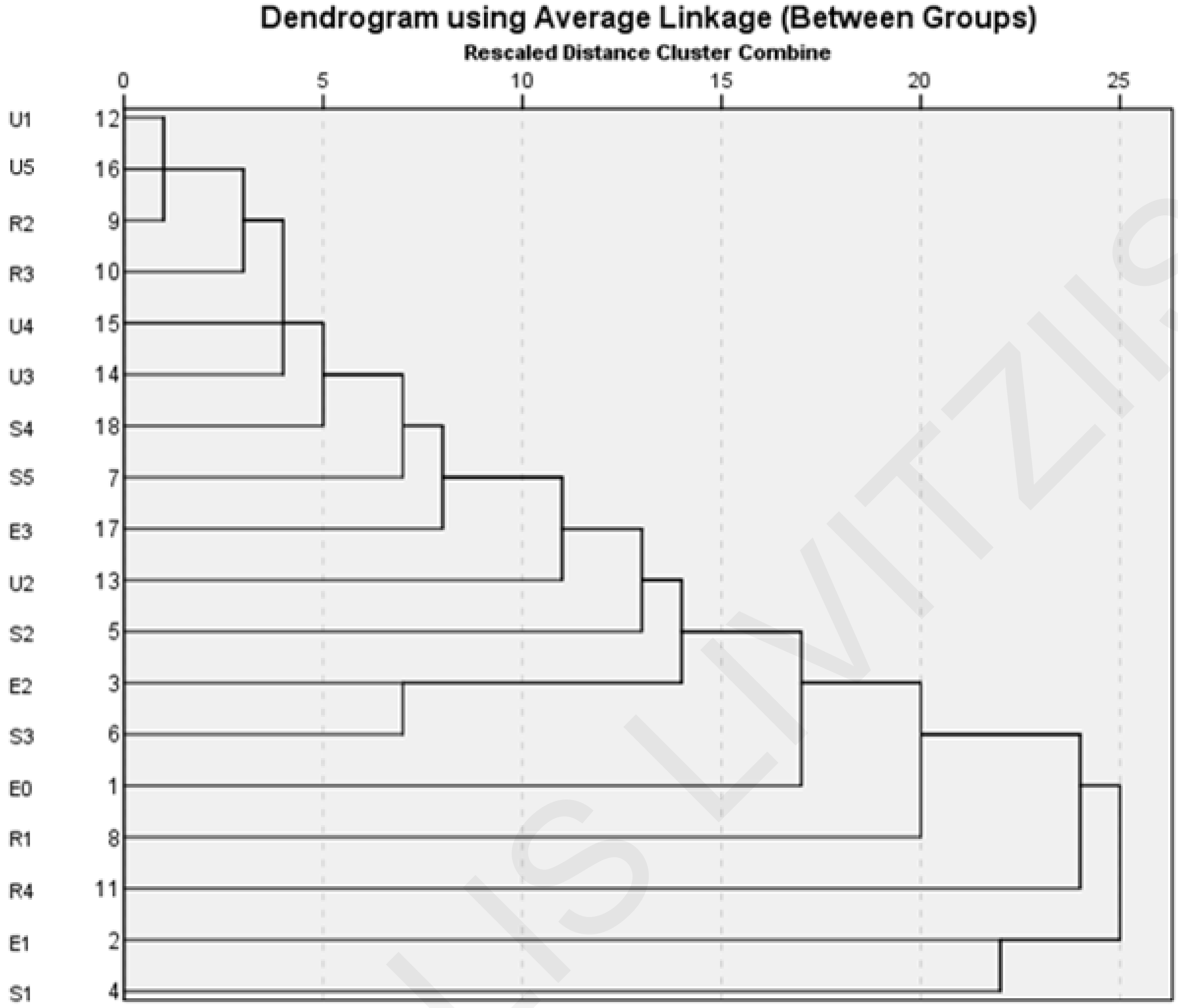

Figure 15: Tree diagram of all subcategories clustering in the 1st implementation

As we can see at the diagram, U1 (teacher suggests an activity that could help students resolve a specific conceptual issue) and U5 (teacher asks the opinion of other students about an expressed idea) subcategories cluster together with R2 (teacher readily offers the right answer to a question posed by him/herself or by a student). This might reflect a strategy the teacher follows whereas she suggests an activity or asks the opinion of other students only to confirm what she has already answered. However, the fact that also R3 (teacher provides disconfirmation) clusters also with the aforementioned subcategories, indicates that a similar strategy is followed in general when the students' answer is not the expected (correct) one. The following excerpt (Table 16) is an example where the teacher disconfirms student's suggestion regarding the extent of the experimental error (turn 3) and then asks for a process that could identify it turn 4.

Table 16: *Excerpt from 1st implementation, Lesson 2, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What could be another way that experimental errors existed?	E	E1
2	Student 1	Maybe the height was differing for some centimeters.	S	S1
3	Teacher	Well, maybe not centimeters, it is too big difference.	R	R3
4	Teacher	How could you identify how much is this difference?	U	U1

At further distance, U3 (teacher seeks to engage students in deeper reasoning on something) and U4 (teacher articulates the consensus from series of contributions that were exchanged) also combine with S4 (student explicates an inference/poses a question about an aspect of the topic under discussion). This depicts different strategies the teacher follows when students actively engage in the dialogue, sometimes opening the discussion asking for further analysis or explanation and sometimes bringing to a closure by summing up ideas. The next excerpt (Table 17) is an example where the teacher opens up the discussion after a student's inference (turn 4) by asking him to further explain himself (turn 4).

Table 17: *Excerpt from the 1st implementation, Lesson 3, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	If the spring was 10m, would the elongation still increase proportionally as in the graph you drew?	E	E1
2	Student 1	Yes, as long we increase the spring's length, it will increase too	S	S1
3	Teacher	So, the elongation will keep increasing proportionally.	R	R4
4	Student 2	No, I believe the length is influencing up to a particular value.	S	S4
5	Teacher	What do you mean?	(R)U	U3

2nd Implementation (Lyceum)

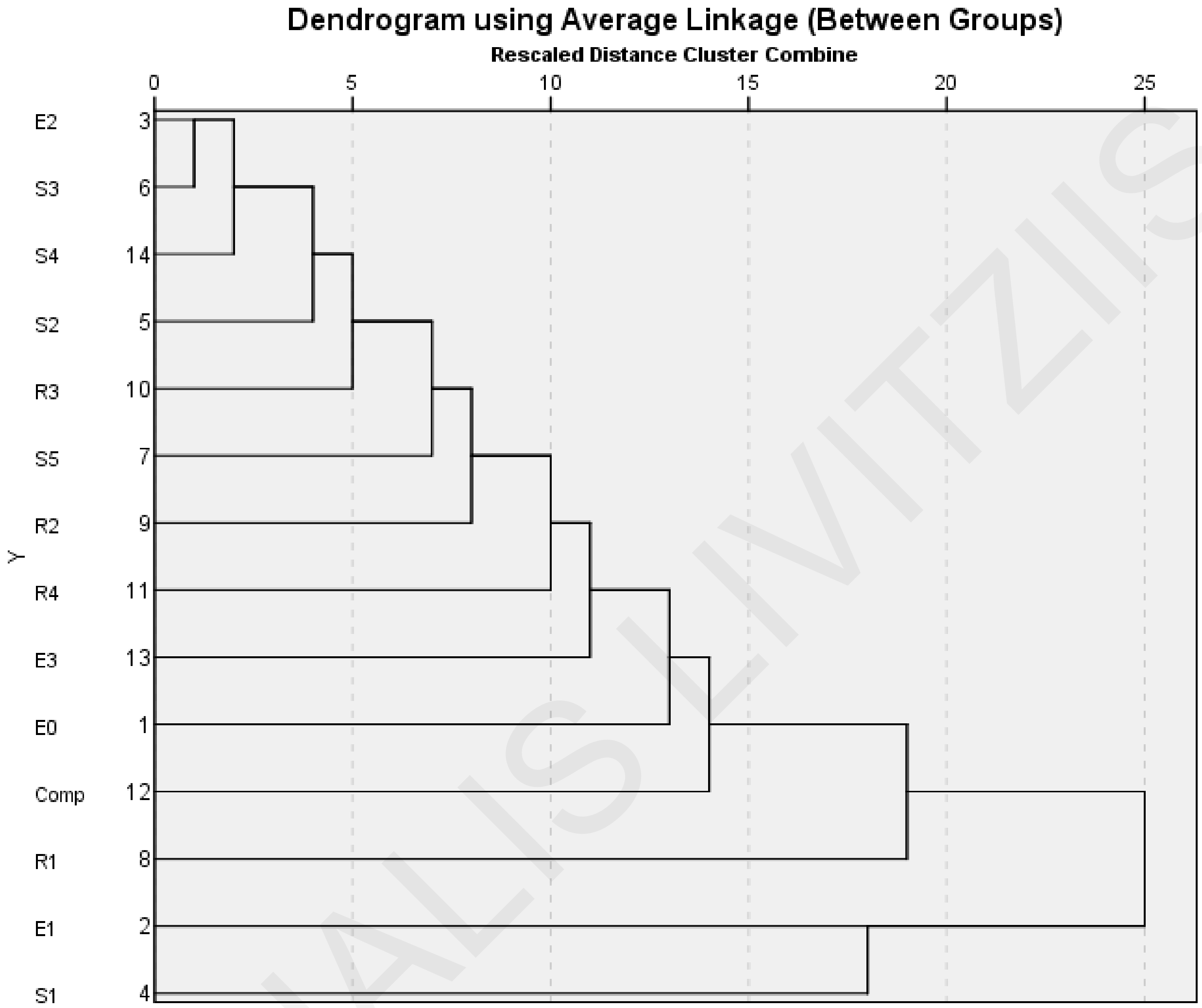

Figure 16: Tree diagram of subcategories and ESRU completeness clustering in the 2nd implementation

As it seems from the tree diagram in Figure 16 of the implementation at Lyceum, there is quite strong relation between E2 (teacher asks students an open-ended question) and S3 (students provide an example or report data) in this implementation also. This is verified from a Phi correlation between the two items, where a moderate and statistically significant correlation is revealed ($\varphi=0.5$, $p<0.01$). The example below (Table 18) illustrates how the prompt for an example by the teacher (turn 4) triggers a series of students' contributions (turns 5-7) who provide relevant examples.

Table 18: *Excerpt from the 2nd implementation, Lesson 1, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Any object can be a material body?	U	U3
2	Student 1	A moving object	S	S1
3	Student 2	An object we could exert force on it	S	S1
4	Teacher	Tell us an example	E	E2
5	Student 2	A car	S	S3
6	Student 3	The chair	S	S3
7	Student 4	The bottle	S	S3

Subcategory S4 (student explicates an inference/poses a question about an aspect of the topic under discussion) also combines in the aforementioned cluster as well as S2 (student offers justification for his/her reasoning) does so too. This indicates that in the instances that the teacher prompted more active engagement of students in the discussion, they responded accordingly.

Table 19: *Excerpt from the 2nd implementation, Lesson 4, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Right, while weight is increasing, acceleration remains the same.	U	U4
2	Student 1	But Dt (time the fall lasted) is decreasing.	S	S4
3	Teacher	Right, but what else is decreasing? Du (variation in velocity). So, acceleration, which is Du/Dt, proportionally remains the same.	R	R3
4	Student 2	But when the height is increasing, the acceleration decreases.	S	S4
5	Teacher	By how much does it drop?	R	R3
6	Teacher	This 8.4 could be 8.5 so if we just ignore the decimal part of the number this would round up to 9. This is also the case for these two measurements (points to other 2 measurements).	R	R2

However, it seems that R3 (teacher disconfirms students' response) clusters with the aforementioned cluster which might reflect the attitude of the particular teacher against students' expressed ideas in some cases. This illustrated on the extract below (Table 19), where two students have objections (turns 2 and 4) regarding the conclusions of the experiment as expressed from their classmate and affirmed by the teacher (turn 1). In both cases the teacher disagrees and illustrates for the students the “correct” answer (turns 3, 5 and 6).

On the other hand, more specific questions to elicit students’ ideas about a new concept/idea/relation (E1), usually result on shorter, unreasoned suggestions of a concept/relation by the students (S1) (Figure 15). Phi crosstab verifies this relation ($\varphi=0.2$, $p<0.01$) despite the weak. The example below (Table 20) is a characteristic one where the teacher asks for a specific term (resultant force) and the students respond only by providing the term.

Table 20: *Excerpt from the 2nd implementation, Lesson 5, Episode 5.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Which force causes the car to move?	E	E1
2	Student 1	The resultant	S	S1
3	Teacher	Neither the weight or the vertical but, the resultant.	R	R1

Finally, it seems from the Dendrogram (Figure 15) that the completeness of ESRU cycles (Comp) combines with E0 which is a code that stands for the cases where elicitation took place by using information from a previous cycle. This indicates that when emergent information is used, then it is likely that more useful information can emerge and be used productively. In addition, in this implementation also, examining the crosstabs between the cases that the ESRU cycle actually begins from using the information arose from the previous cycle (E0) and the various ways students respond (S), the only subcategory that is significantly related ($\varphi=0.2$, $p<0.01$) is S2 (student offers justification for his/her reasoning). This is the case in the excerpt below (Table 21), where the teacher prompts students’ thinking (turn 2) to elaborate experimental results and then, during the new cycle that emerge, she asks them to explain their thinking more thoroughly (turn 5).

Table 21: *Excerpt from the 2nd implementation, Lesson 5, Episode 5*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What happened to acceleration when we increased the slope?	E	E1
2	Student 1	It increased	S	S1
3	Teacher	It increased, didn't it?	R	R1
4	Teacher	So, what could we say about the relation between force and acceleration?	U	U2 (E0)
5	Student 2	That as larger the force is, the larger the acceleration gets.	S	S2
6	Teacher	How do we know that? Which other variable increases with larger the slope?	(R)U	U3

Cluster analysis using the subcategories of the "use" category was also carried out. As we can see at the diagram, U1 (teacher suggests an activity that could help students resolve a specific conceptual issue) and U5 (teacher asks the opinion of other students about an expressed idea) subcategories cluster together with E2 (teacher asks students an open-ended question) and S3 (students provide an example or report data). This indicates that productive use of the emergent information often took place when students were given the opportunity to present experimental data or examples to support their arguments.

At some further distance, U4 (teacher articulates the consensus from series of contributions that were exchanged) and S2 (student offers justification for his/her reasoning) also combine with the aforementioned subcategories, suggesting that this collection of information might have helped the teacher to lead discussion to a consensus. Also, R3 (teacher provides disconfirmation) combines with U2 (teacher seeks to focus students' attention on something with the intent to facilitate the discussion). This indicates that sometimes the teacher disconfirms students' ideas but, at the same time, tries to redirect their thinking towards the scientifically correct ones.

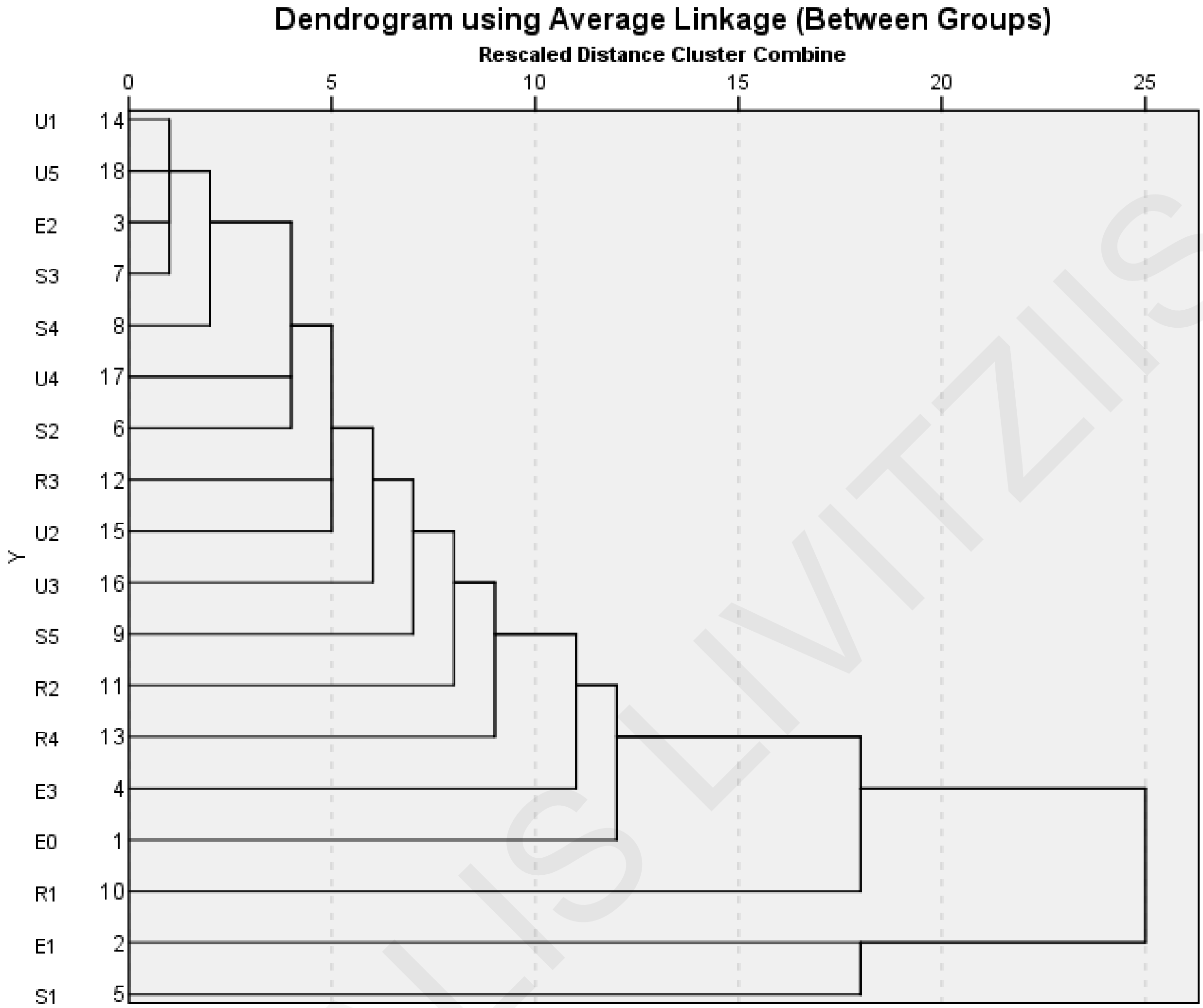

Figure 17: Tree diagram of all subcategories clustering in the 2nd implementation

3rd implementation (Gymnasium)

As is apparent from the tree diagram of the implementation at the Gymnasium, this implementation was no different when it comes to the quite strong linkage between E2 (teacher asks students an open-ended question) and S3 (students provide an example or report data). This verifies that when the teacher prompted more active engagement of students in the discussion, they usually responded accordingly. Phi correlation verifies this quite strong relation between these subcategories ($\varphi=0.6$, $p<0.01$). This is also illustrated at the example below (Table 22) where the teacher asks students to describe the experiment they did and the measurements they collected (turns 1 and 3) and the students respond accordingly (turns 2 and 4).

Table 22: *Excerpt from the 3rd implementation, Lesson 3, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Can somebody remind me what we did?	E	E2
2	Student 1	We had two vehicles, one at the bottom of the slope and one at the top. We put different masses on the vehicles and let the one move downwards and hit the other.	S	S3
3	Teacher	So, what did you observe? Regarding the measurements of the forces in all cases?	E	E2
4	Student 2	The force exerted from the one vehicle to the other has the same magnitude with the force exerted from the 2nd vehicle to the 1st.	S	S3

Despite that the teacher sometimes disconfirms students' ideas (R3) and provides the right answer (R2), students are given the opportunity to reason themselves (S2) or pose their questions and objections (S4) and then, the information is used in many of these cases. This can be seen on the diagram whereas, both the variable of the completeness of the ESRU cycle (Comp) and the code E0 that stands for elicitation that took place using information from previous cycle, combine with all the aforementioned subcategories (Figure 17).

Table 23: *Excerpt from the 3rd implementation, Lesson 3, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	In our case, objects A and B are the vehicles, but it could be an elephant and an insect, it would be the same.	U	U4
2	Student 1	I don't understand how this is possible.	S	S4
3	Teacher	It does sound strange...	R	R4
4	Student 2	If a small object hits us, we won't feel the same as if a big object did.	S	S4
5	Teacher	You need to differentiate between the magnitude of the force and the impact to the object.	R	R2
6	Teacher	Meaning, despite that the same force applies on the two objects, this doesn't mean they have the same result.	U	U4

Nevertheless, it is again the code R4 (neutral recognition) that combines closer to completeness of ESRU cycle and this is also verified by the moderate correlation between these items ($\varphi=0.4$, $p<0.01$).The extract below (Table 23) is an example were the teacher is summing up what it was discussed before (turn 1), giving the opportunity to students to express their queries (turns 2 and 4), to reveal their understanding and provide them more feedback (turn 6).

In addition, also in this implementation, examining the correlation between the cases that the ESRU cycle actually begins from using the information that arose from the previous cycle (E0) and the various ways students respond (S), the only subcategory that is significantly related ($\varphi=0.2$, $p<0.01$) is S2 (student offers justification for his/her reasoning). On the other hand, more specific questions to elicit students` ideas about a new concept/idea/ relation (E1) usually result on shorter, unreasoned suggestions of a concept/relation by the students (S1), which are only affirmed by the teacher (R1). The relation between E1 and S1 is confirmed as weak but significant by Phi correlation ($\varphi=0.2$, $p<0.01$) as well as the one between S1 and R1 ($\varphi=0.3$, $p<0.01$).The extract below (Table 24) is a characteristic one where the teacher asks for a specific interaction (turn 1), the student responds accordingly (turn 2) and the teacher recognizes the contribution by affirming it.

Table 24: *Excerpt from the 3rd implementation, Lesson 3, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Which object exerts this force, and to which is it applied on?	E	E1
2	Student 1	The man to the box.	S	S1
3	Teacher	Good! The man to the box.	R	R1

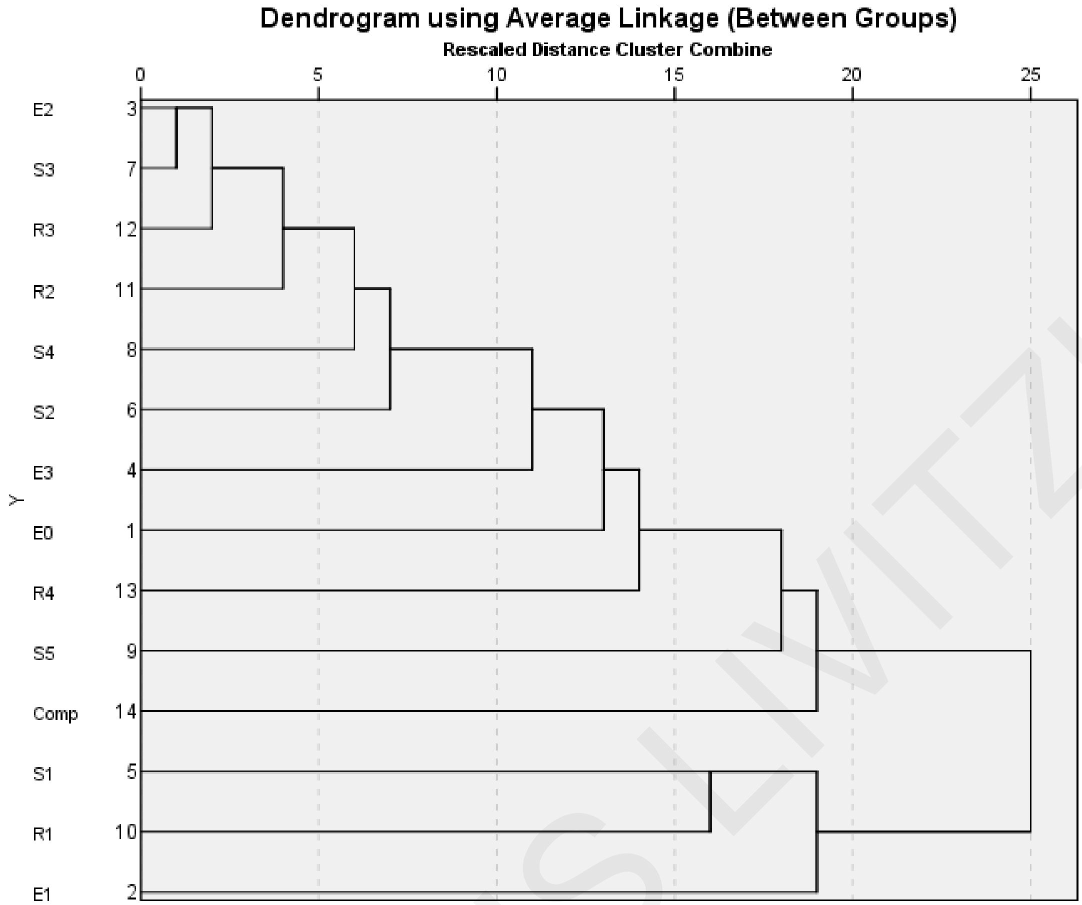

Figure 18: Tree diagram of subcategories and ESRU completeness clustering in the 3rd implementation

Cluster analysis using the subcategories of the “use” category was also carried out. As we observe in the diagram (figure 19), U5 (teacher asks the opinion of other students about an expressed idea) clusters with E2 (teacher asks students an open-ended question), S3 (Student provides an example or reports data) and R3 (Provision of disconfirmation). In the extract below (Table 25), the student is providing an example to make an argument (turn 1) and the teacher, despite she has already affirmed his idea (turn 2), she still asks for alternative opinions in an effort to engage other students too (turn 3).

Table 25: *Excerpt from the 3rd implementation, Lesson 5, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Student 1	In the example you told us before with the elephant and the man to demonstrate that the force is independent to the size of the objects…	S	S3
2	Teacher	Yes, right, we have measured it.	R	R1
3	Teacher	Does anybody have another opinion?	U	U5

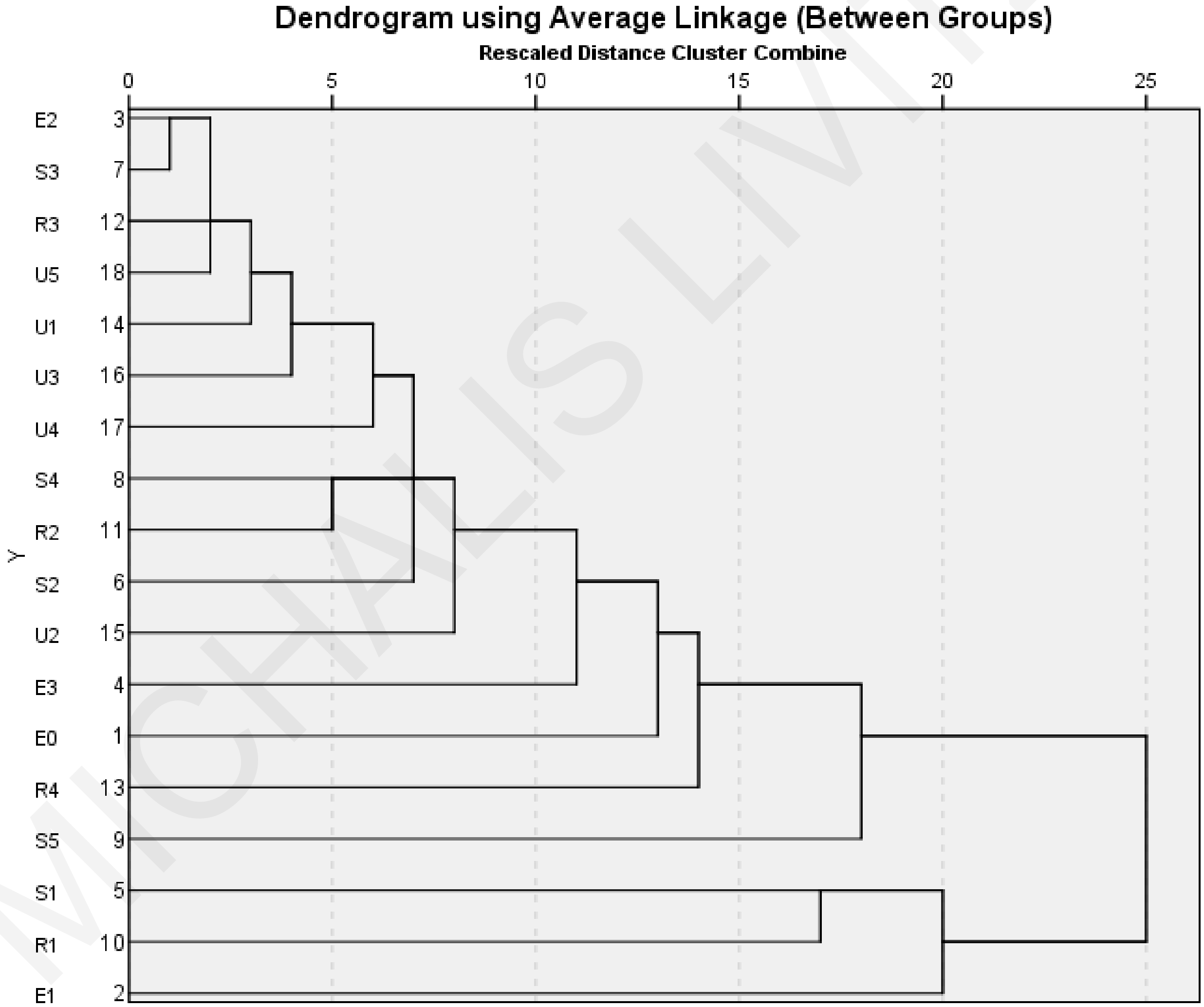

Figure 19: Tree diagram of all subcategories clustering in the 3rd implementation

This indicates that when the discussion is enriched with data or examples students provide, even if the teacher disagrees with their interpretation, the discussion remains open. This is supported by the fact that U1 (Teacher suggests an activity that could help students resolve a specific (conceptual) issue) and U3 (Teacher seeks to engage students in deeper reasoning on something) also cluster with those on the next level.

At a more distanced level, S4 (student explicates an inference/poses a question about an aspect of the topic under discussion) and R4 (teacher acknowledges a contribution made by the students in a neutral way) also cluster with all the different ways of using the emergent information, indicating that students' inferences are seen from the part of the teacher as opportunities to promote learning.

On the other hand, subcategories S5 as well as E1, S1 and R1 remain in separate clusters for a long distance.

Through all implementations

Comparing Profiles

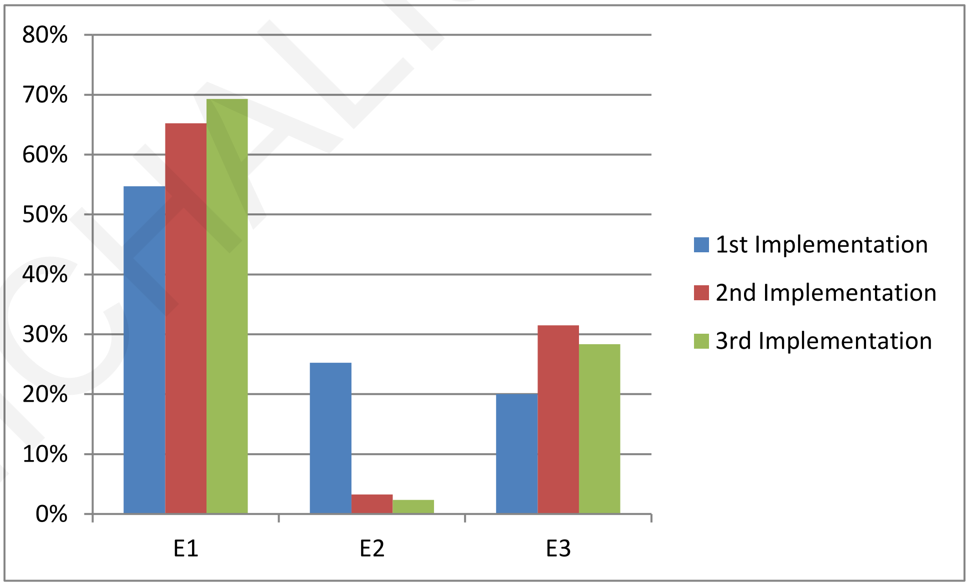

Figure 20: Percentage of Elicit subcategories in each implementation

Looking at the ways the teacher at the 1st implementation tried to elicit students' ideas; we observe that they seem mostly distributed between codes E1 and E2. This indicates an effort to engage them more actively in the dialogue by asking them to offer examples or report data from their experimental work while teachers at the 2nd and 3rd implementation seemed to ask mostly closed questions.

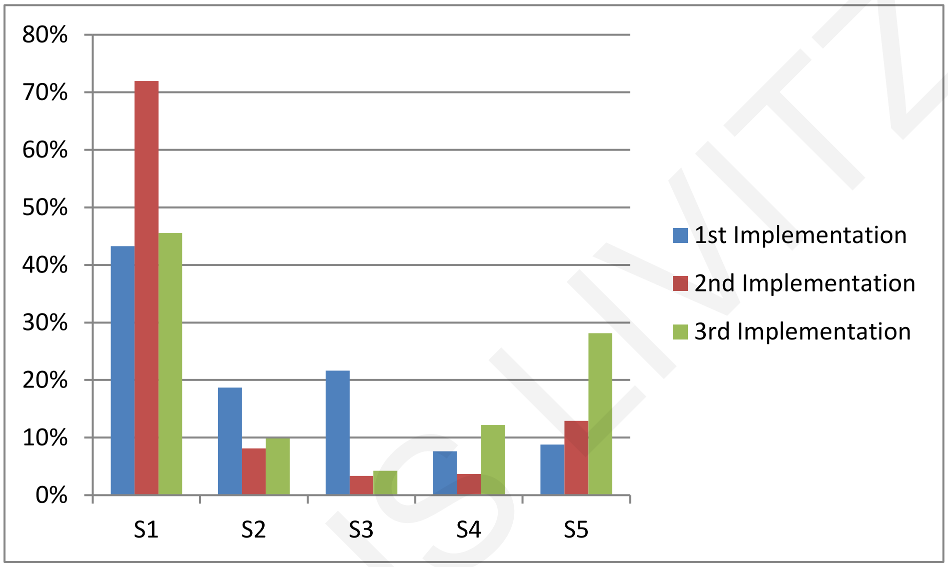

Figure 21: Percentage of Student's response subcategories in each implementation

There is a similar picture with students from the 1st implementation offering justification for their reasoning, providing examples or reporting data much more often from those of the 2nd and 3rd who are limiting their answers mostly on suggesting a concept/relation in response to questions posed by the teacher. This is an indication that the quality and richness of students' contributions might relate to the kind of question posed by the teacher.

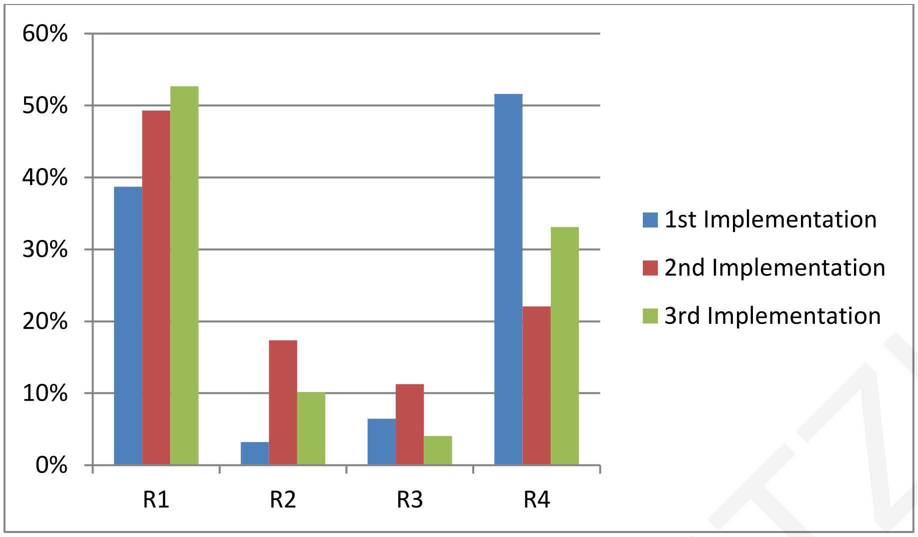

Figure 22: Percentage of Recognition subcategories in each implementation

Regarding the ways teachers recognized students' contributions in the dialogue it seems that all of them were keen to provide affirmation at about half of the cases. However, the teacher of the 2nd implementation was more willing to readily offer the right answer to a question posed by herself or by a student, while the teachers of the 1st and 3rd implementation acknowledged students' contribution in a neutral way often.

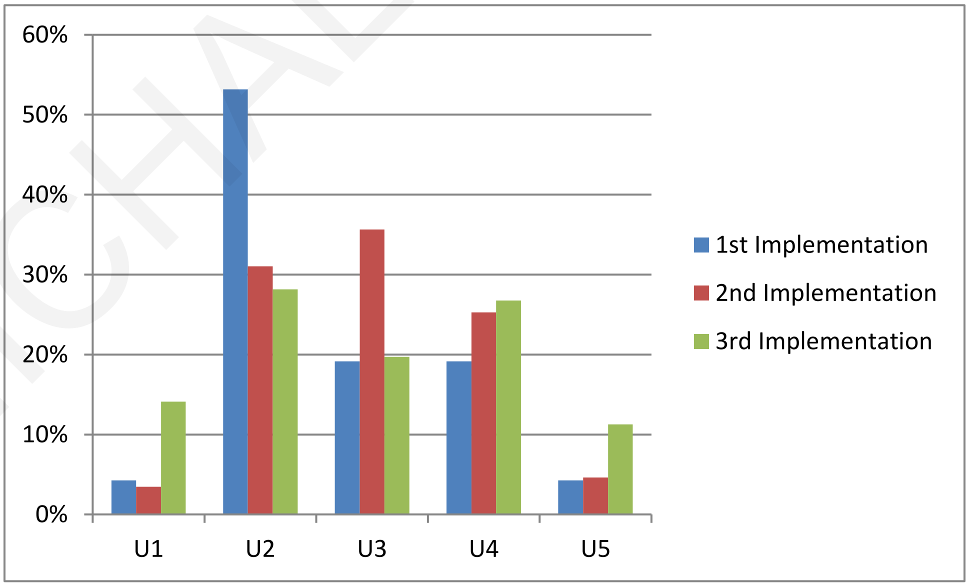

Figure 23: Percentage of Use subcategories in each implementation

The ways the teachers used the information emerged in the dialogue is quite common in the three implementations with codes U2 (teacher seeks to focus students' attention on something with the intent to facilitate the discussion), U3 (teacher seeks to engage students in deeper reasoning) and U4 (teacher articulates consensus from series of contributions that were exchanged) dominating in all cases. Nevertheless, in 1st implementation it was much more often that the teacher tried to focus students' attention on something with the intent to facilitate the discussion (U2) while in 2nd implementation more often the teacher tried to engage students in deeper reasoning.

However, the picture is different if we examine how often emergent information is used and the ESRU cycle is completed.

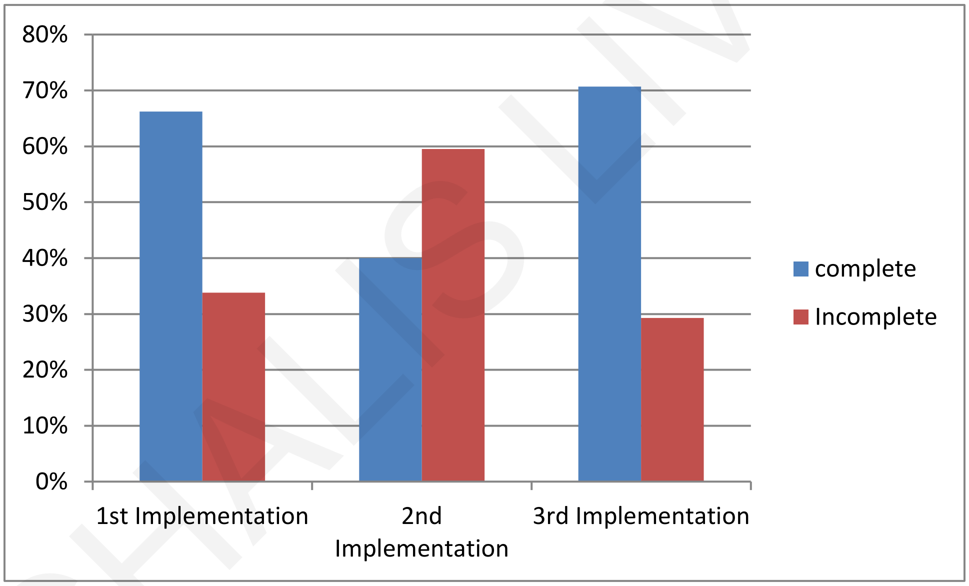

Figure 24: Percentage of complete and incomplete ESRU cycles in each implementation

In the case of the first implementation, we identified three cases of classroom dialogue that seemed consistent with the assessment method (i.e.., interactions on the fly). The coding of these dialogues revealed that the ESRU cycle was completed for 47 times, a number that corresponds to 66% of the instances, while for 24 times it remained incomplete, either as ESR or ES.

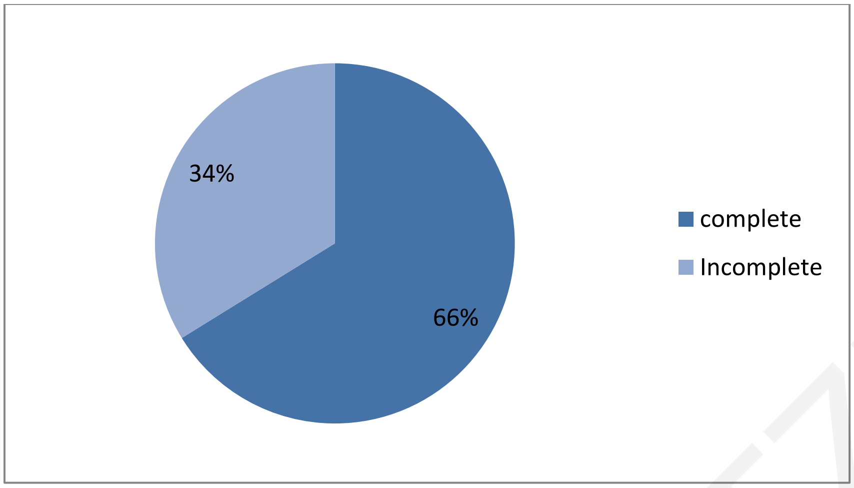

Figure 25: Percentage of complete and incomplete ESRU cycles in the 1st implementation

During the second implementation we identified 6 instances of dialogue, which involved a total of 81 completed ESRU cycles (41% of the instances) and 119 broken ESRU cycles (Table 2).

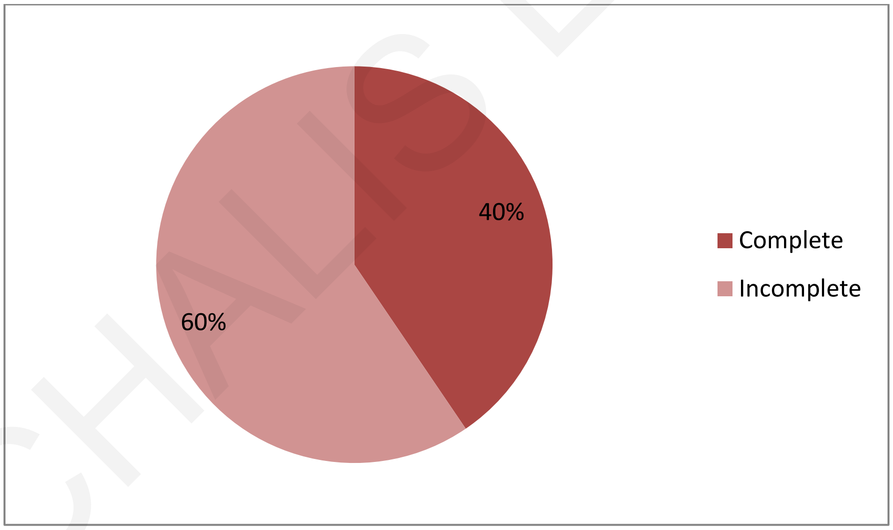

Figure 26: Percentage of complete an incomplete ESRU cycles in the 2nd implementation

During the third implementation we identified 6 instances of dialogue, which involved a total of 70 completed ESRU cycles (71% of the instances) and 29 broken ESRU cycles (Table 30).

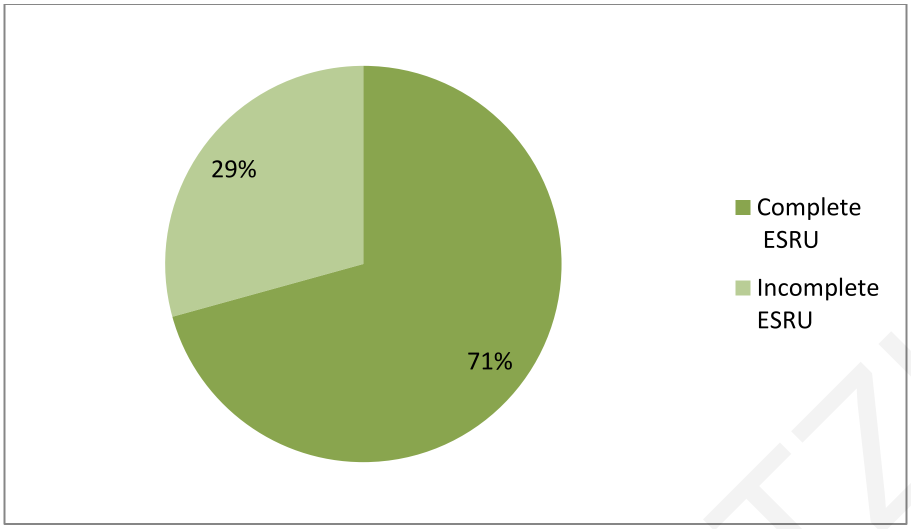

Figure 27: Percentage of complete an incomplete ESRU cycles in the 3rd implementation

It is revealed that the information is much more often used at the cases of the 1st and 3rd implementation than is not used, while the situation is completely reversed at the 2nd. We will discuss how the various patterns we have identified for the appearance of ESRU subcategories might influence the rate of completion of ESRU cycles.

Table 26: *Complete and incomplete ESRU cycles in each implementation.*

Variant of ESRU cycle	1st Implementation (frequency)	2nd Implementation (frequency)	3rd Implementation (frequency)
Complete ESRU	51 (68%)	81 (41%)	70 (71%)
Incomplete ESR or ES	24 (32%)	119 (59%)	29 (29%)

Looking at the cluster analysis of all the 3 implementations we notice that, in all of them subcategories E2 (teacher asks students an open-ended question) and S3 (students provide an example or report data) combine at short distance. S4 (student explicates an inference/poses a question about an aspect of the topic under discussion) also combine in the aforementioned cluster as well as S2 (student offers justification for his/her reasoning) does so. This is also verified by the cluster analysis we ran from all the data of the three implementations merged (Figure 28), supporting our hypothesis that when the teacher prompts more active engagement of students in the discussion, they usually responded accordingly. This may justify the high rate of completion of ESRU cycles at the 1st implementation where the teacher has a relatively high percentage (25%) eliciting students' ideas through asking them an open-ended question (E2). We suggest that this gave the opportunity to rich information about students' understanding to be revealed and the chance to the teacher to use it.

Another factor that might have helped towards the high rate of completion of ESRU cycles at the 1st implementation might be the quite high percentage (52%) of neutral recognition of students' responces, from the part of the teacher (R4). This is verified again at the diagram of all implementations (Figure 28) where we see R4 (neutral recognition of students' responces) combining with the variable Comp that stands for complete ESRU cycle. R4 was relatively high at 3rd implementation (33%) but quite low at the 2nd one (22%).

On the other hand, E1 (Teacher poses a closed-ended question to elicit students` ideas about a new (although interrelated) concept/idea/relation), S1 (Student suggests a concept/relation in response to question posed by the teacher) and R1 (Provision of affirmation) remain in separate clusters for a long distance (Figure 28) and in particular, they don't seem to lead to complete ESRU cycles. These categories had much higher percentage of appearance than the others at the case of the 2nd implementation where the rate of completion of ESRU cycles remained very low. It can be suggested that asking mainly closed questions about specific concepts and then, when a student response correctly, only affirming without prompting for further explanation, allows less information to emerge and be used during the dialogue.

As is apparent from the tree diagrams, especially for the 1st and 2nd implementations (Figures 14 & 16), R2 (teacher readily offers the right answer to a question posed by him/herself or by a student) along with R3 (teacher provides disconfirmation) combines with S4 (student explicates an inference/poses a question about an aspect of the topic under discussion). This might reflect the need of the teachers to respond with an immediate answer to students' questions instead of using them for a productive dialogue. However, frequency of S4 was much higher at the 3rd implementation than the others (especially than the 2nd implementation) and this can be a factor that influenced positively the rate of completion of ESRU cycles at the partcular implementation. An interpration for this might be that the attidute of teacher against their responses to the dialogue encouraged them to express their ideas, allowing better collection and use of information and consequently the rate of completion of ESRU cycles.

On the other hand, the relatively high percentage of R2 (teacher readily offers the right answer to a question posed by him/herself or by a student) along with R3 (teacher provides disconfirmation) at the 2nd implementation (17% and 11% respectively) (Figure 8), might had discouraged students from freely expressing their alternative ideas, negatively influencing the available information for the teacher to use, thus the rate of completion of ESRU cycles was also quite low.

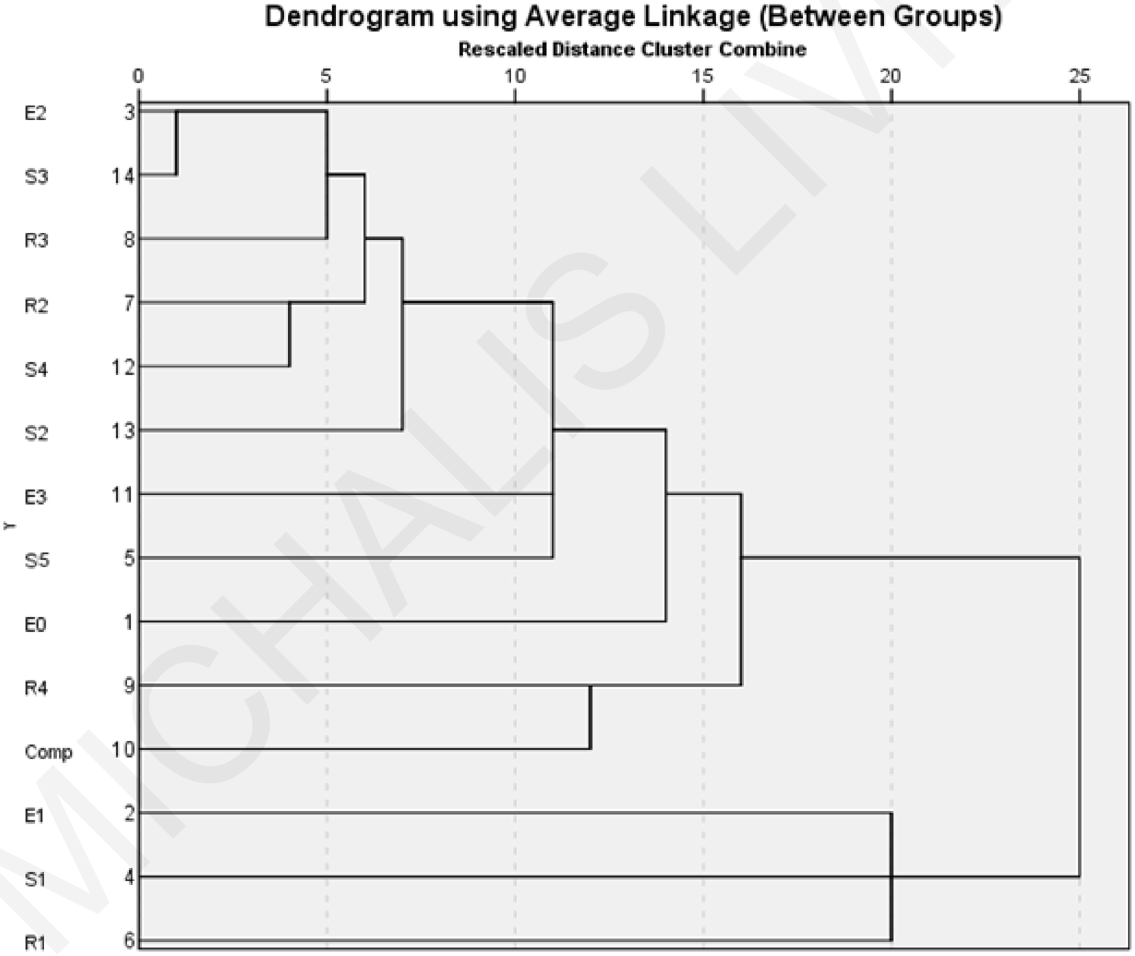

Figure 28: Tree diagram of subcategories and ESRU completeness clustering through all implementations

When it comes to the dendrogram from all implementations that includes the subcategories of Use (Figure 29), we observe again a connection between the subcategories U1 (Teacher suggests an activity that could help students resolve a specific issue), U5 (teacher asks the opinion of other students about an expressed idea) and R3 (teacher provides disconfirmation). This indicates that the discussion can be fruitful and students' contributions can be used to promote their learning even if the teacher expresses her disagreement when students don't respond the she expects.

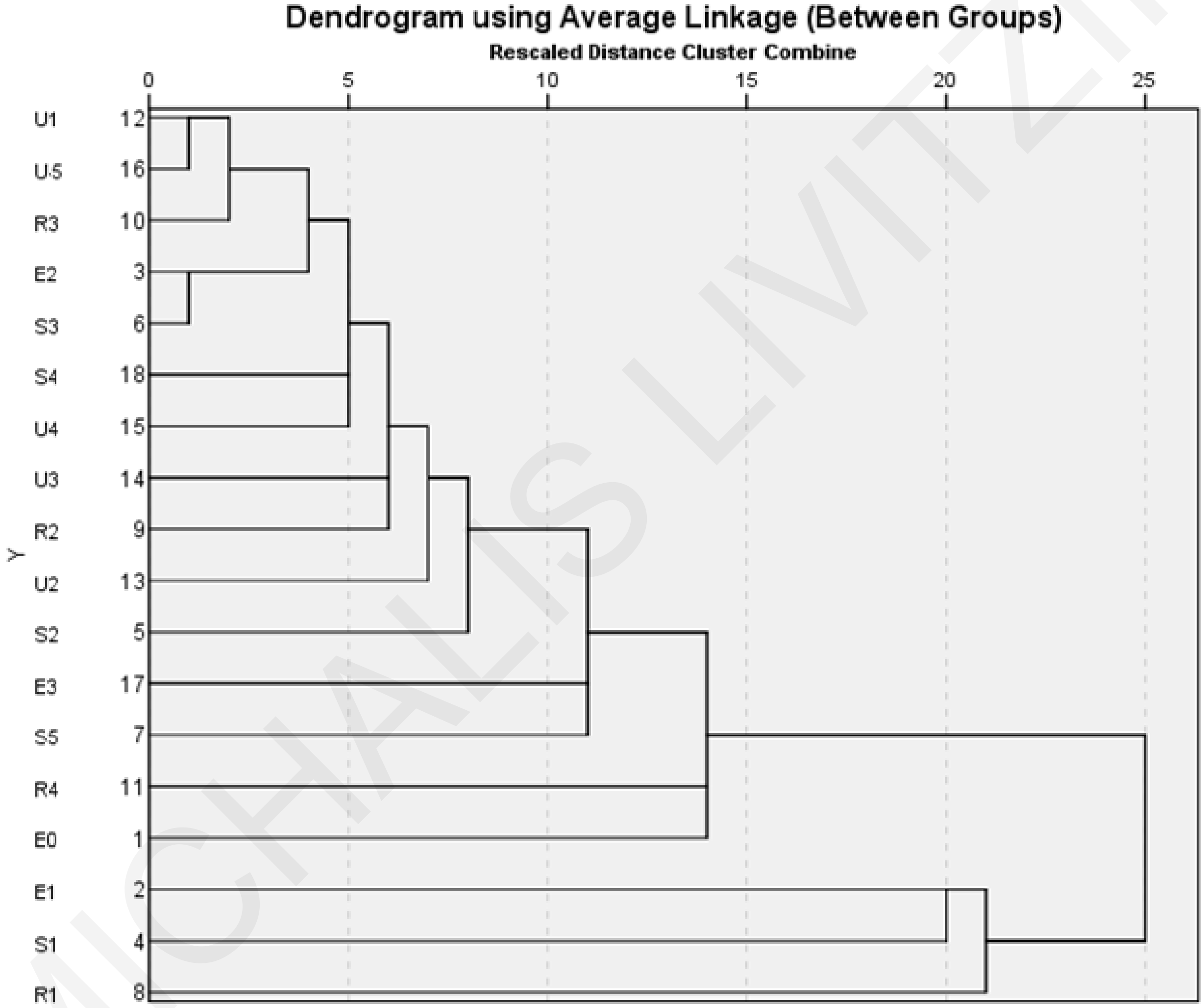

Figure 29: Tree diagram of all subcategories clustering through all implementations

4.2. Research Question 2: Factors facilitating or impeding teachers’ attempts to use *interactions on the fly* and missed opportunities.

Elaborating the instances where the various subcategories of Use appeared in the dialogue, we were able to categorize our interpretations, for the cases that information that comes up in discussion is used productively and complete ESRU cycles are realized, in 5 categories (Table 27). This categorization reveals that there are different ways of utilizing the feedback information. These extend from just summing up what is already said, to posing appropriate questions that could promote students thinking or, suggesting activities that could meet students’ needs as diagnosed by the teacher.

Table 27: *Variation of complete ESRU cycles.*

	Complete ESRU cycles
1	Teacher suggests an activity in order to investigate a hypothesis that emerged during the discussion.
2	Teacher poses a question that is intended to promote students' thinking about the topic being discussed.
3	Teacher poses clarification questions to help students further articulate a contribution they made.
4	Teacher takes the opportunity to use what student said in order to sum up.
5	Students express several ideas that allow the teacher to ask them to compare them.

The excerpts presented next aim to illustrate how the aforementioned categories appear in the dialogue. At the first one (Table 28), a student who observes that the velocity of an object increases while it moves downwards a declined rump, wonders whether the resultant force, from the weight and the vertical force the surface acts on it, also increases. The teacher takes the opportunity and asks the students to think, how to use the diagram they drew before to find the resultant force at the upper part of the rump, in order to compare it with resultant force at the lower part of the rump. That was an implicit way to suggest an activity that could help them to solve a possible misunderstanding and can be included at the 1st category of complete ESRU cycles (Table 27).

Table 28: *Excerpt from the 2nd implementation, Lesson 6, Episode 3*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Student 2	Further on, when the object goes faster, is the resultant force bigger?	S	S4
2	Teacher	That is a very good question: at the lower part of the ramp, velocity is larger than at the upper part as you have measured.	R	R4
3	Teacher	Is it because the resultant force is larger at the lower part? How could you investigate that? Could you use the diagram you drew before to do so?	U	U1

The next excerpt (Table 29) presents the effort of the teacher to help students realize the vertical force a surface acts on an object placed on it. In order to do so, she reminds students that Earth pulls her downwards (turn 5) and, while she is changing the place she stands on, she questions students what is preventing her from falling (turns 7, 9, 12, 14). This example corresponds with the 2nd category of complete cycles that is when the teacher poses a question that is intended to promote students' thinking about the topic being discussed.

Table 29: *Excerpt from the 3rd implementation, Lesson 4, Episode 5*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Is there any force acting upwards on the robot? [teacher asks while showing a picture of a robot standing on the floor]	E	E1
2	Student 1	No	S	S5
3	Teacher	Is anyone pushing me upwards right now?	E	E3
4	Student 1	No	S	S5
5	Teacher	Why am I not falling then? The Earth pulls me down, isn't it?	E	E3
6	Student 1	Yes	S	S5

7	Teacher	So, why I don’t fall?	U	U2
8	Student 1	Because you stand on your feet.	S	S3
9	Teacher	Who is holding me from falling?	U	U2
10	Student 1	You	S	S1
11	Student 2	Gravity	S	S1
12	Teacher	Now? [teacher stands on the chair]	E	E3
13	Student 1	The chair	S	S1
14	Teacher	Now? [teacher gets on the floor]	E	E3
15	Student 1	The floor, the floor!	S	S1

In the following excerpt (Table 30) the teacher is posing a question after student’s response in order to encourage the student to better explain him/herself and make his/her thinking more transparent. In the particular case the teacher asks the student to explain whether s/he defines as environment of an object the room that the object is placed in (turn 7) when the student mentioned the place around it (turn 5). This example corresponds with the 2nd category of complete cycles that is where the teacher poses clarification questions to help students further articulate a contribution they made.

Table 30: *Excerpt from the 2nd implementation, Lesson 1, Episode 3*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What would you call environment for an object?	E	E1
2	Student 1	The room.	S	S1
3	Teacher	The room.	R	R4
4	Teacher	What else?	E	E3
5	Student 2	The place around.	S	S1
6	Teacher	The place around the object.	R	R4
7	Teacher	So you mean the classroom is the environment of this particular object? [the teacher shows a pen that she holds in her hand]	U	U3

At the next excerpt (Table 31), a student tries to justify why the direction of the forces acted from the table on an object placed on it is upwards. The teacher uses his response and tries to provide the class with a more complete justification. This refers to the 4th category of complete cycles that is where the teacher takes the opportunity to use what student said in order to sum up.

Table 31: *Excerpt from the 2nd implementation, Lesson 1, Episode 6*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Which are the forces on the object? The first is the weight, from the Earth. The other is right this way? Is it acted from the table on the object?	E	E1
2	Student 1	Yes.	S	S5
3	Teacher	Yes.	R	R1
4	Teacher	Why is it pointing upwards?	E	E3
5	Student 2	Because the table acts this force.	S	S3
6	Teacher	The table acts this force.	R	R1
7	Teacher	The table tries to push the object upwards.	U	U4

Finally, the next excerpt (Table 32), depicts how the teacher is using different opinions expressed by students giving them the chance to present their arguments (turn 4), encouraging also other students to participate in the discussion (turn 11). Thus, these contributions refer to the 5th category of complete ESRU cycles, which is when the students express several ideas that allow the teacher to ask students to compare them.

Table 32: *Excerpt from the 3rd implementation, Lesson 4, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What do you in the 2nd team think about the astronaut?	E	E1
2	Student 1	When he is in the atmosphere the Earth attracts him, but when he is outside the atmosphere, it doesn’t.	S	S1
3	Teacher	So, if I understand well your idea, outside the atmosphere there is no weight acting on him, so he will be suspended, but in the atmosphere the Earth attracts him.	R	R4

4	Teacher	Does anybody have a different opinion?	U	U5
5	Student 2	Yes, even outside the atmosphere, the nearest planet will act some force on him.	S	S2
6	Teacher	Either outside or inside	R	R4
7	Student 2	Outside might be less than inside	S	S1
8	Teacher	So you also think that there will be a difference at the point the atmosphere ends, but still the force won't be zero outside.	R	R4
9	Student 2	Yes	S	S5
10	Teacher	There will be some.	R	R4
11	Teacher	Who agrees with Student 1? Who agrees with Student 2? Who has a different idea? Who thinks that in general the atmosphere plays some role in gravity?	U	U5

Similarly, we categorized our interpretations for the cases that information coming up in discussion is not used productively and complete ESRU cycles are not realized, in 6 categories (Table 33). We are going to present an example for each category to explain the categorization.

Table 33: *Variation of incomplete ESRU cycles*

	Incomplete ESRU cycles
1	Teacher ignores/rejects an answer, repeating the question.
2	Teacher provides the right answer
3	Teacher ignores alternative ideas, choosing to respond to a correct one
4	Teacher stops a discussion on an emerged matter
5	Teacher poses a problematically expressed question/leading to yes/no/guided answer
6	Teacher accepts right answer without prompting for further interpretation

The 1st excerpt (Table 34), which corresponds to the 1st category of incomplete ESRU cycles, is an example where the teacher expects the students to understand that an alteration of the motion of an object corresponds to alteration of its velocity. Students' answers reveal that either they didn’t understand the question (turn 2), or they didn’t understand the connection between motion and velocity (turns 4 and 6). Nevertheless, the teacher neither encourage them to explain themselves nor poses a new question to promote their thinking, but rather seems to ignore their responses and she repeats the question (turns 3 and 5) which is considered as not fruitful use of students’ contributions and thus, these ESRU cycles (turns 1-2 and 3-4) are considered as incomplete.

Table 34: *Excerpt from the 1st implementation, Lesson 1, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What we mean by changing the kinetic status of the object? What is the physical magnitude that is changing?	E	E1
2	Student 1	I don’t know, I don’t remember.	S	S5
3	Teacher	What is changing?	E	E3
4	Student 2	Motion	S	S1
5	Teacher	What is the physical magnitude that is changing?	E	E3
6	Student 3	The direction	S	S1

The next excerpt (Table 35) is the continuing of the previous one (Table 35), where the teacher eventually decides to provide the right answer herself instead of using students’ answers in a way that could help them understand what she aimed for. That is why this ESRU cycle is considered as incomplete and is categorized in the 2nd category (table 33).

Table 35: *Excerpt from the 1st implementation, Lesson 1, Episode 1*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Which magnitude’s is the direction changing?	U	U3
2	Student 1	The object’s	S	S1
3	Teacher	How is the direction of the object changing? Isn’t the velocity that is changing when motion is changing?	R	R2

In the next excerpt (Table 36), it is presented a part of the dialogue where is discussed what a material body is. The teacher presents the cases of water and air and in both cases, there are students who think they cannot be considered as material bodies (turns 2, 8) and students who think that they can (turns 3, 9). Teacher does not ask the students in the first case to explain their idea and only confirms the idea of the second (turns 4, 10).

Table 36: *Excerpt from the 1st implementation, Lesson 1, Episode 2*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Is water a material body?	E	E1
2	Student 1	No, it is liquid.	S	S1
3	Student 2	Yes, it is (Material body)	S	S5
4	Teacher	It is a material body isn't it?	R	R1
5	Teacher	Could we touch it?	E	E1
6	Student 2	Yes	S	S5
7	Teacher	The air that we cannot touch isn't a material body?	E	E1
8	Student 1	No!	S	S5
9	Student 2	But air has molecules also.	S	S4
10	Teacher	That's right!	R	R1
11	Teacher	Any material in universe can be "material body"	U	U4

The next excerpt (Table 37) is from a part of the dialogue where the discussion is about the term "environment" for an object, in terms of physics. A student suggests that the air in the room is environment for the object the teacher holds in her hand (turn 2). The teacher seems to accept the answer initially but then she has doubts whether student's answer is correct (turn 3). However, at the end, she decides not to further discuss the issue. Thus, we coded this cycle as incomplete and categorized it at the 4th category of Incomplete ESRU cycles which refers to cases where teacher stops a discussion on an emerged matter (Table 33).

Table 37: *Excerpt from the 1st implementation, Lesson 1, Episode 3*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Any other body that can be considered as environment for this object [the pen that teacher holds in her hand]?	E	E1
2	Student 1	The air in the room	S	S1
3	Teacher	The air in the room. Correct, because it is around it. But… ok, let’s say that air can be considered as environment for this object.	R	R1

Table 38: *Excerpt from the 1st implementation, Lesson 1, Episode 4*

Turn		Speaker	Transcript	ESRU category	ESRU sub-category
1		Teacher	So the object falls because of gravity?	U	U3
2		Student 1	Yes	S	S5
3		Teacher	And why does gravity exist?	E	E1
4		Student 2	So that objects do not suspend!	S	S1

The excerpt (Table 38) above is showing a case where a student responds with yes/ no answer (turn 2) because the question is closed-ended one, that only leads to such an answer (turn 1). There is also the case that the student provides an irrelevant answer because the question is vague. The question in turn 3 is rather philosophical as it is expressed than physical. Probably the teacher wanted to ask what causes gravity and not why does it exist. Nevertheless, we consider these as examples where no valuable information is emerging during the particular cycles to be used and they have been categorized at the 5th category of incomplete ESRU cycles (Table 33).

Finally, the next excerpt (Table 39) presents an example where the teacher accepts as correct the student’s response (turn 4) while the student seems quite unsure and the teacher could encourage him/her to better explain his/her thinking. This cycle falls in the 6th category of incomplete ESRU cycles since teacher accepts the right answer without prompting for interpretation (Table 33).

Table 39: *Excerpt from the 1st implementation, Lesson 4, Episode 2.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	Motion	R	R1
2	Teacher	What kind of motion?	U	U3
3	Student 1	Acceleration?	S	S1
4	Teacher	Acceleration wonders Student 1. Do you agree? Acceleration, correct.	R	R4

Next, we present the frequencies in which each category of incomplete ESRU cycles appeared in each of the three implementations.

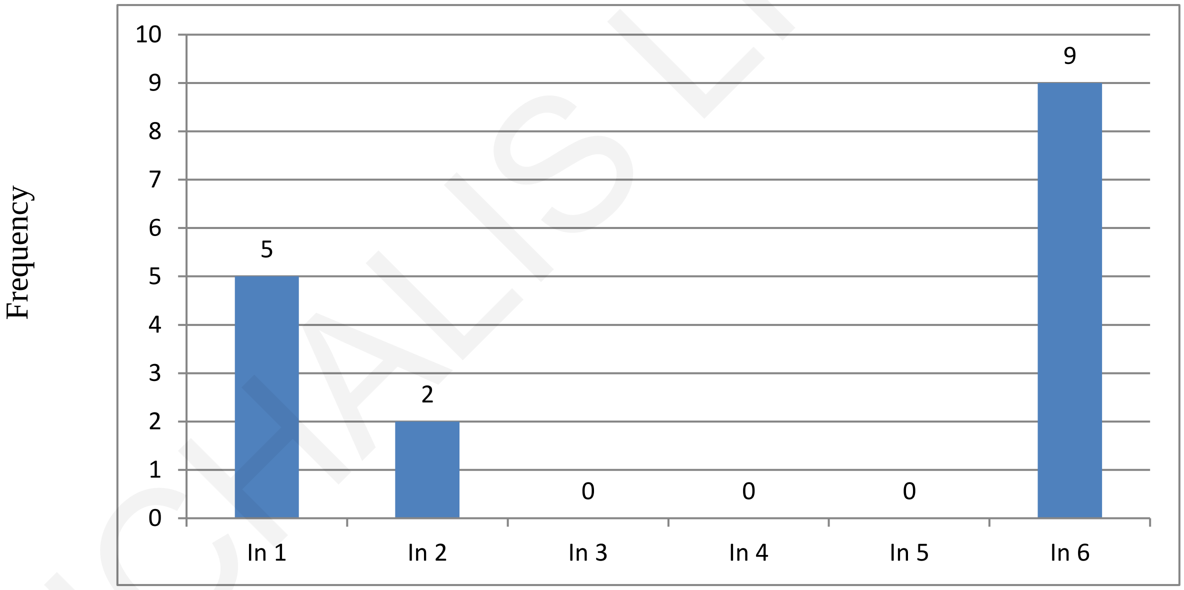

Figure 30: Frequency of various types of incomplete ESRU cycles in the 1st implementation

In the case of the 1st implementation, most of the instances where the ESRU cycle remained incomplete were those where the teacher accepted a right answer without prompting for further interpretation (In6, 56%,), while it seemed worthy to do so. In many cases she ignored or rejected an answer and then repeated the question (In1, 31.5%) or responded to a correct one (In2, 12.5%).

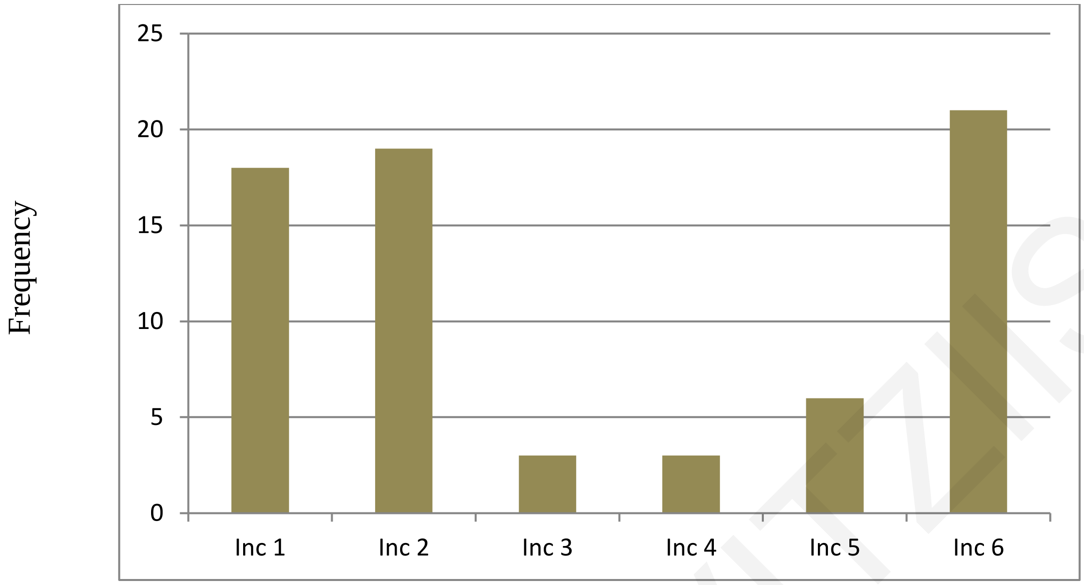

Figure 31: Frequency of various types of incomplete ESRU cycles in the 2nd implementation

In the 2nd implementation, the most frequent category where the ESRU cycle remained incomplete were the case where the teacher accepted a right answer without prompting for further interpretation (In6, 30%,), while it seemed worthy to do so. In similar number of cases she ignored or rejected an answer and then, repeated the question (In1, 26%) or responded to a correct one (In2, 27%). In some cases the teacher posed a question leading to yes/no or other closed answer (In5, 9%). In fewer cases, she ignored or rejected an alternative idea, choosing to respond to a correct one (In3, 4%) or stopped a discussion on an emerged matter (In4, 4%).

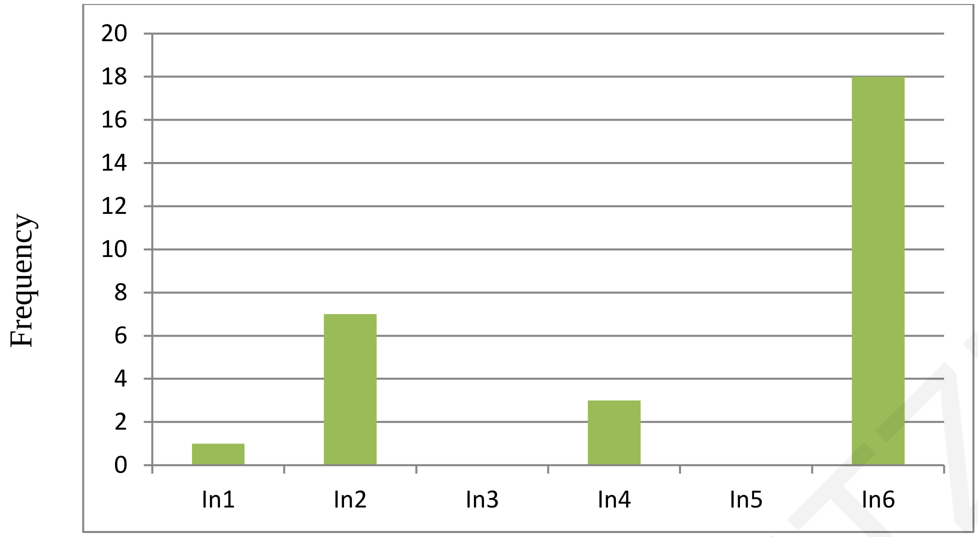

Figure 32: Frequency of various types of incomplete ESRU cycles in the 3rd implementation

In the 3rd implementation, the most frequent category where the ESRU cycle remained incomplete were the case where the teacher accepted a right answer without prompting for further interpretation (In6, 62%,), while it seemed worthy to do so. In less cases she provided the answer herself (In2, 24%) and even less times she just stopped a discussion on an emerged matter (In4, 10%). Only once (4%) ignored or rejected an answer and repeated the question (In1).

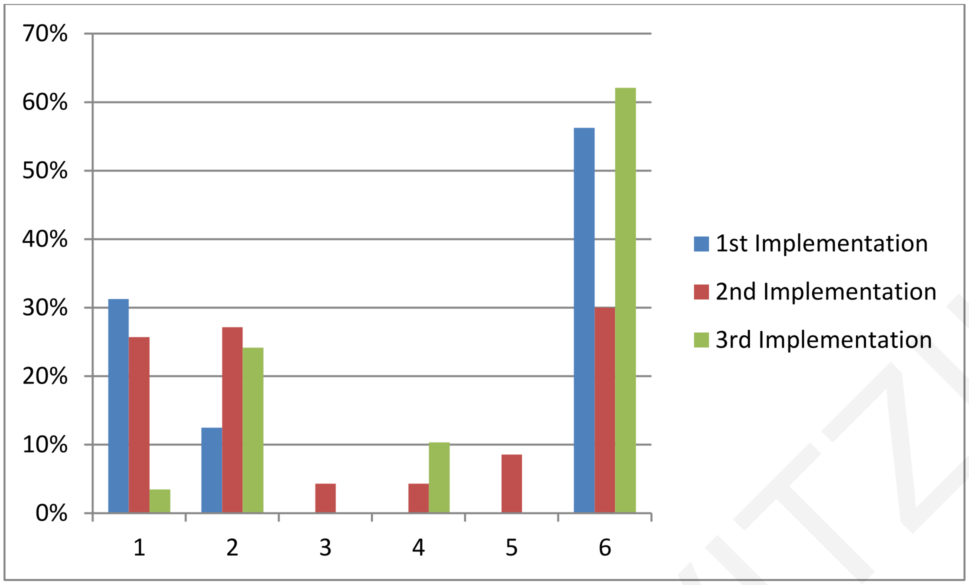

Figure 33: Frequency of various types of incomplete ESRU cycles in all implementations

Distribution at the categories of incomplete ESRU cycles is quite similar between the second and third implementation with cases of In6 (teacher accepts right answer without prompting for further interpretation) being more frequent at the 1st one and 3rd ones and cases of In2 (teacher provides the right answer) being more frequent at the 2nd and 3rd. On the other hand, cases of In3 (teacher ignores/rejects an alternative idea, choosing to respond to a correct one), In4 (teacher stops a discussion on an emerged matter) and In5 (teacher poses a problematically expressed question/leading to yes/no/guided answer) appeared only in 2nd and 3rd implementation for a few times.

The cases above highlight instances where the dialogue either halted or did not reach fruition resulting in broken ESRU cycles. There were also instances where the cycle was broken but the discussion got redirected or ideas got picked up later. In these latter broken cycles, it was clear that the teacher, having provided thinking time both for themselves and for further deliberation by the students, redirected the talk back to pick up unfinished ideas. So, while sometimes the breaks in the cycle indicated instances where teachers manufactured more thinking time within the flow of inquiry ideas, there were also instances where the breaks in cycles identified points where there might have been opportunity to guide or consider ways forward.

Missed opportunities

In an attempt to shed more light into the intricacies underlying the teachers' attempt to employ interactions on the fly as a formative assessment method, we focused on instances where either important information (i.e., contributions made by the students) went unnoticed during the discussion or was used in a non-optimal manner. In order to do so, we look at each episode at a macroscopic level so that we can realize the impact of teacher's choices during the dialogue on students' understanding.

In some cases, it seems that is a strategic choice made by the teacher not to address issues that come up during discussion, despite that misconceptions might underlie there, which influence students' understanding on the subject under discussion, when she considers that those issues are not part of the particular lesson.

Table 40: *Excerpt from the 1st implementation, Lesson 1, Episode 6.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	What can we think (which forces are applied on the object)?	E	E1
2	Student 1	Earth and the table apply forces on the object.	S	S1
3	Student 2	And the object applies force on the table.	S	S1
4	Teacher	Leave that for now.	R	R3
5	Teacher	What are we interested now?	E	E3
6	Student 1	The forces that are applied on the object.	S	S1

In the above excerpt (Table 40) from the classroom dialogue, the teacher asks about the forces that are applied on an object placed on a table (turn 1). Student 1 responds giving the correct answer (turn 2) but then Student 2 indicates that the object is also applying a force on the table (turn 3). The teacher asks to skip this for the moment and focuses the discussion on the forces that are applied on the particular object. She prefers not to discuss anything about action-reaction despite that this issue appears also in another part of the dialogue, indicating that it influences students understanding on the subject under discussion (which forces are applied on an object and the direction of each one).

Table 41: *Excerpt from the 1st implementation, Lesson 1, Episode 6.*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	So, this force is due to the direct contact with the table. Because the table is in contact with the object, some said that it exerts a force downwards and others upwards.	R	R4
2	Teacher	Who is correct?	U	U5
3	Student 1	Since the object is still, it means that the body exerts a force on the table and the table exerts a force of the same magnitude on the object.	S	S3
4	Teacher	You are right, but these forces are not exerted on the object. We want to isolate the object and the forces exerted on this.	R	R3
5	Teacher	Which are these?	E	E3

At the excerpt above (Table 41), the teacher recognizes that there is a confusion regarding the interaction between the table and the object placed on it (turn 1) and she prompts students to elaborate more on this (turn 2). Student 1 tries to explain the stability of the object through the action and reaction forces between the object and the table (turn 3), which is incorrect because they are exerted on different objects. The teacher indicates the last but, she doesn't provide feedback on why these forces exist (action-reaction) and why they cannot be accounted for the stability of the object. Instead, it seems that, the need to focus on the forces exerted on the particular object is rather procedural (turn 4). Despite that the issue of action-reaction occurs repeatedly in the discussion, the teacher decides not to discuss it although it could be related to the difficulties they have in matters relating to the discussion like the point of force application and the direction of the force.

As the teacher revealed when she was asked about this episode during the follow-up interview, she realized that the student was referring to the "reaction" on the other object. However, she chose to discuss this issue at a following lesson as part of teaching Newton's 3rd law of motion:

> *"I couldn't discuss the action-reaction issue before they understand well the 2nd law of motion. They (the students) could get really confused".*

Next, we consider the following excerpt (Table 42) from the whole class discussion, where the teacher tries to help students define the notion of the "environment" for an object in the context of dynamics, discussing about a small ball that she holds with her hand.

Table 42: *Excerpt from the 2nd implementation, Lesson 1, Episodes 3-4*

Turn	Speaker	Transcript	ESRU category	ESRU sub-category
1	Teacher	T: Are plants, sun, atmosphere part of the environment for this object?	E	E1
2	Student 1	S: No.	S	S5
3	Teacher	T: Why not?	(R)U	U3
4	Student 1	S: Because we are not interested in whatever is outside.	S	S3
5	Teacher	T: Right…	R	R1
		[…]		
6	Teacher	What is pulling the object downwards?	E	E1
7	Student 1	The ground	S	S1
8	Student 2	The Earth	S	S1
9	Teacher	T: What is the difference between ground and Earth?	(R)U	U3
10	Student 1	S1: The Earth in general.	S	S1
11	Teacher	T: The Earth in general and the ground in particular;	(R)U	U3
12	Student 1	S1: Yes, because the ground is in its (the object's) environment.	S	S3
13	Teacher	T: The Earth isn't in its environment?	E	E1
14	Student 1	S1: The Earth is outside.	S	S1
15	Student 2	S2: It (Earth) is natural environment.	S	S1
16	Teacher	T: Hmm. If I left it, it falls… Why? Because of the ground?	E	E1

17	Student 3	S3: The gravity.	S	S1
18	Teacher	T: What is gravity?	U	U3
19	Student 3	S3: The force towards Earth’s center.	S	S1
20	Teacher	T: Towards Earth’s center.	R	R4
21	Teacher	So, the object falls because of the gravity?	U	U2
22	Student 3	S3: Yes	S	S5
23	Teacher	T: Where does the gravity comes from? From the ground?	E	E1
24	Student 4	S4: From Earth’s center.	S	S1
25	Teacher	T: So, if we somehow remove Earth’s center there won’t be gravity?	U	U2
26	Student 4	S: From Earth in general	S	S1
27	Teacher	T: From Earth in general. Ok…	R	R1

In the above excerpts from the classroom dialogue we were able to identify five complete ESRU cycles. The discussion relates to the concepts of the object’s environment (turns 1-5) and the gravity as the reason for the object falling (turns 6-27). At the 1st part of the dialogue (turns 1 to 5), it seems that students believe that the environment for an object is what is spatially close to the object. This students’ conceptualization does not seem to be realized by the teacher. Consequently, both the students and the teacher use the same term but in a different sense, despite the teacher typically is using emergent information and one complete ESRU cycle is identified during coding the episode. When asked for this episode at the interview, the teacher said that she identified that students considered as environment of the object the natural environment and tried to deal with this issue: *“…They were talking about the natural environment that is around. I wanted them to realize that environment for the object is what relates to it, what interacts with it”*.

This miscommunication between teacher and students comes to surface later in the dialogue when students classify Earth as not a part of the object’s environment (in terms of physics) but only as an object of the natural environment (turns 12-15). At this point, formative assessment gives the teacher the opportunity to realise students’ misconception and to try through discussion to help them overcome it. That is why she introduces the concept of gravity, in order to help them understand that Earth interacts through the gravitational force with the object, despite that the Earth as a whole is not spatially close to the object (turn 23). Nevertheless, the term “environment” is never explicitly linked with interaction through forces between the objects. The teacher assumes that students have understood that, through the discussion about Earth and gravity, but it is still under question whether students made the connection she expected. Hence, the last extended cycle remains again incomplete, despite the teacher used the emergent information previously (turn 25).

During the dialogue episode above (Table 42), the teacher asks students to elaborate more and express their thinking. At the same time, crucial misconceptions like that the environment for an object is what is close to the object and the fundamental knowledge that gravity is caused from mass are not elaborated, despite the opportunities that emerged for initiating a discussion about this. Consequently, while coding shows that ESRU cycles are realized, at the same time valuable information is not utilized.

It is important to note that the underestimated by the teacher importance of students’ reasoning about gravity came into play a little bit later during the same dialogue as illustrated below (Table 43), when the teacher suggests ignoring the air to avoid discussion about its resistance when the ball is falling (turn 1).

Table 43: *Excerpt from the 2nd implementation, Lesson 3, Episode 2.*

Turn			Speaker	Transcript	ESRU category	ESRU sub-category
1			Teacher	There is air here, but things are quiet, no? In these environments we will consider that there is vacuum.	R	R3
2			Student 1	If it was vacuum, the ball wouldn't fall.	S	S4
3			Teacher	I won't answer directly your question, consider that there is air, but we ignore it.	R	R4
4			Teacher	The gravitational force exists. If it was vacuum, wouldn't he earth still be there?	U	U2
5			Student 1	Yes.	S	S5
6			Teacher	Well, where is the weight depended on?	E	E1
7			Student 1	On Earth's gravity	S	S1
8			Teacher	Ok, period. Let's go to the next…	R	R1

In the excerpt above (Table 43), the teacher identifies a student's misconception that gravity needs air as a mean to cause interaction between objects (turn 2). The teacher responds to the student's objection but doesn't really address the issue that is raised by the student. Firstly, she doesn't ask the student to justify his argument. Instead, she just refers to Earth as the cause of gravity (turn 4), but still not to mass. Since the student probably considers atmosphere as part of the Earth and maybe essential for the gravitational interaction, teacher's argumentation doesn't address the issue raised by the student. An explanation for that might be that the teacher did not appreciate the importance of this statement at that point or, she did not want to initiate this discussion becuase it extended beyond what she had planned to teach. According to her words during the interview, she already said too much about gravity:

> *“The discussion regarding gravity was beyond the lesson plan. I just wanted them to distinguish between the contact force from the ground and the gravitational with Earth. I think, I told them later that it is because of the mass, but they didn’t get it anyway”.*

Another factor that impedes the effectiveness of “on the fly” assessment is the significantly less priority given from the teacher to students’ epistemological awareness and on the development of their scientific skills. Students’ weaknesses on applying scientific methods are ignored and emphasis is given on transmitting the accepted scientific knowledge, underestimating the experimental results and their interpretation. At the next excerpt (Table 44) there is a discussion about the results found in their experiment for the acceleration of an object that is let free to fell from various heights.

Table 44: *Excerpt from the 2nd implementation, Lesson 4, Episode 1.*

Turn	Speaker	Transcript	ESRU code	ESRU Sub-categorie
1	Teacher	What is your conclusion?	E	E1
2	Student 1	When the height increases, the acceleration drops.	S	S1
3	Teacher	By how much does it drop? This 8.4 could be 8.5 so if we just ignore the decimal part of the number this would round up to 9. This is also the case for these two measurements (points to other 2 measurements).	R	R3
4	Student 2	So, they are the same…	S	S1

Looking at the whole episode at a macroscopic level, it seems that the teacher meets the dilemma whether to deal with students’ difficulties with certain aspects of experimentation (systematic/random error and possible sources of error in a given experiment) or, to keep the focus on the conceptual meanings directing the experimental results to what is theoretically expected. The teacher recognises that students are confused by some variation at the calculation of acceleration using measurements of objects falling from different heights (turn 2). Yet, in this instance, the teacher avoids discussing experimental errors; how they might emerge and how they could influence measurements, as well as whether the variation has a specific trend or not. Rather, she skips the potentially useful discussion on this and, instead, seeks to illustrate for students how the observed variation in the measurements of acceleration is sufficiently small to be dismissed as insignificant. In particular, she suggests that putting aside the decimal parts of the numbers, the measurements seem to round up to the same whole number, which implies that height does not influence acceleration (turn 3).

In the above excerpt from the classroom dialogue, the teacher avoids discussing about experimental error and how is it possible to estimate the magnitude of this. Instead, she manipulates the measurements in order to say that acceleration is stable. Identifying experimental error and distinguish from causal variation was one of the goals of the lesson, as these were defined previously between the teacher and the researcher. However, during the interview, the teacher admitted that she guided the students because of lack of time:

> *"If I had more time I would let them discuss more about the results. But the range of the experimental error was beyond the goals of the lesson. However, it seems that putting aside the decimal parts of the numbers didn't convince them".*

The next excerpt (Table 45) is a part of the classroom discussion regarding the results from a set of experimental trials where objects of different mass were released from the same height. At some point in this discussion a student had noticed that they were essentially extending the scope of the validity of the causal relation they had detected between mass and acceleration, to a range of values other than those they had explored. He then expressed his reservations about the validity of this extrapolation from the data.

Table 45: *Excerpt from the 2nd implementation, Lesson 4, Episode 2.*

Turn	Speaker	Transcript	ESRU code	ESRU sub-categories
1	S1	Mass has an influence because they put 7gr and we put 13gr.	S	S4
2	S2	But mass doesn't have an influence.	S	S4
3	S1	It does influence.	S	S4
4	T:	So, it does…	R	R4
5	T	But here we found the same acceleration whether we put 4gr, 7gr or 14gr.	U	U2

6	S1	We added only 4gr, how would that make any difference?	S	S4
7	T	What are you saying? Does mass influence?	U	U3
8	S2	If we had added 1 Kg it would still be the same, yes?	S	S4
9	T	If we had added 1 Kg would it still be the same?	R	R4
10	T	Why does it have to be the same?	E	E1
11	S3	Because the acceleration of gravity is the same.	S	S1
12	T	That is what we wanted to prove. We know it, but we wanted to prove it.	R	R1

Looking at the episode as a whole, we realize that, the teacher is applying formative assessment in an effort to understand what has confused the student and she uses the emergent information trying to address the issue raised by the student (turn 5). The student's objection is reasonable; certain pairs of variables might indeed have a different relation at different ranges of values (e.g. the force that extends a spring with the extension of the spring). However, the use of the emergent information by the teacher might not be the optimal since, she seems reluctant to discuss it sufficiently, imposing what is correct according to canonical physics knowledge, without further discussing this issue.

Studying all the missed opportunities of utilizing important information we were able to identify in the dialogues we analyzed, either because it went unnoticed during the discussion or was used in a non-optimal manner, we could categorize them to three different types (Table 46).

Table 46: *Typology of missed opportunities*

	Type of missed opportunity	Example of the of missed opportunity
1	The teacher does not respond on an issue that comes up during discussion, considering it as not relevant to the particular lesson.	The issue of action and reaction comes up but the teacher chooses not to discuss it for the moment, despite the fact that it seems that there are misconceptions which influence students' understanding on the subject under discussion.
2	Teacher has an assumption that might differ from students' understanding on a	A) Teacher assumes that students understand Earth as a massive object that causes measurable gravitational force.

	particular concept. (Miscommunication between teacher and student)	There are indications in the lesson that this is not shared by students though. B) Teacher doesn’t realize that students give a spatial meaning to the term “environment of an object”.
3	Underestimation of students’ weaknesses on scientific methods	There was an opportunity for discussing random vs systematical error that was not taken.

The only kind of missed opportunities that were identified in the 1st implementation (Figure 34) related with cases where the teacher chooses not to explore further students’ weaknesses on certain aspects of scientific processes or on epistemological issues (MO3).

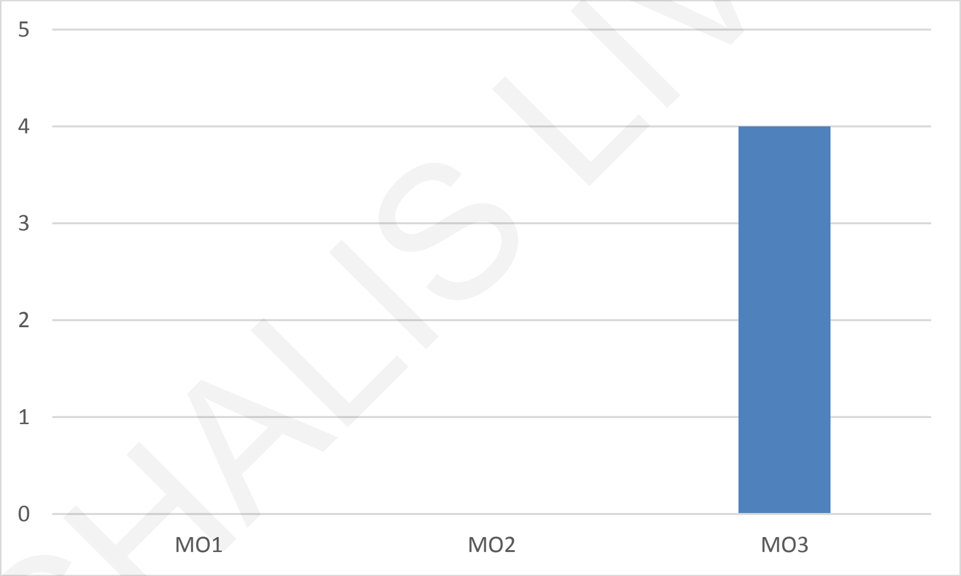

Figure 34: Frequency of missed opportunities identified in the 1st implementation

Most of the missed opportunities that were identified in the 2nd implementation (Figure 35) related with cases where the teacher and the students attached a concept they are using with a different meaning (MO2, 59%). In fewer cases the teacher chose not to explore further students’ weaknesses on certain aspects of scientific processes or on epistemological issues (MO3, 24%) or not to response on a (conceptual) issue that comes up during discussion, considering it as not part of the particular lesson (MO1, 17%).

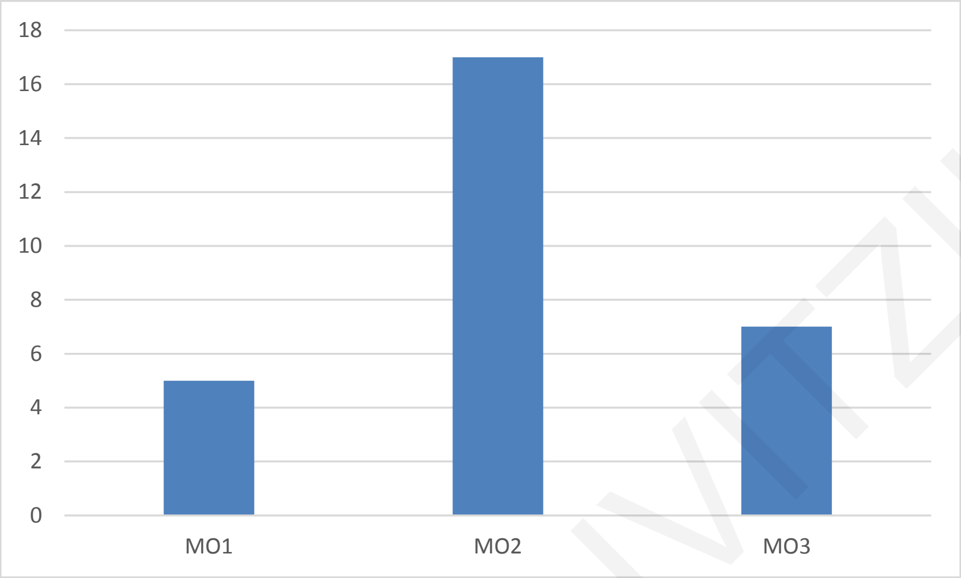

Figure 35: Frequency of missed opportunities identified in the 2nd implementation

At the 3rd implementation, there were found very few missed opportunities; two cases where the teacher and the students attached a concept they are using with a different meaning (MO2) and two cases the teacher chose not to explore further students' weaknesses on certain aspects of scientific processes or on epistemological issues (MO3).

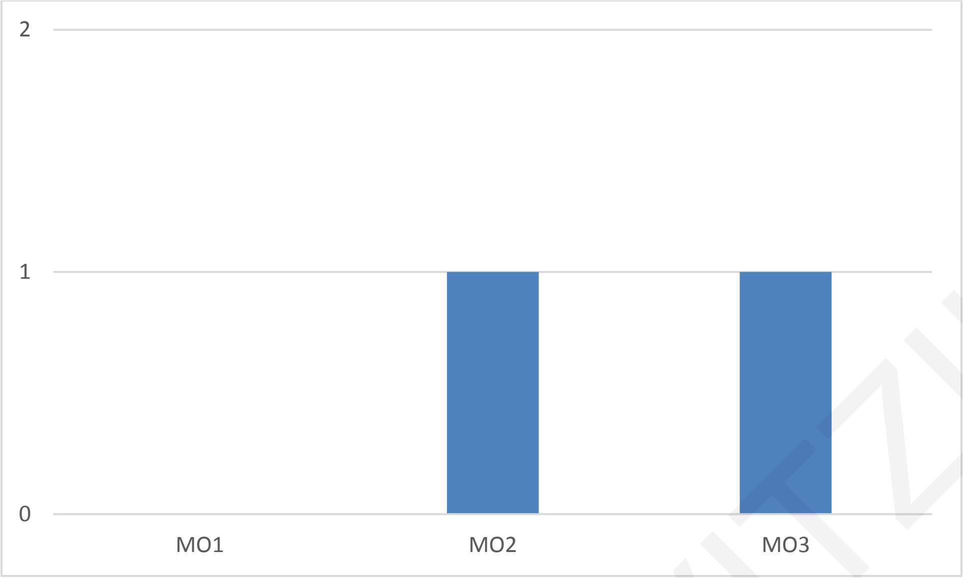

Figure 36: Frequency of missed opportunities identified in the 3rd implementation

We believe that further study of this typology of missed opportunities is likely to help us understand what tends to impede teachers' attempt to employ "interactions on the fly" as a formative assessment method as they are related with the challenges that teachers meet when they apply the formative assessment method of on the fly interactions.

4.3. Research Question 3: Conceptual completeness, depth and coherence of the classroom dialogue.

In an effort to represent the conceptual coherence and depth of the classroom dialogue, we have developed conceptual maps for each episode of the classroom dialogues. The main objectives of these concept maps are to depict:

- The variety of concepts that emerged in discussion.
- How much are these interpreted,
- How well are linked with each other.
- The necessary for understanding the phenomenon concepts that miss from the discussion or were not interpreted at all.

For this purpose, 47 concept maps were developed; 9 for the 1st implementation, 22 for the second and 16 for the third. Following we are going to present representative examples of these concept maps in order to illustrate their functionality.

The following concept map (Figure 37) is one that depicts an episode from the 2nd implementation. The theme under discussion is "object's environment" in terms of force interactions. The main concept (object's environment) is introduced by the teacher and that is why it is coloured with light green. Students link this concept with others, like the atmosphere, the air in the room, the surround in general, the ground, the Earth, the hand that holds it and has contact with it. These concepts are colored with light blue because are brought in the discussion by the students. The teacher also mentions that "object's environment" in Physics is different from what we usually call "natural environment" and includes the sun, the plants etc. However, during the episode there is no explicit link between "object's environment" and the interaction through forces with other objects, which in our opinion should have been made as a key concept in the discussion. That is why the term "environment through" interactions remains unlinked and red colored in the concept map.

The direction of the arrows depicts the order the concepts appeared in the dialogue, while the number on the arrow declares the number of ESRU cycles that this link was made for (if it was made more than once). This number also denotes the depth in which the particular concepts were discussed.

In addition, we can observe that some concepts are connected in some kind of loop; for example the terms "object's environment", "ground" and "Earth" in the concept map in Figure 37. These loops can be an indicator of the coherence of the dialogue since concepts are not just connected one by one but in larger groups, forming a more coherent framework related to the phenomenon under discussion.

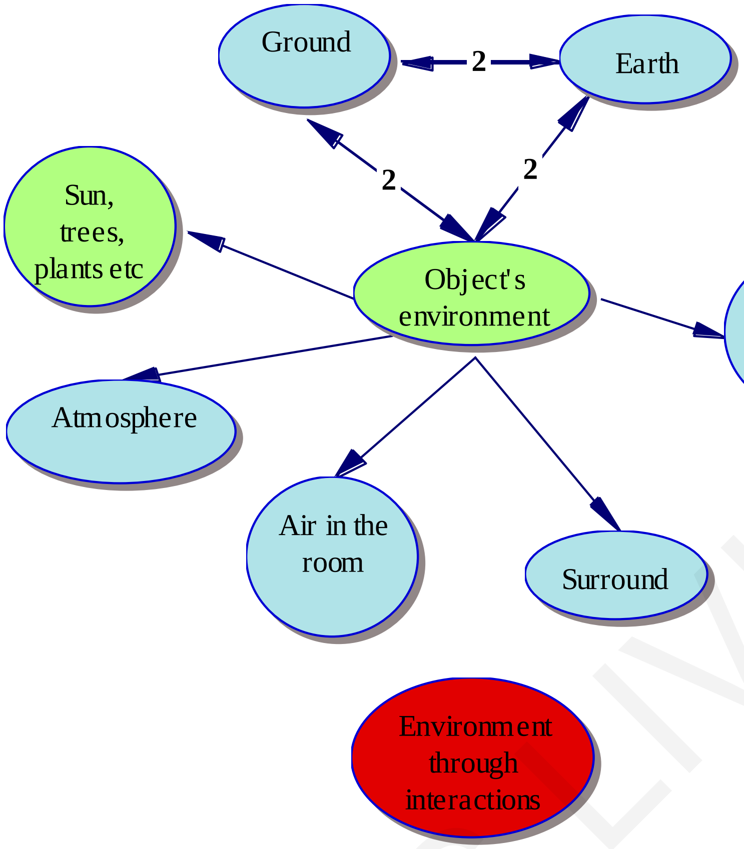

Figure 37: Conceptual map of an episode from the 2nd implementation.

The next concept map (Figure 38) is also illustrating an episode from the 2nd implementation. The theme of this episode is the discrimination between contact force and force from distance. As we can see, most of the concepts are provided from the teacher. However, weight is thoroughly discussed as an example of field force and it is linked with Earth and mass. This is depicted on the map by the loop between the concepts "force from distance, Earth, weight and field, and the multiple visits of some of the links between these concepts in a number of ESRU cycles. Similarly, the term contact force is also mentioned (by students) and well explained. Nevertheless, despite that the concept of reaction force came up in the particular episode twice by students, the teacher avoided to discuss it. Therefore, this term is depicted crossed on the concept map.

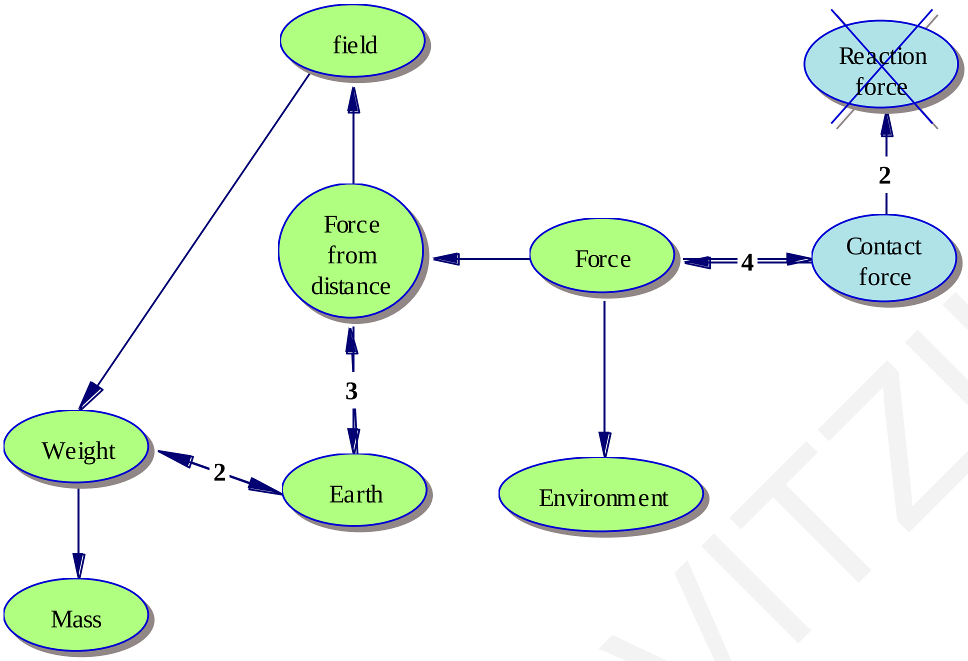

Figure 38: Conceptual map of another episode from the 2nd implementation.

The following concept map (Figure 39) is also illustrating an episode from the 2nd implementation. The teacher is trying to assess students' understanding on the results of an experiment they did, regarding the factors that influence free fall. As we observe, it resembles a quite rich discussion, with students introducing half of the concepts that appeared, 3 loops between some of the concepts and multiple revisits to the links among some of them. However, when students suggested two factors that seemed to them that influence acceleration (because they noticed small differences in measurements for different weight and height, the teacher suggested to round measurements to fewer decimal digits, without discussing the significance of any of the digits. Thus, this might seemed to students mostly like manipulating the numbers the way it suits the teacher. These instances are depicted on the map through the unlinked red boxes. We chose to draw concepts that relate to scientific processes in rectangular shape instead of the oval we use for the concepts that relate with content knowledge in order to discriminate them.

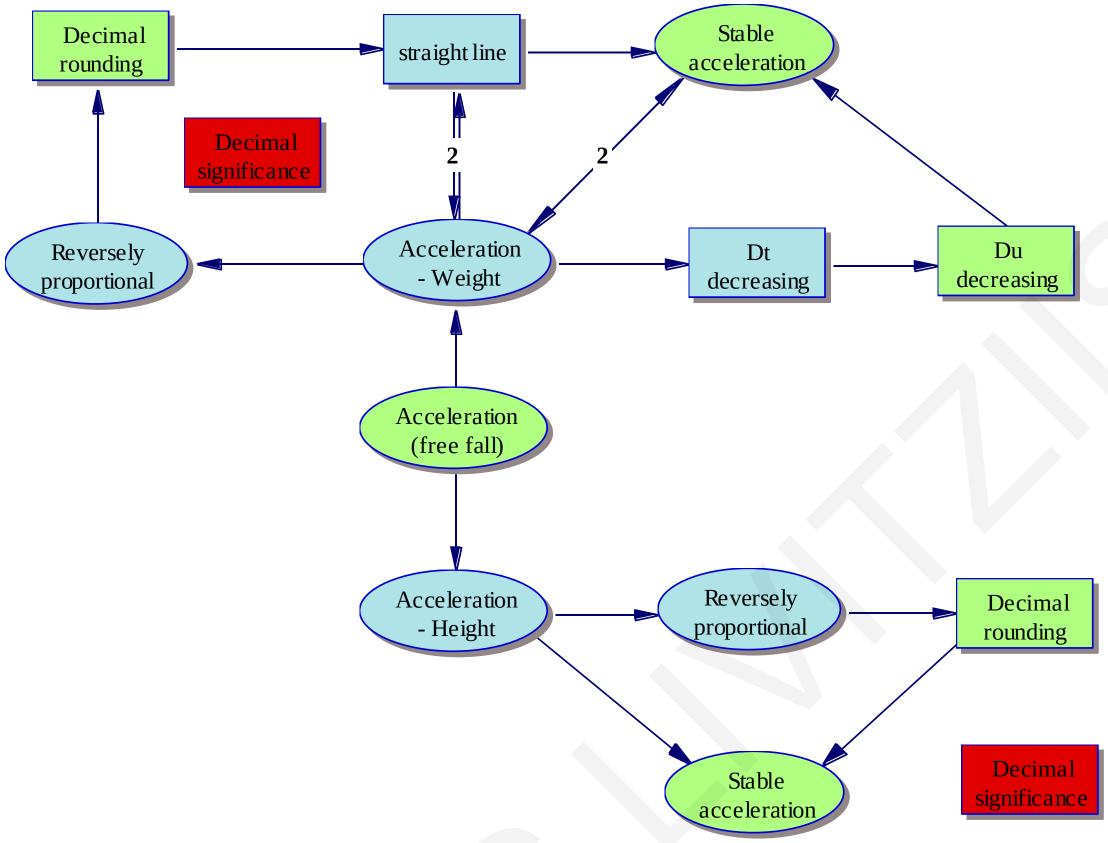

Figure 39: Conceptual map of an episode from the 3rd implementation.

It can be noticed that the previous 3 examples illustrate how different kinds of missed opportunities as they were categorized previously (Table 46) of using the emergent information can be depicted in these concept maps. At the first and the third case (Figures 37 & 39), concepts that could bridge obvious gaps in students' understanding either related with conceptual knowledge or the scientific method, are depicted as red unlinked boxes, elliptical or rectangular respectively. In addition, the kind of missed opportunity that takes place when the teacher avoids or postpones discussing a concept that emerges in dialogue and seems to influence students' understanding, is depicted by crossing the particular concept on the concept map.

As mentioned before, the number on the arrow that shows the links between concepts declares the number of ESRU cycles that this particular link was discussed. Thus, these numbers are indicative for the depth of the discussion, in other words, the emphasis that was given regarding the particular concepts and the relations between them. The following concept map (Figure 40) from an episode from the 1st implementation is an example where most of the concepts and the links among them were revisited and probably explained well enough. On the other hand, the next concept map (Figure 41), from an episode of the 3rd implementation, is an example where most of the links between concepts were made only once.

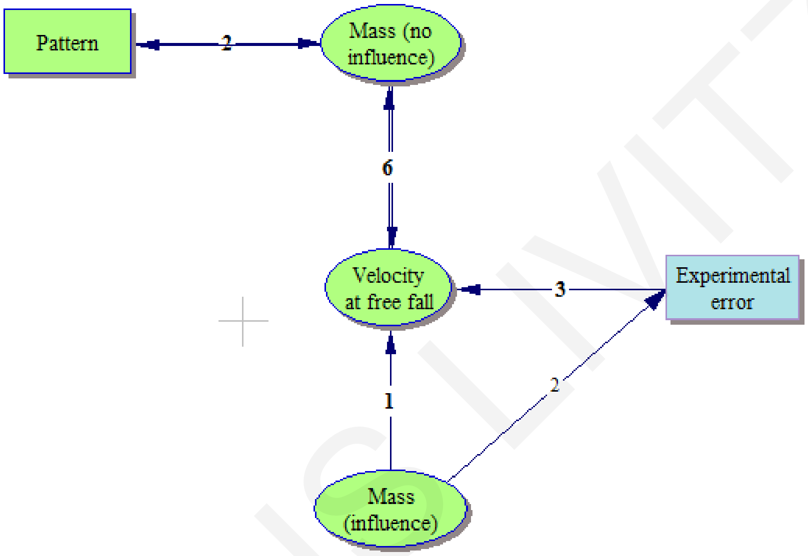

Figure 40: Conceptual map of an episode from the 3rd implementation.

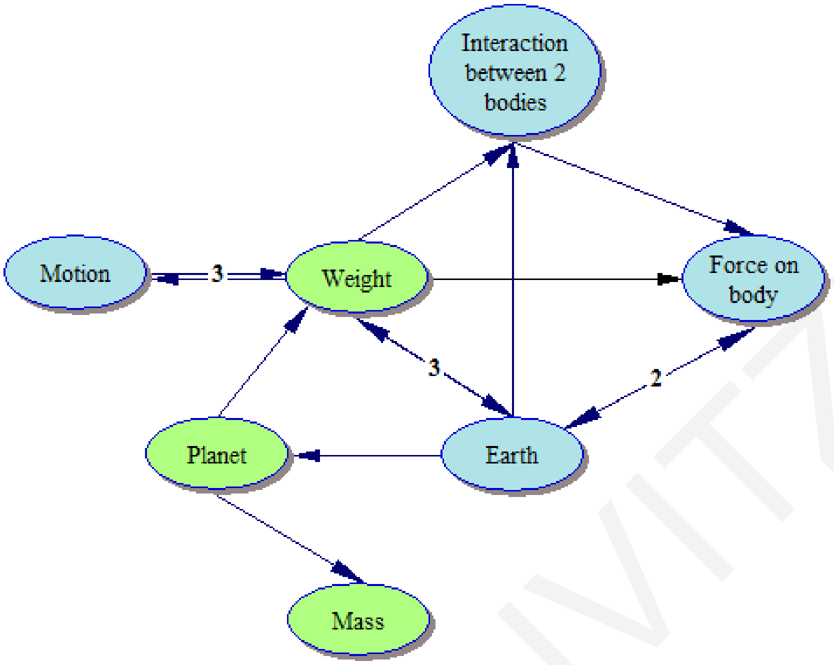

Figure 41: Conceptual map of an episode from the 3rd implementation.

Nevertheless, the dialogue in this episode (Figure 41) can be characterized as highly coherent, since most of the concepts discussed were connected to each other as the loops in the concept reveal. This suggest that the discussion was not linear, going from one concept to another, but each concept was discussed in relation to the others that affect the phenomenon under discussion.

4.4. Research Question 4: How do identified patterns of *on the fly interactions* relate to the conceptual coherence and depth of the classroom dialogue?

In order to be able to respond to the 4th research question, regarding whether the patterns identified during on the fly interactions using ESRU coding relate to the conceptual coherence and depth of the classroom dialogue, we quantified the conceptual completeness, the depth and the coherence as they are depicted on the concept maps. Specifically, we did so in the following three dimensions:

- The rate of the essential concepts that were actually used in each episode of the dialogue (completeness).

- The rate of connections between concepts that were revisited in at least one more ESRU cycle than the one they appeared (depth).
- The rate of concepts in each episode that were connected with each other forming loops on the conceptual maps (coherence).

The next table (Table 47) presents the rate of the three dimensions as well as the total value that we consider as measurement of the conceptual coherence and depth of the classroom dialogue.

Table 47: *Rate of the three dimensions and the total value of conceptual completeness, depth and coherence of the classroom dialogue.*

Episode	Concepts discussed	Depth of discussion	Coherence of discussion	Total (conceptual completeness, depth and coherence)
1.1.1	1,00	1,00	0,60	2,60
1.1.2	1,00	0,71	0,57	2,29
1.2.1	1,00	1,00	1,00	3,00
1.2.2	0,67	0,50	0,75	1,92
1.2.3	1,00	0,83	0,57	2,40
1.3.1	1,00	0,80	0,80	2,60
1.3.2	1,00	0,75	0,75	2,50
1.3.3	1,00	0,40	0,80	2,20
1.3.4	1,00	0,25	1,00	2,25
2.1.1	1,00	0,00	0,00	1,00
2.1.2	1,00	0,25	0,71	1,96
2.1.3	0,90	0,33	0,30	1,53
2.1.4	1,00	0,14	0,43	1,57
2.1.5	1,00	0,11	0,33	1,44
2.1.6	0,89	0,33	0,44	1,67
2.2.1	0,92	0,23	0,67	1,81
2.2.2	0,83	0,33	0,67	1,83

2.3.1	1,00	0,50	0,50	2,00
2.3.2	0,88	0,00	0,00	0,88
2.4.1	0,86	0,13	0,86	1,85
2.4.2	0,89	0,30	0,56	1,74
2.4.3	1,00	0,00	0,00	1,00
2.4.4	0,67	0,67	0,00	1,33
2.5.1	0,75	0,00	0,38	1,13
2.5.2,4	1,00	0,67	0,50	2,17
2.5.3	1,00	0,00	0,00	1,00
2.5.5	1,00	0,50	1,00	2,50
2.5.6	1,00	0,14	0,67	1,81
2.6.1	1,00	0,17	0,67	1,83
2.6.2	1,00	0,43	0,67	2,10
2.6.3	0,80	0,50	0,80	2,10
3.1.1	1,00	0,33	0,71	2,05
3.1.2	1,00	0,14	0,71	1,86
3.1.3	1,00	0,50	1,00	2,50
3.1.5	1,00	0,33	1,00	2,33
3.2.1	1,00	0,42	0,83	2,25
3.2.2	1,00	0,20	1,00	2,20
3.2.3	1,00	0,00	0,00	1,00
3.2.4	1,00	0,33	0,50	1,83
3.3.1	0,86	0,00	0,14	1,00
3.3.2	1,00	0,10	0,86	1,96
3.4.1	1,00	1,00	1,00	3,00
3.4.2	1,00	0,33	0,67	2,00
3.4.3	1,00	1,00	1,00	3,00
3.4.4	1,00	0,29	1,00	2,29
3.5.1	1,00	0,50	1,00	2,50
3.6.1	1,00	0,50	1,00	2,50

Next, this value for the conceptual coherence and depth of the classroom dialogue was examined if it is related with the rate of completion of ESRU cycles of each episode using Pearson correlation. A moderate but significant ($r=0.44$, $p<0.01$) correlation has been found when examining all episodes from all implementations together. However, looking at the correlation of the rate of completion of ESRU cycles of each episode with each one of the three dimensions of conceptual coherence and depth of the classroom dialogue, we realize that it is only significant for the two of them. The correlation it is moderate for the coherence of the dialogue ($r=0.45$, $p<0.01$) and weak for the depth of discussion ($r=0.32$, $p<0.05$), while it is insignificant ($p=0.41$) for the rate of the concepts discussed.

These results indicate that the rate of completion of ESRU cycles of each episode can be an indicator for the conceptual coherence and depth of the classroom dialogue but, still support our state presented previously that there might be missed opportunities to improve students' understanding, particularly because specific critical concepts were not brought and discussed during the dialogue.

For the previous analysis, we have used the assumption that each one of the three dimensions of conceptual coherence and depth of the classroom dialogue has the same weight for calculating a single value for the conceptual coherence and depth of the classroom dialogue. However, this might not be valid. We argue that, the rate of the concepts discussed against the number of concepts that was essential to be discussed in order that students have a complete understanding of the phenomenon is more important than the other two dimensions. This is based on the argument that there is little meaning to discuss in depth particular concepts of a phenomenon when the same time other critical concepts are missing from the discussion. In addition, we consider the depth of the dialogue to be more important than the coherence of concepts. The argument is that firstly there is a need to establish a good and meaningful connection between each concept and the main focus of the discussion before illustrating other connections between concepts. Based on these assumptions we developed an eight level rubric for categorizing each episode according the level of conceptual completeness, coherence and depth of the dialogue that occurred during the episode (Table 48).

Table 48: *Levels of conceptual coherence and depth of the classroom dialogue.*

	Concepts completeness	Elaboration of concepts	Coherence of concepts
1	All essential concepts	High (more than half were revisited)	High (more than half in loop)
2	All essential concepts	High (more than half were revisited)	Low (half or less in loop)
3	All essential concepts	Low (half or less were revisited)	High (more than half in loop)
4	All essential concepts	Low (half or less were revisited)	Low (half or less in loop)
5	Essential concepts missing	High (more than half were revisited)	High (more than half in loop)
6	Essential concepts missing	High (more than half were revisited)	Low (half or less in loop)
7	Essential concepts missing	Low (half or less were revisited)	High (more than half in loop)
8	Essential concepts missing	Low (half or less were revisited)	Low (half or less in loop)

After placing each episode at the appropriate level of conceptual coherence and depth of the classroom dialogue, we examined whether this level is related with the rate of completion of ESRU cycles of each episode using Spearman correlation. A medium but significant ($r=0.34$, $p<0.05$) correlation has been found when examining all episodes from all implementations together.

5. DISCUSSION

5.1. Patterns of interaction

Regarding the first research question and the patterns we can identify in “on the fly interactions” between the teacher and the students, we used the ESRU categories and their subcategories to analyse the dialogues on a microscopic level. This allowed us to identify the various ways each teacher used to elicit student’s ideas and understanding, to recognize their responses and to use them in a productive way to promote their learning. We also were able to identify the various ways the students responded to these efforts of the teacher. The frequencies in which each ESRU subcategory appeared in the dialogue episodes in each implementation are indicative of the manners that each teacher uses during the dialogue in class and how the students respond to them. Ruiz-Primo and Furtak (2007) reported also a relation between complete ESRU cycles and student learning.

Broken cycles either in ESR format or ES are sometimes dominant structure in a class, i.e. the U component is rarely used by the teacher. However, this does not mean that incomplete cycle, such as ESR, would be an inevitably poor move in terms of formative assessment. Occasionally, it can be the teacher's pedagogical choice to stand aside without giving feedback or probing immediately but doing so after a while. Looking more carefully at the structure of the dialogue, we were able to identify what we called "extended cycles" where the cycle does not follow the regular form of ESRU but Use is coming after some repetition of S and R subcategories (i.e. ESRSRU or ESSRU). In other cases, codes might be missing from a cycle, for example E (Elicit) because elicitation took place through the previous cycles in form of Use. These differentiations indicate the complexities that a classroom dialogue includes.

The application of the ESRU coding system enabled us to differentiate teacher's action between Eliciting and Using. Eliciting is associated with creating opportunities to collect evidence of learning and Use is associated to creating opportunities to make use of that evidence in order to promote learning. This distinction is very important when characterizing teachers' formative assessment practice, because effective formative practice requires collecting evidence of learning and acting on it leading to responsive teaching (Harrison, 2015).

Regarding the ways the teachers tried to elicit students' ideas; we observe that, at the 1st implementation there were fewer cases where the teacher poses a closed question to elicit students' ideas about a new concept/idea/relation and more cases where the teacher asks students to offer an example or report data, comparing with the 2nd and the 3rd implementation (Figure 19). This indicates an effort to engage them more actively in the dialogue by asking them to offer examples or report data from their experimental work while teachers at the 2nd and 3rd implementations seemed to ask more often close-ended questions.

The quality of classroom questioning is a matter for concern, as expressed in the work of Stiggins et al. (1989) who studied 36 teachers over a range of subjects and over grades 2 to 12, by observation of classroom work, study of their documentation, and interviews. At all levels the questioning was dominated by recall questions, and whilst those trained to teach higher-order thinking skills asked more relevant questions, their use of higher-order questions was still infrequent. An example of the overall result was that in science classrooms, 65% of the questions were for recall, with only 17% on inferential and deductive reasoning. A review of work of this type (Pressley et al., 1992) establishes that requiring learners to compose answers with explanations to explore their prior knowledge of new work does improve learning, and that this may be because it helps the learner to relate the new to the old and to avoid superficial judgments about the new content. Despite the fact that it seems likely that most teachers are aware of the benefits of richer questioning styles, they believe that such approaches are difficult to implement in 'real classrooms' (Dassa, 1990). This might justify the low rate of open-ended questions in the cases of 2nd and 3rd implementation in our study.

On the other hand, students from the 1st implementation offered justification for their reasoning and provided examples or reported data much more often from those of the 2nd and 3rd who were mostly limiting their answers on suggesting a concept/relation in response to questions posed by the teacher (Figure 20). This led us at the hypothesis that the quality and richness of students' contributions might relate to the kind of question posed by the teacher. In her discussion of classroom question and answer sequences, Kennedy (2005) states that "teachers devise standard ways of posing questions and students learn that there are standard ways of responding to these questions" (p.95). Gutierrez calls these standard ways of posing questions and responding "instructional scripts" (1994). Scripts are "normative patterns of life within a classroom" that students use for "interpreting the activity of others and for guiding their own participation." They are negotiated after "repeated interactions … with particular social and language patterns constructed both locally and over time" (p.340).

The cluster analysis we ran among the subcategories of ESRU verified the aforementioned hypothesis since in all three implementations subcategories E2 (teacher asks students to offer an example or report data) and S3 (students provide an example or report data) combine at short distance (Figures 14, 16, 18). At the same time, S4 (student explicates an inference/poses a question about an aspect of the topic under discussion) also combine in the aforementioned cluster as well as S2 (student offers justification for his/her reasoning) does so. These are also verified by the cluster analysis we ran using all the data of the three implementations together (Figure 28), supporting our hypothesis that when the teacher prompts more active engagement of students in the discussion, they usually respond accordingly, providing useful information about their understanding that could be used by the teacher to promote their learning. This may also justify the high rate of completion of ESRU cycles at the 1st implementation where the teacher has a relatively high percentage (25%) eliciting students' ideas through asking them to offer an example or report data (E2). We suggest that this gave the opportunity to rich information about students' understanding to be revealed and the chance to the teacher to use it. These are instances that the teacher is being proactive in deliberately seeking formative assessment information from students (Bell & Cowie, 2010) in an effective way.

According to the literature, the kind of questions the teacher asks influences the classroom norms, participants' beliefs, classroom epistemology, and opportunities available for students to engage with curriculum content. Duschl (2008) claims that the scientific knowledge we hold needs to be put into practice and tested. It is important how and when the important dialectical discourses about data representations, data and conceptual models, evidence, explanatory theories, and methods are incorporated into science learning environments. For science learning, the conversations should mediate the transitions from evidence to explanations, or vice versa, and thereby unfold discovery and inquiry. The use of only one type of question, such as asking students to provide their observations, is not as useful as a combination of other types of questions; for example, asking students to evaluate the quality of those observations (Ruiz-Primo & Furtak, 2006b). Evidence-based reasoning is a cornerstone of effective formative assessment practice in the context of scientific inquiry (National Research Council, 2001b; Duschl, 2003). Research shows that certain discursive moves such as high-level questioning, encouraging argumentation, and focusing on student thinking create opportunities for conceptual understanding and, in some instances, are positively related to student achievement (Empson, 2003; Nystrand et al., 1997; O'Connor & Michaels, 1993; Wood et al., 2006; Yackel, 2002).

However, both in questioning and written work, teachers' assessment focuses on low-level aims, mainly recall (Black and William, 1998). There is little focus on such outcomes as speculation and critical reflection (Senk et al, 1997), and students focus on getting through the tasks and resist attempts to engage in risky cognitive activities (Duschl & Gitomer, 1997). In a classroom where the teacher's questioning has always been restricted to 'lower-order' skills, such as the production of correct procedures, students may well see questions about 'understanding' or 'application' as unfair, illegitimate or even meaningless (Schoenfeld, 1985).

This is depicted in our analysis where E1 (Teacher poses a question to elicit students` ideas about a new (although interrelated) concept/idea/relation), S1 (Student suggests a concept/relation in response to question posed by the teacher) and R1 (Provision of affirmation) remain in separate clusters for a long distance (Figures 14, 16, 18 & 28) and in particular, they don't seem to lead to complete ESRU cycles. These categories appeared more often than the others at the case of the 2nd implementation where the rate of completion of ESRU cycles remained very low. It can be suggested that asking mainly closed-ended questions about specific concepts and then, when a student responses correctly, only affirming without prompting for further explanation, allows less information to emerge and be used during the dialogue.

Webb and colleagues (2006) developed a coding scheme to describe the cognitive demand present in teacher-student (and small group) interactions. In their study, approximately 80% of teachers' responses and questions required low or medium level cognitive processes (which they define as looking up or recalling information, performing calculations or problem steps, and confirmation of student responses without elaboration). A question-asking style of recitation will reveal only a limited part of the students' understanding. If one of the broad goals of recitation is to provide the teacher with an assessment of the students' knowledge of the subject matter so that the next learning activity may be planned, then a question-asking style of recitation may actually frustrate the primary goal of recitation itself (Dillon, 1983). The nature and quality of the questions teachers pose matter for the nature and quality of the student thinking they reveal and promote (Black & Wiliam, 1998a). Substantial research supports the value of open-ended questions, such as in the excerpt from Black and Wiliam (above, that require and afford more than single word responses). Such questions elicit more information from students (Nystrand et al, 2003), which provides teachers with more data and sparks deeper student thinking.

Ruiz-Primo & Furtak (2006) also noticed that very few assessment conversations involved formulation of explanations, evaluation of quality of the evidence (to support explanations), or comparing or contrasting others' ideas, explanations. It is possible that by omitting these important steps in favor of focusing on predictions and observations may provide students with an incomplete experience in inquiry learning. This reflects that many teachers pay attention to the procedural aspects of the epistemic frameworks more than to the development of the criteria to make judgments about the products of inquiry (e.g., explanations). Their findings regarding broken cycles (ESRs) are consistent with those of the IRE/F studies, in that teachers conduct generally one-sided discussions in which students provide short answers that are then evaluated or provided with generic feedback by the teacher (Lemke, 1990).

The ways teachers recognized students' contributions in the dialogue seems to play an important role on the amount of information that is allowed to emerge during the dialogue and is used to promote learning. The teacher of the 2nd implementation, where the rate of completion of ESRU cycles remained very low, was more willing to readily offer the right answer to a question posed by herself or by a student, while the teachers of the 1st and 3rd implementation acknowledged students' contribution in a neutral way more often. This is verified again at the dendrograms of all implementations and especially at the combined one (Figures 14, 16, 18 & 28) where we see R4 (neutral recognition of students' responces) combining with the variable Comp that stands for complete ESRU cycle. R4 was relatively high at 3rd implementation too (33%) but quite low at the 2nd one (22%).

This suggests that regognizing students' ideas in a more neutral way allows more time and space for them to think and express in a better way their thinking. The information can be used to build up a picture of the student learning by the teacher (Bell & Cowie, 2000) giving the opportunity to the teacher to use the emerging information more often and provide more productive feedback.

Ruiz-Primo & Furtak state that when teachers practice revoicing, ''they work students' answers into the fabric of an unfolding exchange, and as these answers modify the topic or affect the course of discussion in some way, these teachers certify these contributions and modifications'' (Nystrand & Gamoran, 1991, p. 272). Revoicing, then, is not only a recognition of what the student is saying, but also constitutes, in a way, an evaluation strategy because the teacher acknowledges and builds on the substance of what the student says. Furthermore, it has been found that this type of engagement in the classroom has positive effects on achievement (Nystrand & Gamoran, 1991). Some indicators of exploratory talk are hesitations, rephrasing, false starts, expressions of tentativeness and fairly low level of explicitness (Cazden, 2001).

On the other hand, the relatively high percentage of R2 (teacher readily offers the right answer to a question posed by him/herself or by a student) along with R3 (teacher provides disconfirmation) at the 2nd implementation (17% and 11% respectively), might had discouraged students from freely expressing their alternative ideas, negatively influencing the available information for the teacher to use, thus the rate of completion of ESRU cycles was also quite low. When following-up to evaluate, the evaluative move can signal the end of a discussion (Pierson, 2008); there is often no space for a productive exchange to develop (especially when evaluation emanates from an authoritative source). In a classroom where the teacher routinely withholds evaluative follow-up comments, the discourse moves focus on answers only and gives students responsibility for determining the correctness of responses, thereby placing a larger cognitive burden on them (Cazden, 2001; Hammer, 1995; O'Connor & Michaels, 1996; vanZee & Minstrell, 1997). Bell and Cowie (2001) also state that, in discussing the extent to which they disclosed their ideas to the teacher, the students commented on the nature of the assessment strategies used by the teachers, the relationship between teachers' rights and disclosure, disclosure as a source of potential harm, and trust as mediating the disclosure. An important finding of that research was that the validity of formative assessment relied on the extent of student disclosure (Cowie, 2000). Students' perceptions that their ideas were subject to evaluative judgement by teachers perhaps explains their sensitivity to whether a teacher's assessment purpose was to 'check on' or 'find out' about their learning (Cowie, 2005). In all, 40 students indicated they deliberately refrained from asking questions, thereby disclosing their ideas because of concerns about potential harm. They reported teachers did not always respond to the content of their questions or to them, when they were seeking help to understand ideas.

For completing the ESRU cycle, a teacher can ask another question that challenges or redirects the students' thinking, promotes the exploration and contrast of students' ideas, or makes connections between new ideas and familiar ones. Teachers also can provide students with specific information on actions they may take to reach learning goals. In this manner, viewing whole-class discussions in an inquiry context through the model of ESRU cycles allows the reflective and adaptive nature of inquiry teaching to be highlighted (Ruiz-Primo & Furtak, 2006).

In our research, the teachers used the emergent information in various ways but more often were the cases where they sought to focus students' attention on something with the intent to facilitate the discussion (U2), tried to engage students in deeper reasoning (U3) and articulated consensus from a series of contributions that were exchanged (U4). Nevertheless, there were also instances where the teacher suggested an activity that could help students resolve a specific issue (U1) or asked the opinion of other students about an expressed idea from their classmates (U5). Similar actions by the teachers were identified in Bell's and Cowie's work (2000), for example, suggesting further questions, suggesting further activities, questioning of a student's ideas, explaining the science, giving feedback as to the students' scientifically acceptable or unacceptable ideas.

Maybe the most interesting subcategory of Using is U1, because through their participation and reactions to instruction, students influence classroom experiences so that teachers must adapt their instruction accordingly (Sawyer, 2004). Gains in students' concepts and problem-solving performance appeared to be directly related to changes in teachers' instruction" – changes aligned with principles of formative assessment (Fennema et al., p.430). Movement toward instructional practices that allowed students to solve a wide range of problems, elicited their thinking, and adapted instruction based on students' thinking was related to increases in student learning in that year or the following year (up to half a standard deviation increase in class's mean achievement scores) (Parsons et al., 2010). Results describing the thoughtfulness of teachers' adaptations and rationales were quite surprising, with a majority identified as minimally thoughtful or reflective of fragmented pedagogy. However, two subsequent studies (Allen et al., 2013; Parsons, 2012) following teachers purposefully selected as reflective professionals found more substantial adaptations almost entirely in response to students' behavior which is in line with this research where, the teachers adapt their instruction by introducing an unplanned activity to meet students' need, in very few occasions.

Subcategory U2 relates with cases where the teacher uses the information gathered from students' contributions with prompting questions which either help them make sense of the data they had presented or elaborate their ideas in a way that will help them to improve their understanding.

These moves serve to move thinking forward through acknowledging, shaping and extending student language and ideas, rather than imposing science ideas in a manner unconnected with student experience and thinking. As Heritage (2010) notes, scaffolding occurs when the more expert other provides support through a process of interaction. For example, a teacher asking leading or probing questions to elaborate the knowledge the learner already possesses, or providing feedback that assists the learner to take steps to move forward through the Zone of Proximal Development (Vygotsky, 1978).

From a socio-constructivist perspective on learning (Vygotsky, 1978), students learn through a negotiation on meaning making. From a formative assessment perspective, the use of open-ended questions in a genuine dialogic environment is more likely to support this meaning making process. This is because, in these contexts, teachers create opportunities for students to engage in deeper thinking, to articulate and reveal their thinking to others and to be open in considering next steps in learning. Research shows that open-ended questions tend to elicit more elaborate answers that rely on an articulation of ideas, and therefore involve deeper thinking (Chin, 2007).

Asking students questions that relate with subcategory U3 like “Why do you think so?” or “What does that mean?” can easily help teachers to explore in more detail their students’ level of understanding. The centrality of inference in formative assessment becomes quite clear when we consider the distinctions among errors, slips, misconceptions, and lack of understanding. Each of these causes implies a different instructional action, from minimal feedback (for the slip), to re-teaching (for the lack of understanding), to the significant investment required to engineer a deeper cognitive shift (for the misconception). The key point, however, is that any attribution of an underlying cause is an inference, a ‘formative hypothesis’, that can be tested through further assessment. That further assessment might, for example, involve asking for the student’s explanation as to why he or she chose to respond in a particular way (thereby making the student a partner in formative assessment); administering more tasks and looking for a pattern of responses consistent with the hypothesis; or relating the error to other examples of the student’s performance (Bennet, 2011). Encouraging students to provide more information or to consider why they are making a statement can help students to recognize the elements of good scientific explanations so that they will have a better idea of what to include on the next occasion (Ruiz-Primo & Furtak, 2006).

To support learning while it is occurring, teachers need to provide descriptive feedback in the form of ideas, strategies, and tasks the student can use to close the “gap” between his or her current learning level and the next level, a strategy that it is represented in our study by subcategory U4. In this sense, feedback becomes instructional scaffolding in the ZPD (Heritage, 2010). Teachers also often elaborate and reformulate the contributions made to classroom dialogue by students (for example in response to a teacher’s questions) as a way of clarifying what has been said for the benefit of others and also to make connections between the content of children’s utterances and the technical terminology of the curriculum (Mercer, 2004). Nystrand’s studies of eighth and ninth grade English classrooms suggest that feedback does affect student learning when the purpose of follow-up is to elaborate on student ideas. Their research shows a statistically significant positive correlation between achievement scores and high-level evaluation where high-level evaluation is defined as certification or acknowledgment of a response and the subsequent incorporation of that response into class discourse (Nystrand et al., 1997). In another example (Cowie, 2005), a student claimed she could be influenced by the coherency of an idea, by her teacher’s view as someone with authority and by empirical evidence in support of a particular view. Helpful feedback of this kind given to one student can also be a secondary source of feedback for the rest of the students.

The way forward is to ensure that feedback is provided in a timely manner (close to the act of learning production), that it focuses not just on strengths and weaknesses but also on offering corrective advice, that it directs students to higher order learning goals, and that it involves some praise alongside constructive criticism. Nieminen et al (2015) found that, when a teacher guides students to take the next step, this can be done after a quick interpretation or after further probing. In the former, the guidance is based on correspondence of the students’ and the teacher’s ideas. In the latter, guidance is based more on the students’ ideas. Thus, the former is more authoritative guidance and the latter more dialogic guidance.

Another strategy that has been considered essential in engaging students in assessment conversations is comparing and contrasting students’ responses to acknowledge and discuss alternative explanations or conceptions (Duschl, 2003). This strategy is essential in examining students’ beliefs and decision-making concerning the transformation of data to evidence, evidence to patterns, and patterns to explanations. Considering evidence through a collaborative frame allows the time and opportunity for students to compare ideas and therefore provides more evidence for students and the teacher about individual and group understanding. Such an approach enables the teacher to tailor any follow-on activities in response to the evidence that signals to individuals and to all the learners that her interest lies in moving their own and other students' ideas forward, rather than simply searching for correct answers from the class (Cowie, Harrison & Willis, 2018). However, in our study only very few times teachers asked their students to compare their ideas with their classmates as indicated by the low rate of subcategory U5 we have identified. Ruiz-Primo and Furtak also found that comparing and contrasting students' responses was not a strategy frequently used by these teachers.

Another interesting result that comes up from our analysis is that, as is apparent from the tree diagrams (figures 14 and 15), especially in the 1st and 2nd implementation, R2 (teacher readily offers the right answer to a question posed by him/herself or by a student) along with R3 (teacher provides disconfirmation) combines with S4 (student explicates an inference/poses a question about an aspect of the topic under discussion). This might reflect the need of the teachers to respond with an immediate answer to students' questions instead of using them for a productive dialogue. The notion of the teacher as a neutral facilitator is often not seen as part of formative assessment. As teachers in Bell and Cowie's research stated, being a neutral facilitator isn't how they think themselves during formative assessment, but rather they like to take action. They may choose to do nothing because they want to leave the kids for a while to see if they can find their way through it, but if they can't, a teacher needs to make another decision. Traditional models of schooling and many of the familiar, taken-for-granted practices of education position the teacher as the subject-matter expert whose job is to provide clear explanations of difficult mathematical topics which most children will probably never understand. The students' job is to remember what they have been told (Lemke, 1990). Although it is possible for tutors to infer student knowledge from student answers, it is important to acknowledge that student contributions are normally very fragmentary, semi-coherent, and distributed over many turns during the collaborative exchange. In fact, the tutors ended up supplying more information in route to an answer than did the students, even though the tutors asked the original questions (Graesser, 1995).

Nevertheless, code S4 (Student explicates an inference or poses a question about an aspect of the topic under discussion) was much more frequent at the 3rd implementation than at the others (especially than the 2nd implementation) and this can be a factor that influenced positively the rate of completion of ESRU cycles at the particular implementation. In their detailed qualitative study of the classroom characteristics of two outstandingly successful high-school science teachers, Garnett & Tobin (1989) concluded that the key to their success was the way they were able to monitor for understanding. A common feature was the diversity of class activities—with an emphasis on frequent questioning in which 60% of the questions were asked by the students. The students also construed themselves as active participants in formative self-assessment through their description of asking a question as a first response to a formative self-assessment they did not understand (Cowie, 2005). In this case, it appeared that the students acted to utilize the people around them as a 'learning resource' (Wenger, 1998). An interpration for this might be that the attitude of teacher against their responses to the dialogue encouraged them to express their questions and queries, allowing better collection and use of information and consequently the rate of completion of ESRU cycles. Trust was related to student experience of teachers and peers as well intentioned and or the potential for benefit from interaction. Students formed impressions of a teacher's likely actions and reactions overtime by monitoring whether she/he was willing to re-explain ideas and by drawing on information gained from peers and siblings. They needed to feel comfortable with a teacher before they disclosed their ideas by asking questions.

When classroom talk is accountable to the learning community, students listen to one another, not just obediently keeping quiet until it is their turn to take the floor, but attending carefully so that they can use and build on one another's ideas. Students and teachers paraphrase and expand upon one another's contributions. If speakers aren't sure they understood what someone else said, they make an effort to clarify. They disagree respectfully, challenging a claim, not the person who made it. Students move the argument forward, sometimes with the teacher's help, sometimes on their own.

The research by van Zee and Minstrell (1997) shows the positive gains in learning that come about when the authority for classroom conversation shifts from the teacher to the students. Peers more readily engage in question-answer exchanges with each other than with superiors; and student responses to fellow students are longer and more complex than responses to the teacher (Mishler, 1978). Employing a technique they call the "reflective toss," van Zee and Minstrell found that students become more active in the classroom discourse, with the positive consequence of making student thinking more visible to both the teacher and the students themselves. By commenting on the work of peers, students develop detachment of judgement (about work in relation to standards), which is transferred to the assessment of their own work (Nicol & McFarlande, 2006).

Students often suppress evidence of what they know and not ask questions for fear of embarrassment and because they do not want to appear too ‘brainy’ (Wiliam, 1992). Thus, it seems that students may evaluate the benefits and risks of disclosing their ideas as they juggle competing and often contradictory social and academic goals as part of the process of task engagement and learning (Claxton 1984). In this view, the assessment process itself is inherently both a social and a learning situation, one in which important power relationships between teacher/assessor and learner greatly influence the learners performance, and in which learners are continually reappraising their relationship to the assessment process, and are learning what is required and how to deliver or withhold it (Gipps 2002).

A limitation of the ESRU coding system and the subcategories we developed is that the number of students participating actively in discusion is not depicted. In other words each time code S is appeared it could be either the same or different student contributing in the dialogue.

5.2. Factors that afford *on the fly interactions*

Elaborating the instances where the various subcategories of Use appeared in the dialogue, we were able to respond at the first part of the second research question, regarding the factors that seem to afford productive use of the emergent information. In particular, the categorization in five categories of our interpretations for the cases that information that comes up in discussion is used productively (Table 27), reveals that there are different ways of utilizing the feedback information. These extend from summing up what is already said to posing appropriate questions that could promote students thinking or encourage them to make their thinking more transparent, suggesting activities that could meet students’ needs as diagnosed by the teacher or, inviting students to compare their ideas.

Research has shown that there is a variety of moves that teachers can use to ensure purposeful, coherent, and productive conversation. A number of moves have been identified that help students in building understanding of complex ideas, while at the same time keeping students engaged and "on the same page”. Similarly with Ruiz-Primo and Furtak (2006), the using strategy most frequently observed in this study was asking why and how questions to challenge or redirect students' thinking. This type of question allows teachers to get more information about students' understanding easily. Teachers can improve the quality of classroom dialogue by using probing questions to explore students' ideas, through encouraging students to express their ideas using their own words and by helping students to elaborate and justify their views (Mercer & Howe, 2012). Howe and Mercer (2007) have shown that positive learning outcomes arise from students engaging in dialogue where they have to justify their views, discuss and resolve differences in opinion. In such cases, both teacher and students need to be responsive to ideas that are being put forward.

In classroom talk that is accountable to generally accepted standards of reasoning, students use data, examples, analogies, and hypothetical "what-if" scenarios to make arguments and support claims. Students are encouraged to seek out different kinds of supporting evidence, strengthening an argument by using a variety of sources to support it. Students and teachers assess and challenge the soundness of each other's evidence and quality of reasoning, often posing counter-examples and extreme case comparisons to illustrate a point. Hidden assumptions are uncovered and examined. Students and teachers consistently ask one another to show why the evidence used to support a claim is accountable to rigorous thinking (Michaels et al., 2010). The action taken by the teacher in cases where teachers are beingresponsive by undertaking formative assessment in response to other information they had gathered about the students' learning is addressed by Bell and Cowie (2000) as reactive.

In Tytler and Aranda's research (2015) the teachers were selected as representing expert practice as judged by professional norms. These teachers rarely evaluated in any judgmental sense, but prompted and responded to student input in a variety of ways. The moves have been identified as falling within three broad categories of purpose - Eliciting/Acknowledging, Clarifying, and Extending - which are all well represented in these expert teachers' practice, and provide shape to the conceptual intent of lessons. The overwhelming impression in each case is one of strategically planned and executed practice in which the shaping of knowledge was shared between the students and teacher.

Ruiz-Primo and Furtak (2006b) have provided evidence that the teacher whose whole-class conversations were more consistent with the ESRU cycle had students with higher performance on embedded assessments and this difference was also reflected on the posttest. However, as we showed in the previous chapter and we will discuss further later on, there could be cases during the classroom dialogue where, despite ESRU cycles are completed, nevertheless crucial information that emerges might go unused or used in non-optimal way.

5.3. Factors that impede teachers' attempts to use *interactions on the fly*

Identifying and categorizing (Table 37) the cases where the ESRU cycles remained incomplete while there was an opportunity for the teacher to sustain the dialogue in a fruitful manner was a first step to respond at the second part of the second research question and identify possible factors that impede teachers' attempts to use "interactions on the fly".

The first factor we identified that impedes teachers' attempts to use "interactions on the fly" is the way the teacher recognizes and reacts to students' responses to her questions. Many cycles remained incomplete because the teacher ignored or rejected a student's answer because it wasn't the one she expected and simply repeated the question. While teachers can plan a specific activity to create opportunities to make visible what the students are thinking and focusing on, how the teacher responds to this evidence communicates to students what the teacher deems important. There is little discussion about the substance of student thinking (Coffey et al, 2011). The teacher assessed student contributions for their alignment with that; ideas that did not align were ''wrong'' and rejected. A teacher's response frames what counts as valued learning and determines if and how their actions advance or inhibit student learning and learning motivation (Cowie, Harrison & Willis, 2018). Especially, when teacher rejects a student's response, providing negative evaluation, this affects students' attitude against the classroom dialogue and learning in general. Not surprisingly, in a comprehensive review of feedback, Kluger and DeNisi (1996) found that one-third of the studies showed negative effects: feedback about performance actually harmed learning outcomes.

The second factor we identified that impedes teachers' attempts to use "interactions on the fly" is the hurry of the teacher to provide the 'right' answer either when getting a 'wrong' one or when responding to a student's question. This can be accounted on teacher's priority to present the established scientific knowledge in favor of investigating students' alternative ideas. Day and Cordon's (1993) study of two 3rd grade classes found that students given a 'scaffolded' response (given as much or as little help as they needed) out-performed those students given a complete solution as soon as they got stuck, and were more able to apply their knowledge to similar, or only slightly related, tasks.

A comparison of the prompts in writing versus discussion (Furtak & Ruiz-Primo, 2008) reveals that the diversity of students' responses in writing is not reflected in classroom discussions. First, students with "correct" answers are usually more likely to share them during class, whereas those students who are less sure of their responses stay silent. Second, teachers may intentionally or unintentionally provide students with cues to the answers they are seeking, and may end up guiding the conversation away from other student ideas (Furtak, 2006).

A similar explanation can be given to the next kind of teacher's move that seems to lead to incomplete ESRU cycles, that is when several ideas were heard from different students at the same time and the teacher chose to respond to the one that is closest to the 'right' answer. This is rather convenient for the teacher in order to keep the pace of the lesson according to her plan leaving aside the exploration of other students' alternative ideas.

Another factor that seems to impede the effort for formative assessment is the teacher's need to stop in some cases an emerged discussion, considering it irrelevant or not timely aligned with the lesson's plan. This can be attributed to the pressure teachers feel to complete the curriculum on time or their need to stick to the lesson plan.

The aforementioned factors are related to the degree of responsiveness teachers are willing to employ during their formative assessment practice. Teachers in Bell's and Cowie's study (2000) said that they had to manage the degree of responsiveness when doing formative assessment. They were aware that they had to manage the behavior and learning of the whole class as well as that of individuals. They also had to manage attending to the students investigating their own interests and ideas, and to the students learning what was listed in the curriculum. In both these situations, responding to one aspect meant that they could not respond to the other. They could not always be as responsive to a situation as they wished or were able to. However, the essence of formative assessment is the component of action or responsiveness of the teacher and students to the assessment information gathered or elicited (Bell & Cowie, 2001).

The fifth identified factor that impedes teachers' attempts to use "interactions on the fly" is the way teacher tries to elicit students' ideas. When she poses a problematically expressed question that leads to yes/no answer or it is just guiding students to a specific term, there is very little room for students to express their thinking and make it transparent to the teacher and their classmates. In Graesser's study (1995), the questions posed by the teachers tended to be underspecified and were frequently unclear to the students (as manifested by counter-clarification questions). A similar issue raised by Ames (1992) is the nature of the tasks set, which should be novel and varied in interest, offer reasonable challenge, help students develop short-term self-referenced goals, focus on meaningful aspects of learning and support the development and use of effective learning strategies.

When a teacher questions the class, the teacher's beliefs will influence both the questions asked and the way that answers are interpreted. An important principle here is the distinction between 'fit' and 'match' (von Glasersfeld, 1987, p. 13). For example, a teacher may set student problems in solving systems of simple equations. If students answer all the questions correctly, the teacher may well conclude that the students have 'understood' the topic, i.e. the teacher assumes that the students' understanding matches his/her understanding. However, this is frequently not the case. The relationship between fit and match depends critically on the richness of the questions used by the teacher, and this, in turn will depend on the teacher's subject knowledge, their theories of learning, and their experience of learners (Black and William, 1998).

Finally, the incomplete ESRU cycles indicate another factor that relates with the tendency of the teacher to accept an answer without prompting for further interpretation when the answer is seemingly aligned with the established scientific knowledge. This was the most frequent move of the teachers that led to a broken ESRU cycle and, as we will discuss later, sometimes leads to wrong inference of students' understanding and in any case doesn't assure that the other students share the same understanding as the one that gave the answer.

5.4. Challenges that emerge during teachers' attempts to use *interactions on the fly*

In an attempt to shed more light into the intricacies underlying the teachers' effort to employ interactions on the fly as a formative assessment method, we focused on instances where either important information (i.e., contributions made by the students) went unnoticed during the discussion or was used in a non-optimal manner, despite the fact that in some of these cases the ESRU coding was revealing complete cycles of interaction. Looking at each episode at a macroscopic level so that we could realize the impact of teacher's choices during the dialogue on students' understanding, we were able to identify missed opportunities for promoting students' learning and categorize them to three different types (Table 50). This analysis provided insights to the challenges that teachers meet when applying "on the fly interactions".

The first type of missed opportunities relates with cases where the teacher does not respond to an issue that comes up during discussion, considering it as not part of the particular lesson. It seems that some teachers share the view that teaching should be built in a linear manner and she would not deviate from the lesson plan prescribed by the textbook. Heritage (2007) also concludes that teachers' view of science and their concomitant view of teaching science is dominated by tasks and activities rather than conceptual structures and scientific reasoning. These views prevent teacher from realizing the influence of students' misconceptions on different subjects and the need to address them on time. It is a challenge for the teachers to overcome these views that might prevent them from realising the influence of students' misconceptions on developing scientific ideas. Teachers need support in recognising ways to take students' partial understanding forward in a way that recognizes and addresses such misconceptions as they arise in the classroom.

Another explanation for the teachers' decision to put aside emerging issues is the responsibility they feel to cover the formal curriculum. In Cowie's study (2005), student comments during interviews relating to group/class–teacher interactions focused on teacher concern with curriculum coverage. Unprompted, three groups of year 10 students discussed as unproductive the tendency of some teachers (not their current science teacher) to 'rush' through the curriculum. They acknowledged that the teacher has a responsibility to the class to cover the curriculum but they also perceived this practice as undermining understanding as a goal for learning and class attendance (Cowie, 2005).

There are times when something unplanned but significant happens: an unusual comment by a student, evidence of divergent understandings of a particular term, an unexpected outcome of an experiment. Teachers must make on-the-spot judgments about whether to maintain the focus and coherence of the lesson as planned, or to take advantage of a "teachable moment." They must weigh the costs and benefits of shifting course in mid-stream. They must find ways to balance the challenge of keeping the talk focused and academically rigorous with the challenge of including all members of the classroom community as valued, engaged participants, attending to differences in students' cultural and linguistic backgrounds, previous academic preparation, and interests (Michaels et al, 2010). Teachers need to be able to assess whether students' answers are right, wrong or partially right, and to evaluate the extent to which the ideas students express might advance or impede their learning in the short and or the longer term. They need to be able to generate a range of actions to take student learning forward from their current position. Flexibly connected pedagogical content knowledge is central to this kind of teacher decision-making. Importantly, teachers can also make decisions about which aspects of the discussion to act on immediately, which to leave for later consideration and which to ignore.

This characteristic of formative assessment is addressed as the dilemmas faced by the teachers when doing formative assessment. The word "dilemmas" is used as there was no obvious solution to the situation and the decision made in response to each situation would depend on contextual features and the teacher and students concerned. Unlike problems that can be solved, dilemmas are managed and this management relies heavily on the professional judgment of teachers.

Windschitl (2002) described pedagogical dilemmas as "*dilemmas for teachers that arise from the more complex approaches to designing curriculum and fashioning learning experiences that constructivism demands*" (p.132). Teachers can face pedagogical dilemmas when engaging in formative assessment such as centering student understandings as the focus of classroom practice, managing classroom interactions and conversations, understanding content and assessing students' knowledge. While many of these dilemmas are faced by teachers in other content areas outside science when engaging in formative assessment practices, it is the issue of understanding scientific content and practices that is most specific to science teachers.

Ruiz-Primo and Furtak also suggest that instructional responsiveness is a crucial aspect of scientific inquiry teaching. That is, teachers need to continuously adapt instruction to students' present level of understanding rather than pursuing a more teacher-directed instructional agenda. In the absence of the crucial step of using information about student learning, teachers may follow the IRE/F sequence without gathering information about student learning and following an instructional agenda that is unresponsive to evolving student learning. Doing so is a challenging task, as it involves both planning and flexibility, or as Sawyer (2004) termed it, ‘‘disciplined improvisation.’’

The second type of missed opportunity we identified occurs when the teacher has an assumption that might differ from students’ understanding on a particular concept. This occasion leads to a miscommunication between teacher and student, preventing the teacher to have a clear conclusion regarding students understanding. In the relevant example presented in the previous chapter (Tables 42 and 43), the teacher did ask students to elaborate more and express their thinking, using somehow the emergent information and completing the ESRU cycles. At the same time, crucial misconceptions were not elaborated, despite the opportunities that emerged by applying “on the fly” assessment for initiating a discussion about these issues. It seems that it remains challenging for the teacher to make sure that his/her intention to make links between concepts when providing feedback is effective on students’ understanding. This finding is in line with Coffey’s (2011) skepticism about the need to consider disciplinary substance when evaluating the discourse that takes place in the science learning environment.

It is well accepted that is not easy for the teachers to adjust instruction based on what student responses reveal about student thinking (Feldman and Capobianco, 2008). The discourse that unfolds in the classroom is complicated by the possibility that the inputs contributed by individual students or the teacher, could be resting on tacit assumptions, not necessarily shared by all members of that specific learning community. This might influence how the discussion evolves; thus there is a challenge associated with detecting such cases when they occur. This could have an effect on the demands that are placed on the teacher in terms of detecting and managing such instances. Overall, the examples mentioned before illustrate how challenging is for the teachers to identify hidden assumptions that students have about particular concepts and respond at the time, in a way that could help learners overcome their misconceptions.

Ross and Gibson (2010) analyzed data for differences in expert versus less expert teachers’ noticing ability during observation of literacy lessons. Significant differences were found between the noticing ability of expert and less expert participants. Experts demonstrated consistent application of literacy-related content knowledge, detailed perception of meaningful patterns in students' responses, and reasoning or hypothesizing integrated to these observations. They also noticed and commented on most pivotal events. Experts, then, possessed the expected pedagogical content knowledge for literacy instruction and applied this knowledge in depth and with fluency while observing a lesson. Without specific training in professional noticing ability, it is typical for teachers to respond infrequently or superficially to student attempts (Berliner, 2001).

There is a tacit presumption of ‘‘content’’ as a body of correct information, centered on terminology and selected in advance as lesson objectives. When discussion about the nature of the disciplinary objectives is absent, teachers tacitly support traditional views of content as information students should retain. The target content, then, appears to be the body of correct information, in the form of terminology. Effective assessment in science education should involve genuine, extended attention to the substance of student reasoning, on at least two levels (Coffey et al, 2011). Teachers should elicit and pay ‘‘persistent attention’’ (Strike and Posner, 1992) to students' arguments. What reasons do students have for answering as they do? What evidence and logic are they using? In this, the teachers are not only becoming aware of student reasoning but modeling for students how they should focus their attention in science. In other words, they are assessing student reasoning in ways that are consistent with how students should learn to assess ideas as participants in science.

As the formative assessment done by the teachers is often unplanned and responsive, it involves uncertainties and taking risks. Formative assessment needs the teacher finding out and responding to the diverse views of students; it has indeterminate outcomes; it cannot be planned in detail before the lesson; the effects of the required actions are not usually known beforehand; and usually it requires the teacher to take action in the busyness of the classroom. Their confidence in their professional knowledge and skills was seen by the teachers to influence the degree of risk and uncertainty taken (Bell & Cowie, 2000). Professional noticing research acknowledges that teachers can only attend to some aspects of the learning situation and that they subsequently draw on prior teaching experience (Erickson, 2011) along with specialized discipline knowledge to decide what is salient (Sherin, Jacobs, & Phillip, 2011; Sadler, 1989). It is not sufficient to consider only the teacher's actions. The core of formative assessment lies not in what teachers do but in what they see. The point is teachers' awareness and understanding of the students' understandings and progress; that's what the strategies are for. As Coffey states (2011) “*To appreciate the quality of* *a teacher's awareness, it is essential to consider disciplinary substance: What is happening in the class, and of that, what does the teacher notice and consider?*".

In order to be maximally effective, formative assessment requires the interaction of general principles, strategies, and techniques with reasonably deep cognitive-domain understanding. A teacher who has weak cognitive-domain understanding is less likely to know what questions to ask of students, what to look for in their performance, what inferences to make from that performance about student knowledge, and what actions to take to adjust instruction. To create an improvisational classroom (Sawyer, 2004), the teacher must have a high degree of pedagogical content knowledge - to respond creatively to unexpected student queries, a teacher must have a more profound understanding of the material than if the teacher is simply reciting a preplanned lecture or script. An unexpected student query often requires the teacher to think quickly and creatively, accessing material that may not have been studied the night before in preparation for this class; and it requires the teacher to quickly and improvisationally be able to translate his or her own knowledge of the subject into a form that will communicate with that student's level of knowledge (Sawyer, 2004).

Discussing on the fly assessment that ''occurs when 'teachable moments' unexpectedly arise in the classroom'' (p. 4), Shavelson (2006) wrote: "*Such formative assessment and pedagogical action (''feedback'') is difficult to teach. Identification of these moments is initially intuitive and then later based on cumulative wisdom of practice"*. In addition, even if a teacher is able to identify the moment, she may not have the necessary pedagogical techniques or content knowledge to sufficiently challenge and respond to the students. Other research on formative assessment has argued that, although teachers can often make reasonable inferences about student understanding, they face difficulties in making ''appropriate'' instructional moves (Heritage, Kim, Vendlinski, & Herman, 2007). The distinction, between making evidence-based inferences and subsequently adapting instruction, is crucial. The distinction is crucial because a failure in either step can reduce the effectiveness of formative assessment. If the inferences about students resulting from formative assessment are wrong, the basis for adjusting instruction is weakened. Similarly, if the inferences are correct but instruction is adjusted inappropriately, learning is also less likely to occur (Bennet, 2011).

The third type of missed opportunity we identified relates with underestimation of students' weaknesses on scientific methods deriving probably from the significantly less priority given from the teacher to students' epistemological awareness and on the development of their scientific skills. Studies have revealed that many teachers have weaknesses on elaborating and interpreting experimental results (Boudreaux, 2007) and that might be a reason that they avoid analyzing in depth the experimental results in class. This restricts the potential of the assessment method to identify student's difficulties on elaborating experimental data and to help them develop skills related with scientific methods.

Unfortunately, the conclusion to be drawn from much classroom-based research is that teachers' assessment of information related to cognitive goals is often ignored. Instead, what receives priority is information more frequently aligned with the activity goals of the classroom (Duschl & Gitomer, 1998). However, content understanding alone is not enough. Science has particular ways of considering evidence; generating, testing, and evaluating theories; and communicating ideas. A goal of science education is to help students participate in all the practices of the scientific community's culture.

NGSS Framework's (NRC, 2012, p.30) highlights an understanding that conceptual knowledge development is intertwined with the activities of doing science and that together these can be defined as practices. Second, the report outlined specific "cognitive, social, and physical practices" important for engaging in science and argued that the integration of these practices into the learning of content is essential. They suggest that students should engage in scientific practices instead of just learning about them.

Another challenge for the teacher is to create a collaborative environment in class and positive attitude against formative assessment and the assessment dialogue in particular. Collaborative ways of working grounded in trust and respect allow student thinking and ways of working to be revealed, developed and revised as their ideas are affirmed, challenged, adopted and adapted. For students, some of the benefits are in establishing conditions for productive peer assessment, both as an end in itself and as a scaffold towards independent self-monitoring (Sadler, 2010). For teachers, the process of collaborative learning also sets up a chain of evidence as students move through a variety of social contexts that require students to articulate and justify their ideas. As part of this frame, teachers need to ensure that all their students understand the ‘rules of the game’ for how to contribute to classroom discussions.

Another problem may be that students can fail to recognise formative feedback as a helpful signal and guide (Tunstall & Gipps, 1996a). Purdie & Hattie's (1996). Comparative study of the responses of Japanese and Australian students, which aimed to explore their self-regulation strategies, shows that response can be culturally determined. Many researchers report that positive learning gains secured by formative feedback are associated with more positive attitudes to learning (Black and William, 1998).

There were moments during the implementations of this study, that you could “feel” the unwillingness of some students to engage in the classroom dialogue in an actively and fruitful manner. Formative assessment should involve awareness of how students are engaging in disciplinary practices. Are students reasoning about the natural world, or are they focused on what they are ‘‘supposed to say,’’ playing the ‘‘classroom game’’ (Lemke, 1990) of telling the teacher what they think she wants to hear? In a Guessing Game routine, teachers pose a series of related questions to guide students toward a desired conclusion, not unlike Edwards and Mercer’s (1987) description of cued elicitation. Because the end point is often hidden from students, they attempt to guess the correct answer, relying on verbal and nonverbal cues from the teacher (pauses, intonation, gestures, etc.) and/or outright rejection of their responses (Kennedy, 2005, p.103). As one of our teachers noted during the interview *“the teacher needs communication skills that will help him/her to encourage students to express themselves, to be comfortable to say something that could be wrong. Some teachers might have difficulties on that because of their character. But they can improve through training”.*

5.5. Completeness, depth and coherence of classroom dialogue

Duschl’s (2003) epistemic dimension seems to include both the procedures necessary to generate scientific evidence and the reasoning processes involved with generation of scientific knowledge. Classroom evaluation practices generally encourage superficial and rote learning, concentrating on recall of isolated details, usually items of knowledge which students soon forget (Black and William, 1998). This can be attributed to the fact that concepts are not well linked to each other and thus the learning process lucks in coherence and meaning. Ruiz-Primo and Furtak (2006) found that teachers focused on the procedures involved in scientific inquiry rather than the process of developing scientific explanation.

In many cases, the evidence suggests that teachers are focused on finding out what the students already know of the target information (Coffey et al, 2011). However, perhaps out of a desire to have the widest relevance, many studies have focused on strategies that cut across topics and disciplines, such as wait time or ‘‘stop lighting’’ or questioning, without closely examining the ideas and reasoning they reveal. By not delving into the specific substance of student thinking, the literature - and, subsequently, practice - misses and may undermine its fundamental objective. However, there is evidence that the quality of dialogue in a feedback intervention is important (Graesser & Person, 1995) and can, in fact, be more significant than prior ability and personality factors combined (Clarke, 1988).

An exception is the research by Clarke (1988) on classroom dialogue in science classrooms. He analysed the discourse of three teachers in four classrooms, grading the quality of the discourse by summation over four criteria. These included the numbers of interpretable themes, the numbers of cross-correlations (an indicator of coherence) and proportions of themes explicitly related to the content of the lessons. As he claims, such a sophisticated analysis of dialogue indicates specific areas of weakness which could then be remedied by appropriate training. It could also be used to produce ideal "templates" of various models of teaching (e.g. Brady; 1985) for use as a guide for lesson planning.

We challenge the view that it is difficult for teachers to learn to attend to the substance of student thinking. Recent work in science and math teacher education (Coffey, Edwards & Finkelstein, 2010; Kazemi et al., 2009; Levin, Hammer, & Coffey, 2009; Levin & Richards, 2010; Singer-Gabella et al., 2009; Windschitl, Thompson, & Braaten, 2011), has presented evidence of novice teachers’ attention to student thinking, novices whose preparation emphasized awareness and interpretation of student thinking as evident in video records and written work. By this reasoning, much depends on how teachers frame what they are doing, and a primary emphasis on strategies may be part of the problem. Assignments that direct teachers and teachers-in training to what they are doing may inhibit their attending to what students are thinking.

Our concept maps enabled us to represent the conceptual coherence and depth of the classroom dialogue, in four dimensions:

- The variety of concepts that emerge in discussion.
- How much are these interpreted,
- How well are linked with each other.
- The necessary but missing or not interpreted concepts.

More specifically the various concepts that emerged in discussion were depicted in boxes that were coloured differently (light blue or light green) to denote whether they were brought in the discussion by the teacher or a student. This allows to illustrate whether all the essential concepts for describing the phenomenon or scientific process under discussion at the particular episode are included in the dialogue. As the concept maps reveal, some episodes were long and rich in concepts relevant to the phenomenon under discussion while other episodes were short and included very few concepts. As we observed, in some episodes, necessary concepts for the understanding of the phenomenon were missing from the discussion. These were depicted red coloured in order to illustrate the lack of completeness of the dialogue episode.

These red coloured boxes relate also with the second category of missed opportunities that were previously discussed, when the teacher has an assumption that might differ from students' understanding on a particular concept. Often, the teacher has the impression that the students have made the connection with the missing concept in their minds, without explicitly referring at it. However, as it was shown before from the macroscopic analysis of episodes, this is not always the case.

Concepts are linked with arrows on the concept maps. The direction of the arrows depicts the order the concepts appeared in the dialogue, while the number on the arrow declares the number of ESRU cycles that this link was made for (if it was made more than once). This number is considered as an indicative measure for the depth that the particular concepts were discussed. Again, the concept maps we developed suggest that episodes of classroom dialogue vary between each other. Some of them illustrate that concepts were linked between each other various times, suggesting that the relevance between them and with the phenomenon was well discussed, while others show that concepts were just mentioned one after the other without much discussion about the relation between them. In addition, concepts depicted as crossed on the concept maps, are concepts that were mentioned by students but were ignored by the teacher and thus not discussed at all. These items relate to the first type of missed opportunities when the teacher does not respond to an issue that comes up during discussion, considering it as not part of the particular lesson.

Furthermore, the concept maps we developed depict the coherence of the classroom dialogue during each episode. The indicator for this are the loops that concepts form when each concept is linked with other concepts which are also linked with each other. This robust connections suggest that the dialogue included a holistic approach for the phenomenon under discussion instead of fragmented presentation of various concepts.

Overall we claim that the concept maps we developed allow a representation of the dialogue that indeed reveals the quality of each episode of the classroom dialogue in three dimensions: the completeness of the dialogue, the depth that concepts were discussed and linked to each other and, the coherence of the dialogue. These representations revealed also that the dialogue episodes can be very different between each other regarding several of the aforementioned dimensions, even if they were realized by the same teacher in the same class.

5.6. Relation between the completeness, depth and coherence of classroom dialogue and the quality of informal formative assessment

The next step was to quantify the conceptual coherence and depth of the dialogue as it is depicted on the concept maps; in other words to find a value for each of the three dimensions (completeness, depth and coherence) for each episode. These values were then compared with the quality of informal formative assessment as measured by the completion rate of ESRU cycles in each episode. This comparison revealed that the rate of completion of ESRU cycles of each episode can be an indicator for the conceptual coherence and depth of the classroom dialogue but, still not for the completeness of the dialogue. This supports our statement presented earlier that there might be missed opportunities for improving students’ understanding, particularly because specific critical concepts were not brought up and discussed during the dialogue, even when the teacher takes special care to use the emergent information.

This result is in line with the conclusion presented by Coffey et al. (2011) that researchers and their analysis focus on the strategies teachers use when engaging in formative assessment, while there is less attention on how they manage with disciplinary substance. Similarly, our analysis depicted that the ESRU coding scheme is not sufficiently representing the degree teachers notice students' hidden assumptions regarding critical concepts for understanding the phenomenon under discussion. There is variation between students' ideas and actions (Gibson & Ross, 2016) and variability in what teachers notice (Robertson, Richards, Elby, & Walkoe, 2015). As Mason and Davis (2013) explain, it is what a teacher is attuned to notice and how they connect what they notice with possible pedagogical actions that is central to teachers being adaptive and responsive. Therefore it is important to develop coding schemes and improve our analysis in a way that considers disciplinary substance and noticing students' hidden assumptions and misconceptions during classroom dialogue. We believe we have shown that concept maps are a useful tool for depicting the conceptual completeness, depth and coherence of classroom dialogue during on the fly interactions as a means for formative assessment.

6. CONCLUSIONS

In conclusion, by analysing the dialogue that evolves in class when teachers apply on the fly interaction as a method of formative assessment, we were able to identify patterns that appear and relate between them the various ways the teachers used to elicit student’s ideas and understanding, the ways the students responded to these efforts of the teachers and how the teachers recognised and used them in a productive way to promote students learning. These patterns revealed that the episodes of classroom dialogue were dominated from teachers’ closed-ended questions that lead to short and unexplained answers from the part of students. This kind of answers provide little information regarding their understanding which can be used to improve their learning. However, in those instances that the teachers did encourage their students to explain their thinking or to provide evidence to support their answer, the students usually responded accordingly presenting explanations, examples and data from their experiments to support their opinion.

Another important outcome from the patterns we identified is that where the teacher avoids providing evaluative feedback on students’ responses, he or she allows more time and space for students to evaluate their peer’s opinion and express their own, enriching the dialogue and allowing more fruitful information regarding their understanding to surface and become available for the teacher to use it in a productive way. In other words, when the students are allowed to propose different answers throughout the discussion and the teacher does not evaluate any given answer, but instead facilitates a collaborative improvisation among the students, they are guided toward the social construction of their own knowledge.

On the other hand, ignoring or rejecting students’ contribution discourages them from participating in the dialogue and disclosing their thinking. Even by affirming and accepting an answer that seems aligned with the established scientific knowledge without questioning it, transmits the wrong message to the students that the teacher is only looking for a specific response instead of examining student’s degree of understanding.

In addition, we were able to categorize practices and strategies that teachers follow and lead to productive use of the emergent information in their effort to promote students’ learning. On the other hand, we identified challenges that the teachers confront that relate both with their teaching practice but also with the complexity of the learning environment. The distinctions between some of the categories were difficult to make objectively given that they rely on subtle judgments of conceptual intent and the context of the statement. The categories, however, are distinct in their intended function. We rather argue, that can serve as a useful tool to make apparent teachers' discursive moves in promoting student understanding and reasoning, and to help teachers' articulate powerful ways of doing this.

A limitation of the ESRU coding system we used and the subcategories we developed is that the number of students participating actively in discussion is not depicted. In other words, each time code S is appeared it could be either the same or different student contributing in the dialogue. Consequently, if we are interested on what extend students are actively engaged in the classroom dialogue, there is a need for a modification on the coding scheme that could include this information.

Another limitation of this research is the small number of teachers that participated. We decided to devote more time for observing more lessons of these three teachers in order to better capture their teaching strategies when applying "on the fly" interaction and how their students respond to them. However, future research might examine whether the patterns we identified during classroom dialogue apply in other teachers' teaching practice. In addition, further research could examine whether the patterns we identified apply as such, in different contents of science teaching like electric circuits, electromagnetism, energy etc., or even in teaching of other subjects like history or literature.

Nevertheless, we believe that this study contributes on the way to identify what it means to be an instructionally responsive teacher in the context of scientific inquiry. We aim to help teachers to understand the differences between asking questions for the purpose of recitation and asking questions for the purpose of eliciting information and for improving student learning. Pre- and in-service teachers alike may benefit from learning about the ESRU cycle as a way of thinking about classroom discussions as assessment conversations, or opportunities to understand the students' understanding and to move students toward learning goals in an informal manner.

Improvisational teaching requires a teacher who can facilitate discussion among students. Beginning teachers need routines, but also need to learn how to flexibly apply them. Borrowing a page from scripted instruction, beginning teachers could be explicitly provided with a set of routines; but in creative teaching, those routines would be designed to allow variation and embellishment. In addition, beginning teachers have great difficulty mastering the ability to lead collaborative discussion, and these techniques tend to be used effectively only by experienced teachers who also possess profound content knowledge.

Although an experienced teacher may have encountered most of the potential students' answers in prior years, a teacher cannot know exactly which answer will be proposed on any given day. And even with years of experience, a teacher cannot predict how the rest of the class will respond to a proposed answer; the flow of the discussion is collaboratively determined by all of the students responding to each other, and the knowledge that is co-constructed by the students emerges from the improvisational flow of dialogue. The classroom is collaboratively creative; the teacher is not the sole creative force, but rather a facilitator for the entire group's creativity (Sawyer, 2004). In addition, increased opportunities for students to make sense of their developing understandings can be facilitated through more effective small group conversations. However, many teachers do not engage in small group work as students can so easily be off task and learning can be limited. Thus support for engaging in effective small group talk formats of classroom conversations should also be offered to teachers in practice.

Teachers require the professional knowledge and skills to plan for assessment; observe learning; analyse and interpret evidence of learning; give feedback to learners and support learners in self-assessment. Teachers should be supported in developing these skills through initial and continuing professional development (Assessment Reform Group (ARG), 2002). Providing teachers with tools that will help them to integrate assessment into the course of everyday instruction to meet the goals of a curriculum will help to realize educational reforms. We believe that both the patterns we have identified and the various strategies that teachers employ when using the emergent information during dialogue that we have described, can be useful tools in hands of pre and in-service teachers' trainers that would like to enhance teachers in their effort to apply a formative assessment method like "on the fly" interaction. Teachers need solid examples and practice along with self-reflection to be able to improve their teaching abilities, in particular when it comes to applying such a demanding method that they have to react on time to what comes up during the classroom dialogue.

Some of the most critical aspects of implementation, for example, pushing students to support their claims with reasons and evidence, or encouraging students to argue with each other, are difficult for any teacher, especially those with limited teaching experience and weak backgrounds in science. Teachers need substantial knowledge to implement formative assessment effectively in classrooms. It is doubtful that the average teacher has that particular knowledge, so most teachers will need substantial time and support to develop it. One possible solution may be to avoid focusing on the structure of formative assessments and instead work with teachers to improve their ability to lead whole-class discussions that truly engage learners in sharing and arguing their ideas. A specific way we propose is through watching video of themselves or other colleagues and reflect on them with the guidance of a tutor. Our teachers who were interviewed after their implementations and watched videos with extracts from their teaching practice were able to identify instances where they could manage the classroom dialogue in a more productive way. Prompts from the researcher to focus on specific parts of the dialogue helped them also to think instances of the dialogue from another perspective and notice missed opportunities to better recognize students' difficulties or misconceptions.

Additionally, teachers will need useful classroom materials that model the integration of pedagogical, domain, and measurement knowledge (e.g., developmentally sequenced tasks that can help them make inferences about what students know with respect to key domain competencies, and about what next to target for instruction). Successful enactment of group discussion takes a great deal of practice and coaching from more experienced practitioners and curriculum experts.

Concept maps that depict the completeness, the depth and the coherence of the dialogue might be helpful tools both for teachers' coaches in an effort to assess a lesson they have observed and provide teachers with valuable feedback. On the other hand, the teachers themselves could benefit by developing such a concept map of their own lesson for purpose of reflecting on their teaching practice. Further research could explore possible ways to improve the representation of the completeness, the depth and the coherence of the dialogue using concept maps.

REFERENCES

Airasian, P. W., & Russell, M. K. (2001). *Classroom assessment: Concepts and applications*. Boston: McGraw-Hill.

Alexander, R. (2006). Towards dialogic teaching: Rethinking classroom interaction, York, UK: Dialogos.

Allen, M., Matthews, C., & Parsons, S. (2013). A second-grade teacher's adaptive teaching during an integrated science-literacy unit. *Teaching and Teacher Education*, *35*, 114–125.

Ames, C. (1992) Classrooms: goals, structures, and student motivation, *Journal of Educational Psychology*, *84*, pp. 261-271.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of science teacher education, 13*(1), 1-12.

Araceli Ruiz-Primo, M., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3-4), 237-263.

Assessment Reform Group. (1999). Assessment for learning: Beyond the black box. *Université de Cambridge, School of Education.*

Ayala, C. C., Shavelson, R. J., Araceli Ruiz-Primo, M., Brandon, P. R., Yin, Y., Furtak, E. M. & Tomita, M. K. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education*, *21*(4), 315-334.

Bell, B. (1995). Interviewing: A technique for assessing science knowledge. Learning science in the schools: *Research reforming practice*, 347-364.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science education, 85*(5), 536-553.

Black. (1998). *Testing: Friend or foe? Theory and practice of assessment and testing.* New York, NY: Falmer Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.

Black, P. (2001). Formative assessment and curriculum consequences. *Curriculum and assessment*, 1, 7.

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education (UK).

Black, P. (2004). *Working inside the black box: Assessment for learning in the classroom*. Granada Learning.

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

Black, P. (2010). Formative Assessment. In P. Penelope, E. Baker & B. McGaw (Ed.), *International Encyclopedia of Education* (pp. 359-364). Bristol, UK.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International journal of educational research*, *35*(5), 463-482.

Boudreaux, A., Shaffer, P. S., Heron, P. R., & McDermott, L. C. (2008). Student understanding of control of variables: Deciding whether or not a variable influences the behavior of a system. *American Journal of Physics*,*76*(2), 163-170.

Buck, G. A., Trauth-Nare, A., & Kaftan, J. (2010). Making formative assessment discernable to pre-service teachers of science. *Journal of Research in Science Teaching*, *47*(4), 402-421.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. *Formative classroom assessment: Theory into practice*, 43-62. Chicago

Brophy, J., Good, T., & Wittrock, M. C. (1986). Handbook of research on teaching. *Teacher behavior and student achievement*, 328-375.

Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students, conceptions of assessment and mathematics achievement: Evidence for the power of self-regulation.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1-6.

Carlsen, W. S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of educational research*, *61*(2), 157-178.

Cazden, C. B. (2001). The language of teaching and learning. *The language of teaching and learning*, 348-369.

Chi, M.T.H., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Chin, C., & Teou, L. Y. (2009). Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education, 31*(10), 1307-1332.

Chung, G. K., Shel, T., & Kaiser, W. J. (2006). An exploratory study of a novel online formative assessment and instructional tool to promote students' circuit problem solving. *The Journal of Technology, Learning and Assessment, 5*(6).

Clarke, J. (1988) Classroom dialogue and science achievement, *Research in Science Education, 18*, pp. 83-94.

Claxton, G. (1984). *Live and learn: An introduction to the psychology of growth and change in everyday life*. Harpercollins College Division.

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

Cowie, B. (2005). Student commentary on classroom assessment in science: A sociocultural interpretation. *International Journal of Science Education*, *27*(2), 199-214.

Cowie, B., Harrison, C., & Willis, J. (2018). Supporting teacher responsiveness in assessment for learning through disciplined noticing. *The Curriculum Journal*, 1-15.

Dassa, C. (1990) From a horizontal to a vertical method of integrating educational diagnosis with classroom assessment, *Alberta Journal of Educational Research*, 36, 35-44

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R. T. K. E., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Jama*, *296*(9), 1094-1102.

Day, J. D., & Cordón, L. A. (1993). Static and dynamic measures of ability: An experimental comparison. *Journal of educational psychology*, *85*(1), 75.

Dillon, J. T. (1983). *Teaching and the Art of Questioning. Fastback 194*. Phi Delta Kappa, Eighth and Union, Box 789, Bloomington, IN 47402.

Driver, R., Leach, J., Scott, P., & Wood-Robinson, C. (1994). Young people's understanding of science concepts: Implications of cross-age studies for curriculum planning.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International journal of science education, 25*(6), 671-688.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation, 14*(7), 1-11.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, *4*(1), 37-73.

Duschl, R. (2000). Making the nature of science explicit. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education. The contribution of research* (pp. 185–206). Buckingham, UK: Open University Press.

Duschl, R. A. (2003). Assessment of inquiry. In J.M. Atkin&J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington,VA: National Science Teachers Association Press.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education, 32*(1), 268-291.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, 47(2), 123-182.

Empson, S.B. (2003). Low-performing students and teaching fractions for understanding: An interactional analysis. *Journal for Research in Mathematics Education, 34*(4), 305-343.

Erickson, F. (2011). *On noticing teacher noticing*. In M. Sherin, V. Jacobs, & R. Phillip (Eds.), Mathematics teacher noticing (pp.18–34). New York, NY: Routledge.

Feldman, A., & Capobianco, B. M. (2008). Teacher learning of technology enhanced formative assessment. *Journal of Science Education and Technology, 17*(1), 82-99.

Fennema, E., Carpenter, T.P., Franke, M.L., Levi. L., Jacobs, V.R., & Empson, S.B. (1996). A longitudinal study of learning to use children’s thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434

Freeman, R. & Lewis, R. (1998) *Planning and implementing assessment* (London, Kogan Page).

Furtak, E. M., & Ruiz ‐ Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education, 92*(5), 799-824.

Garnett, P. J., & Tobin, K. (1989). Teaching for understanding: Exemplary practice in high school chemistry. *Journal of Research in Science Teaching, 26*(1), 1-14.

Gipps, C (2002). *Sociocultural perspectives on assessment*. In G. Wells and G. Claxton (eds.) Learning for Life in the 21st Century (Oxford: Blackwell Publishers).

Gibson, S., & Ross, P. (2016). Teachers’ professional noticing. *Theory into Practice, 55*(3), 180-188.

Gómez Puente, S. M., van Eijck, M., & Jochems, W. (2013). Facilitating the learning process in design-based learning practices: An investigation of teachers’ actions in supervising students. *Research in Science & Technological Education, 31*(3), 288-307.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one ‐ to ‐ one tutoring. *Applied cognitive psychology*, *9*(6), 495-522.

Gutierrez, K.D. (1994). How talk, context, and script shape contexts for learning: A cross-case comparison of journal sharing. *Linguistics and Education, 5*(3-4), 335-365.

Harlen, W. (2006). Teachers' and children's questioning.'. *ASE Guide to Primary Science. Hatfield: Association for Science Education*.

Harlen, W. (2010). Assessment in schools - primary science. In P. Penelope, E. Baker & B. McGaw (Ed.), *International Encyclopedia of Education* (pp. 293-299). Bristol, UK.

Harrison, C. (2015). Assessment for learning in science classrooms. *Journal of Research in STEM Education*, *1*(2), 78-86.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do*?. Phi Delta Kappan, 89*(2), 140-145.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment?. *Educational Measurement: Issues and Practice, 28*(3), 24-31.

Heritage, M. (2010). Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity?. *Council of Chief State School Officers*.

Howe, C., & Mercer, N. (2007). Children's social development, peer interaction and classroom learning. *The primary review*. Cambridge: University of Cambridge.

Hickey, D.T., Ingram-Goble, A.A., & Jameson, E.M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology*, *18*(2), 187-208.

Hickey, D. T., McWilliams, J., & Honeyford, M. A. (2011). Reading Moby-Dick in a participatory culture: Organizing assessment for engagement in a new media era. *Journal of Educational Computing Research*, *45*(2), 247-263.

Hickey, D. T., & Filsecker, M. K. (2012). Participatory assessment for organizing inquiry in educational videogames and beyond. In K. Littleton, E. Scanlon, & M. Sharples (Eds.), *Orchestrating inquiry learning: Contemporary perspectives on supporting scientific inquiry learning* (pp. 146–174). London: Taylor and Francis.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in higher education*, *27*(1), 53-64.

Hutchinson, P., & Hammer, D. (2009). Attending to student epistemological framing in a science classroom. *Science Education, 94*(3), 506-524.

Hyland, P. (2000). Learning from feedback on assessment. *The practice of university history teaching,* 233-247.

Ivie, S. D. (1998). Ausubel's learning theory: An approach to teaching higher order thinking skills. *The High School Journal, 82*(1), 35-42.

Jacobs, V., Lamb, L., & Philipp, R. (2010). Professional noticing of children’s mathematical thinking. *Journal for Research in Mathematics Education, 41*, 169–202

Jacobs, V. R., & Empson, S. B. (2016). Responding to children’s mathematical thinking in the moment: an emerging framework of teaching moves. *ZDM, 48*(1-2), 185-197.

Kennedy, M. M. (2005). *Inside teaching*. Harvard University Press.

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance. *Psychological Bulletin, 119*(2), 254–284.

Kyriazi, E., & Constantinou, C. (2004). The Science Fair as a Means for Developing Investigative Skills in Elementary School. In *Proceedings of the 1st International Conference on Hands on Science* (pp. 5-9).

Laurillard, D. (2013). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. Routledge.

Lemke, J. L. (1990). *Talking science: Language, learning, and values.* Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).

Lewis, E. B. (2008). Content is not enough: A history of secondary earth science teacher preparation with recommendations for today. *Journal of Geoscience Education*, *56*(5), 445-455.

Linge, N., & Parsons, D. (2006). Problem-based learning as an effective tool for teaching computer network design. *IEEE Transactions on Education, 49*(1), 5-10.

Linn, M. C., Davis, E. A., & Bell, P. (2004). Inquiry and technology. *Internet environments for science education*, 3-28.

Lyons, J. S., & Brader, J. S. (2004). Using the learning cycle to develop freshmen's abilities to design and conduct experiments. *International Journal of Mechanical Engineering Education, 32*(2), 126-134.

Marshall, J.C., Horton, B., & Smart, J. (2009). 4E × 2 instructional model: Uniting three learning constructs to improve praxis in science and mathematics classrooms. *Journal of Science Teacher Education, 20* (6), 501-516.

Mason, J. (2002). *Researching your own practice: The discipline of noticing.* London: Routledge-Falmer.

Mason, J., & Davis, B. (2013). The importance of teachers' mathematical awareness for in-the-moment pedagogy. *Canadian Journal of Science, Mathematics and Technolog Education, 13*(2), 182-197.

McConnell, D. A., Steer, D. N., & Owens, K. D. (2003). Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education, 51*(2), 205-216.

Mercer, N. (2004). Sociocultural discourse analysis. *Journal of applied linguistics, 1*(2), 137-168.

Mercer, N., and Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. Learning, *Culture and Social Interaction, 1*, 12-21.

Mercer, N., & Dawes, L. (2014). The Study of Interaction between Teachers and Students, from the 1970s until the 2010s. *Oxford Review of Education, 40*(4), 430–445.

Messick, S. (1994). The interplay of evidence and consequence in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2010). Accountable talk sourcebook: For classroom conversation that works. *Pittsburgh, PA: University of Pittsburgh Institute for Learning.*

Mishler, E. G. "Sttidies in Dialogue and Discourse: IL Types of Discourse Initiated by and Sustained Through Questioning." *Journal of Psycholinguistic Research 4* (1975)-99-121.

National Research Council. (2001). Inquiry and the national science education standards. Washington, DC: National Academy Press.

Nelson, J., Robison, D.F., Bell, J.D. & Bradshaw, W.S. (2009). Cloning the professor, an alternative to ineffective teaching in a large course. *CBE Life Sciences Education*, *8*(3), 252-263.

Nicoll, G. (2001). A three-tier system for assessing concept map links: a methodological study. *International Journal of Science Education, 23*(8), 863-875.

Nicol, D. J., & Macfarlane ‐ Dick, D. (2006). Formative assessment and self ‐ regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Nieminen, P., Correia, C. F., Hähkiöniemi, M., Serret, N., Viiri, J., & Harrison, C. (2015). Formative assessment in inquiry-based science education using interactions on-the-fly.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of research in science teaching, 27*(10), 937-949.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* Cambridge University Press.

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 261-290.

Nystrand, M., Wu, L. L., Gamoran, A., Zeisler, S., & Long, D. A. (2003). Questions in time: Investigating the structure and discourse. *Discourse Processes, 35*(2), 135–198

O’Connor, M.C. & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly, 24*(4), 318-335

Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching, 45*(4), 497-523

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., ... & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences, 13*(3), 337-386.

Parsons, S. A., Davis, S. G., Scales, R. Q., Williams, J. B., & Kear, K. (2010). How and why teachers adapt their literacy instruction. *Mentoring literacy professionals: Continuing the spirit of CRA/ALER after, 50*, 221-236.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. *National Center for Education Research*.

Parsons, S. A. (2012). Adaptive teaching in literacy instruction: case studies of two teachers. *Journal of Literacy Research, 44*, 149–170.

Pierson, J. (2008). The relationship between patterns of classroom discourse and mathematics learning. University of Texas at Austin.

Pressley, M., Wood, E., Woloshyn, V.E., Martin, V., King, A. & Menke, D. (1992) Encouraging mindful use of prior knowledge—attempting to construct explanatory answers facilitates learning, *Educational Psychologist, 27*, pp. 91-1.

Purdie,N. &Hattie,J. (1996) Cultural differences in the use of strategies for self-regulated learning, *American Educational Research Journal, 33*, pp. 845-8

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4-13.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13*(3), 273-304.

Robertson, A. D., Richards, J., Elby, A., & Walkoe, J. (2015). Documenting variability within teacher attention and responsiveness to the substance of student thinking. *Responsive teaching in science and mathematics*, 227-248.

Ross, P., & Gibson, S. A. (2010). Exploring a conceptual framework for expert noticing during literacy instruction. *Literacy Research and Instruction*, *49*(2), 175-193.

Roth, W. M., & Roychoudhury, A. (1994). Science discourse through collaborative concept mapping: new perspectives for the teacher. *International journal of science education, 16*(4), 437-455.

Roychoudhury, A., & Roth, W. M. (1996). Interactions in an open‐inquiry physics laboratory. *International Journal of Science Education, 18*(4), 423-445.

Ruiz-Primo, M. A., & Shavelson, R. J. (1997). Concept-Map Based Assessment: On Possible Sources of Sampling Variability.

Ruiz‐Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of research in science teaching, 44*(1), 57-84.

Sawyer, R.K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation. *Educational Researcher*, *33*, 12–20.

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for research in Mathematics Education,* 187-215.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of research in science teaching*, 28(9), 859-882.

Senk, S.L., Beckman, C.E. & Thompson, D.R. (1997) Assessment and grading in high school mathematics classrooms, *Journal for Research in Mathematics Education, 28,* pp. 187-2

Shavelson, R. J. (2006). On the integration of formative assessment in teaching and learning with implications for teacher education. *Stanford Education Assessment Laboratory and the University of Hawaii Curriculum Research and Development Group. Retrieved December, 27,* 2011.

Shepard, A. L. (2005). Linking Formative Assessment to Scaffolding. *Educational Leadership, 63*(3), 66-70.

Sherin, M. G., & Jacobs, V. R. (2011). Situating the study of teacher noticing. In *Mathematics teacher noticing* (pp. 33-44). Routledge.

Schoenfeld, A. H. (1985). Making sense of "out loud" problem-solving protocols. *The Journal of Mathematical Behavior, 4*(2), 171-191.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford Univ Pr.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement, 26*(3), 233-246.

Stoddart, T., Abrams, R., Gasper, E., & Canaday, D. (2000). Concept maps as assessment in science inquiry learning-a report of methodology. *International Journal of Science Education, 22*(12), 1221-1246.

Swain, J.R.L. (1991) The nature and assessment of scientific explorations in the classroom, *School Science Review, 72*(260), pp. 65-76.

Tiberghien, A., & Malkoun, L. (2009). The construction of physics knowledge in the classroom from different perspectives. *Transformation of Knowledge through Classroom Interaction, Routledge, London*, 42-56.

Tillema, H. (2010). Formative Assessment in Teacher Education and Teacher Professional Development. In P. Penelope, E. Baker & B. McGaw (Ed.), *International Encyclopedia of Education* (pp. 563-571). Bristol, UK.

Tippins, D. J., Koballa, T. R., & Payne, B. D. (2002). Learning from cases: Unraveling the complexities of elementary science teaching. *Boston: A Pearson Education Company*.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Springer, Dordrecht.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631.

Tunstall, P., & Gipps, C. (1996). 'How does your teacher help you to make your work better?'Children's understanding of formative assessment. *The Curriculum Journal, 7*(2), 185-203.

Tunstall, P., & Gipps, C. (1996). Teachers' feedback to young children in formative assessment: a typology. *British Educational Research Journal, 22*(4), 389–410.

Tytler, R., & Aranda, G. (2015). Expert teachers' discursive moves in science classroom interactive talk. *International Journal of Science and Mathematics Education, 13*(2), 425-446.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education, 45*(4), 477-501.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Vygotsky, L. (1986). *Thought and language.* Cambridge: The MIT Press

Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *The Journal of the Learning Sciences, 15*(1), 63-119.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker, 9*(5), 2-3.

Wenning, C. (2005). Levels of Inquiry: Hierarchies of Pedagogical Practice and Inquiry Processes. *Journal Physics Teacher Education Online, 2*(3), 3-12.

Wiliam, D. (1992). Some technical issues in assessment: a user's guide. *British Journal of Curriculum and Assessment, 2*(3), 11–20.

Wiliam, D. (2007). Keeping learning on track. *Second handbook of research on mathematics teaching and learning, 2,* 1053-1098.

Wiliam, D., & Thompson, M. (2017). Integrating assessment with learning: What will it take to make it work?. In *The future of assessment* (pp. 53-82). Routledge.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research, 72*(2), 131-175.

Wood, T., Williams, G. & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures. *The Journal for Research in Mathematics Education, 37*(3), 222-255.

Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences, 6*(2), 227-269.

Vosniadou, S., & Ioannides, C. (1998). From conceptual development to science education: A psychological point of view. *International Journal of Science Education, 20*(10), 1213-1230.