PEDAGOGISCHE STUDIËN
TIJDSCHRIFT VOOR ONDERWIJSKUNDE EN OPVOEDKUNDE

Learning outcomes of exploratory talk in collaborative activities

J. T'Sas and S. De Maeyer

Abstract How effective are collaborative activities for students' learning? Among the many studies which suggest an answer to this question, a considerable number proposes one important condition: the use of exploratory talk. Over the last decades significant UK experiments on this matter have been repeated, refined or elaborated upon in other countries. As direct replicator studies are an important means to generalise results in different contexts, a similar study was set up in Flanders, the Flemish part of Belgium. During a two-year quasi-experiment groups of eleven and twelve year old primary school students were taught the basic principles of exploratory talk and put these into practice in group assignments for eight consecutive weeks. Pre- and post-tests included measurement of reasoning and problem solving skills. The results confirm the original UK study as described in Mercer et al. (1999) as well as other, similar studies: students of the experimental group improved their reasoning and problem solving skills at group level significantly, whereas control groups made no such progress.

Keywords exploratory talk, conversation skills, collaborative learning, dialogic teaching

119

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48
(102) 119-160

Artikelgeschiedenis
Ontvangen: 1 mei 2024
Ontvangen in gereviseerde vorm: 19 februari 2025
Geaccepteerd: 31 maart 2025
Online: 25 augustus 2025

Contactpersoon
Jan T'Sas,
jan.tsas@uantwerpen.be

Financiering onderzoek
-
Belangen
-

1 Introduction

Over the last decades many researchers and educators have promoted the use of students' conversational skills in collaborative activities in the classroom (Alexander, 2008; Barnes, 1976; Coultas, 2012; Mercer, 1995; Rojas-Drummond et al., 2013; Sutherland, 2006). Accordingly, there is considerable support for the sociocultural idea that knowledge is a co-constructed activity of students and that collaborative talk can support them to develop high-order thinking, high-level understanding, the voicing of personal opinions and ideas, and argumentation skills (Boyd & Kong, 2017; Fernandez et al., 2001; Harris, 1995; Lofgren et al., 2013; Mercer & Littleton, 2007; Mercer et al., 1999; Soter et al., 2008; Sutherland, 2013; Topping & Trickey, 2014).

These findings notwithstanding, general teaching practice appears to be dominated by rather monologic classroom approaches (e.g. Cazden, 2001; Haneda & Wells, 2008; Mercer & Dawes, 2014). Additionally, for those teachers who do embrace a more dialogic pedagogy, implementing collaborative strategies often turns out to be ineffective, as without sufficient preparation true collaboration between students easily becomes parallel working, making partner or group work a wasted educational opportunity (Dawes et al., 1992; Mercer, 2010b). True collaboration demands proper organization and structure: clear goals and instructions, the right size and composition of groups, shared and individual responsibilities, specific kinds of group work, collaborative strategies and systematic assessment (Kagan, 2014; Slavin, 1996), but it also requires certain conversational skills (Mercer, 1995).

In the early nineties, British researchers started a series of experiments which resulted in the identification and stimulated use by students of a type of conversation which showed high potential for learning: *exploratory talk* (Dawes et al., 1992; Mercer, 1995; Wegerif & Mercer, 1997a). Positive learning effects of this type of talk have inspired researchers all over the world to replicate, refine or elaborate on these early studies, using both quantitative and qualitative methods and often finding similar positive linguistic, cognitive, social, psychological and pedagogical effects (for an overview, see T'Sas, 2018). The term *exploratory talk* was launched by Barnes (1976). Since then associated constructs have been introduced and become the subject of an expanding amount of research, such as *transactive reasoning* (Berkowitz & Gibbs, 1983), *accountable talk* (Michaels et al., 2008) and *dialectic argumentation* (Asterhan, 2013). As research on this topic grew, these concepts were further refined or elaborated upon (for an overview, see T'Sas, 2018). Meanwhile, in Flanders, the Dutch speaking community of Belgium, the notion and learning potential of exploratory talk has remained largely unnoticed. The aim of this study is to determine whether Mercer's

https://doi.org/10.59302/cj4khk48

original experiment and its outcomes as described extensively in Mercer et al. (1999) would be generalisable to the Flemish educational context. For that purpose, this study was replicated in five Flemish primary schools.

2 Literature review

2.1 International research on classroom interaction

In order to establish which competencies (young) people need to function effectively as active and responsible citizens, employees and learners in 21st century society, the United Nations and the European Commission (OECD) have developed 'Key Competencies for the 21st century' (Van den Branden, 2015). These competencies focus heavily on the skillful processing and application of knowledge, language and information and of 'learning power' (Van den Branden, 2015; Van den Branden & Van Gorp, 2000). From an educational point of view, these Key Competencies can be traced back to principles of constructivist thinking about knowledge and learning which find their origin in the theories of Bruner (1960), Piaget (1970) and Vygotsky (1978).

From a constructivist perspective (De Corte, 1996) describes learning as a constructive, cumulative, self-regulatory, intentional, context bound, collaborative and individual process of acquiring knowledge, giving meaning and developing skills. Zooming in on the aspect of collaborative learning, the Russian educational psychologist Vygotsky (1978) was one of the first theoreticians to highlight the importance of verbal interaction in education as a means for learning. Vygotsky's work has laid the foundation for the so-called sociocultural theory which explains cognitive development and learning in a cultural and social context. It focuses especially on the dialogic approach of learning and 'studies how people use language as a social mode of thinking' (Mercer, 1995, p. 4). In Vygotsky's theory sociocultural researchers and educators found common ground to promote a dialogic pedagogy which includes the open exchange of ideas, jointly undertaken inquiry, mastery of disciplinary knowledge and ways of reasoning, engagement with multiple voices and perspectives, and respectful classroom relations (Haneda & Wells, 2008). It includes teacher-student interaction in the form of dialogic teaching on the one hand (Alexander, 2008) and student-student interaction in the form of exploratory viz. accountable talk on the other (Mercer, 1995; Michaels et al., 2008). However, international research shows that in practice students have little opportunity to use language for learning in a dialogic way (Burns & Myhill, 2004; Eke & Lee, 2008; English et al., 2002; Flanders, 1970; Hoetker, 1968; Sinclair & Coulthard, 1975).

2.2 Exploratory talk for learning

Considering interaction as a key element in learning, collaborative activities

https://doi.org/10.59302/cj4khk48

like group work provide teachers with opportunities to pursue both curriculum goals and goals to improve students' conversational skills. In their meta-analysis of 124 studies on either the effect of pedagogy on interactive processes or the effects of interactive processes on learning outcomes or both, Janssen et al. (2010) show that many factors influence interactive processes in a positive way. In turn, interactive processes can positively influence learning outcomes. Earlier, Light (1991) found that using language to make plans explicit, to make decisions and to interpret feedback seems to facilitate problem solving and promote understanding. However, Mercer and Littleton (2007) found that the quality of interaction in most group work does little to promote learning, let alone joint problem solving. Without proper preparation students' group talk easily becomes *disputational* (each student wants to be in his or her own right, arguments are ignored or not provided, students interrupt each other) and/or *cumulative* (students do not discuss matters with a critical mind, avoid proper argumentation and tend to choose the easy way to get things done). So, in order to make collaborative activities really work, teachers need to organise group work in such a way that students use exploratory talk for learning (Mercer, 1995). The first goal of teaching *exploratory talk* is to improve the students' reasoning skills and quality of interaction. The second is making students use these skills for problem solving activities and learning from one another.

Exploratory talk is a form of conversation "in which partners engage critically but constructively with each other's ideas. Statements and suggestions are offered for joint consideration. These may be challenged and counter-challenged, but challenges are justified and alternative hypotheses are offered. Compared with the other two types in exploratory talk knowledge is made more publicly accountable and reasoning is more visible in the talk" (Wegerif & Mercer, 1997a, p. 53). The added value of exploratory talk for learning is that "[…] it provides a platform for students to share and discuss ideas, aiding in developing their understanding of the subject matter. […]. It encourages critical thinking, creativity, and the development of problem-solving skills, as individuals engage in deep and meaningful discussions to uncover new insights and solutions." (Main, 2024, p. 1).

Exploratory talk has typical linguistic features many of which make reasoning visible: what- and why-questions, positive feedback such as 'that's a good suggestion' or 'you are right', thought reflecting utterances like 'I agree/disagree because', 'I think', 'Why do you say that…?' and consensus seeking statements such as 'So, can we agree on the fact that …' etc. (Barnes, 1976; Mercer, 1995). Sutherland (2006) relates these statements and questions with 'improved quality of talk' and categories of higher-order thinking such as analysis, evaluation, synthesis and hypothesis (see also Palincsar & Brown, 1986). Exploratory talk allows students to formulate their thoughts and arguments, which also improves their problem solving skills. Acquiring and improving those

https://doi.org/10.59302/cj4khk48

skills at school is especially important, because they have long been considered as essential components of professional learning as well as tackling daily issues in life (Simon, 1980; Voss, 2012).

The notion of exploratory talk was launched and described by Barnes (1976) and defined by Wegerif and Mercer (1997a) quoted above. Over the years, researchers have added elaborations and refinings of the concept, and synonyms have been proposed as well. Notions that share nearly the same characteristics are *exploratory discourse* (Fernandez et al., 2001; Kumpulainen, 1996; Nussbaum, 2005), *collaborative argumentation/reasoning* (Golanics & Nussbaum, 2008), *dialogic talk* (Wegerif, 2013) and *transactive reasoning* (Kruger, 1993). Very close to exploratory talk come *accountable talk* (Michaels et al., 2008) and *critical discussion* (Keefer et al., 2000). Elaborating or refining terms of exploratory talk are *incipient* and *elaborate exploratory talk* (Rojas-Drummond & Mercer, 2003). Sutherland (2006) speaks of *group exploratory talk*, focusing on the equality of participation, a notion further explored by Rajala et al. (2012) who distinguish between *inclusive* and *exclusive exploratory talk.* Nikolaidou (2012) adds *reflective* and *operational talk* to the triad cumulative-disputation-exploratory talk. Overarching terms are *Initiation Discussion Response Feedback* (IDRF; Wegerif & Mercer, 1997a), *co-constructive talk* (Rojas-Drummond et al., 2006), *critical learning* (Riley, 2006), *collaborative argumentation* (Golanics & Nussbaum, 2008) and *dialogic reason* (Dourneen, 2013). All these suggestions emphasise the huge interest of researchers in dialogic education and their quest for conceptual clarification .

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

It is important to know that students are not born with exploratory talking skills, nor do they acquire them automatically at home. Consequently, they need to learn them at school (Mercer, 2010b). Specifically, students have to master and employ seven ground rules (Wegerif & Mercer, 1997a):

1. All relevant information is shared
2. All in the group are encouraged to speak by other group members
3. Reasons are expected
4. Challenges are acceptable
5. The group seeks to reach agreement
6. Alternatives are discussed before a decision is taken
7. The group takes responsibility for decisions

Mercer et al. (1999) explain that the first three rules are to bind the group, share together and construct knowledge together through seeking agreement. Rules four and five are about explicit reasoning and challenging, two characteristics which distinguish exploratory talk from disputational and cumulative talk. The sixth rule goes back to Kruger (1993) who found that groups that do best are those which consider alternatives before deciding. The seventh rule has been developed by Wegerif, Mercer and Dawes (1999), who found that students need to be actively encouraged by their peers to speak and to put forward ideas. Setati et al. (2002) add that learning from talk is significantly limited if it is not supported or complemented by strategies for learning to talk. The ground rules are a framework students can hold on to in order to develop these strategies.

2.3 Measurement and effects of exploratory talk

In order to measure learning effects of exploratory talk in student-student conversations UK and Mexican researchers set up empirical experiments within a project called Thinking Together. Here, experimental groups of students were first taught these ground rules via 'basic lessons', while control groups did regular group work without such preparation (Rojas-Drummond et al., 2006; Rojas-Drummond & Mercer, 2003; Rojas-Drummond & Zapata, 2004; Wegerif et al., 1999; Wegerif & Mercer, 1997a, 1997b). After the basic lessons followed a usually extended period of regular and systematic collaborative activities during which feedback and feed forward reflection stimulated students to master exploratory talk and apply it while making assignments for different school subjects. Pre- and post-testing identified learning effects and often included measurements of the development of problem solving skills.

Mercer et al. (1999) recorded students' conversations before and after the intervention. Qualitative discourse analysis of the students' conversations focused on the use of key words in context (KWIC). These are specific utterances which, used in a proper context, can be labelled as indicators of exploratory talk. An additional count of these indicators was then processed statistically in order to compare students' conversations before and after the intervention. Correlation was then examined between the students' improvement of exploratory talk and their score progression on a pre- and post-reasoning test. Direct observed effects were that students increased their reasoning skills: as students were stimulated to do more explaining of certain beliefs or solutions and use more arguments, their utterances became longer and they used keys words in context more frequently (Mercer et al., 1999). Accordingly, Mercer et al. (1999) found an increase of the students' problem solving abilities during which reasoning skills were required. These effects were observed more than once. Based on a number of similar subsequent experiments Wegerif et al. (2005) suggest that the Thinking Together approach 'reliably leads to gains on reasoning tests of between 5% and 10% for individuals and between 10% and 15% for groups' (Wegerif et al., 2005, p. 43; see also Fernandez et al., 2001; Mercer, 1995; Mercer et al., 1999; Rojas-Drummond et al., 2006; Rojas-Drummond & Mercer, 2003; Webb et al., 2016; Wegerif & Mercer, 1997a, 1997b).

An important question then is to what extent these studies can be replicated in other contexts (in this case: the Flemish). Replication is considered to be 'a key aspect of knowledge building in many fields of research' (Cai et al., 2018, p. 2). Schmidt (2016) distinguishes between conceptual replication and direct

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

replication. In conceptual replication different methods are used to test the underlying hypothesis, while in direct replication the same methods are used as in the original research. Only direct replication, Makel and Plucker (2014) add, can disconfirm or corroborate previous claims. Not only does it contribute to the generalisability of the original findings (Morrison, 2019) and thus accumulates understanding, it can also help to "identify, diagnose and minimise […] methodological biases" (Makel & Plucker, 2014, p. 2). In many research domains the concern over replication exists and has existed for generations (for an overview, see Makel & Plucker, 2014). In their analysis of the publication history of the 2013 top 100 education journals ranked by 5-year impact factor Makel and Plucker (2014) found that 71,4% of all direct replications was successful. However, the success rate was considerably higher (88,7%) when the study was undertaken by the same research team that did the original research and significantly lower (54%) when there was no overlap in authorship between the original and replicating articles. These findings emphasise the importance of third-party direct replications, which is one of the reasons why the current study was undertaken.

2.4 The present study: context and research questions

In Flanders (the Dutch-speaking part of Belgium) and in The Netherlands the concept of exploratory talk was introduced by Van der Aalsvoort and Van der Leeuw (1992), but it was not embraced by Flemish education policy, as did happen in the UK. Even now exploratory talk largely remains uncharted territory in Flemish education. Flemish and Dutch research which focuses on the quality of talk for learning at school has long been scarce. In their reviews Hoogeveen & Bonset (1998), Bonset & Braaksma (2008) and Bonset & Hoogeveen (2011) found only a handful of studies on speaking and listening skills between 1969 and 1997, and as good as none between 1997 and 2008. According to Bonset & Braaksma (2008) this lack of interest has to do with a general lower priority feeling among both teachers and researchers, as well as a number of practical problems which make speaking and listening skills more difficult to teach and assess, while leaving researchers with less opportunities to set up proper studies. Not much seems to have changed since Lammers (1993) expressed his concern that there is no real tradition in teaching oral skills in Flemish and Dutch education, nor is there any coherent framework of knowledge and insights concerning didactic methods and assessment strategies.

Since 2008, though, there is a renewed interest in the study of oral skills, more specifically in dialogic teaching and in active, collaborative learning as a means to co-construct knowledge (Van Gorp, 2010; Van der Veen et al., 2017; Damhuis & Van der Zalm, 2017; Jonkman et al., 2021; De Schrijver et al., 2021; Lippens, 2023). One of the reasons for this is an expanding international research interest: before, roughly speaking, 2005 studies on exploratory talk

https://doi.org/10.59302/cj4khk48

were mainly conducted in the UK and the US, but since then the topic has become the subject of research in 27 countries worldwide (T'Sas & Daems, 2024).

These studies notwithstanding, the concept of exploratory talk is not mentioned in either Flemish and Dutch attainment targets nor in its curricula (T'Sas, 2018; T'Sas & Daems, 2024). In The Netherlands attainment targets on oral skills are described separately for listening, speaking and interaction. Interaction includes conversational skills which enable students to learn from one another. The characteristics of these skills are very similar to the ground rules needed to develop exploratory talk (Prenger & Pleumeekers, 2023). In the Flemish attainment targets no such distinction is made and conversational skills are integrated in targets for speaking and listening skills (Vlaamse_overheid, 2024). Summarising, we assume that British and Dutch education policy put a more explicit emphasis on the development of conversational skills for learning than the Flemish. This might make it more difficult to implement exploratory talk in Flemish schools successfully. In order to clear this out, this replicator study was carried out in the Flemish educational context, addressing the following research questions:

RQ 1. To what extent do Flemish students of the third level (primary school) use exploratory talk in group assignments after a 12 week intervention?

RQ 2. What effects does the use of exploratory talk have on these students' reasoning and problem solving skills at group level?

Conforming the original study, we hypothesised that students of the experimental group would use significantly more key words in context after the intervention. Also, and this we added to the original study more explicitly, the quantity and quality of their arguments would augment (Rojas-Drummond & Mercer, 2003). Consequently, the students' conversations would be more exploratory than conversations which contain either few or none such key words and (high quality) arguments (Mercer, 1995; Mercer et al., 1999; Wegerif & Mercer, 1997a etc.). Both key words in context and (high quality) arguments are indicators for reasoning. As this improves the experimental group is expected to perform significantly better on a problem solving test after the intervention.

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

3 Materials and methods

3.1 Sociocultural research

In a review on methods and methodology Mercer (2010a) discusses relevant methods for analysing classroom talk, making a distinction between linguistic ethnography and sociocultural research and explaining why the latter is preferable for studies of the use of exploratory talk for learning. Sociocultural researchers bear on research traditions in social and developmental psychology and pedagogical studies. They strongly attach to the ideas of Vygotsky (1978) who considers language to be a cultural and psychological tool: knowledge and understanding are created together. Language is used to learn by joint reasoning (intermental) and what is learned is eventually integrated into the individual mind (intramental). Since education is considered to be a dialogic process, the way talk – that is dialogue – is organised in the classroom could have an important influence on students' reasoning and reasoning skills. Therefore sociocultural researchers 'are positively inclined towards the use of pre/post interventional designs, seeking to measure differential effects of talk on problem solving, learning and conceptual change' (Mercer, 2010a, p. 3). While often combining qualitative and quantitative methods, their studies are mostly observational, interventional and/or quasi-experimental. Research questions focus on the occurrence of types of classroom talk and the way these types promote learning and develop understanding. Mercer's 1995 research and - mutatis mutandis - our replication of it are examples of this approach.

3.2 Experimental design and data collection

Based on Mercer's original research (Mercer et al., 1999) a quasi-experiment was executed with pre- and post-testing in 5 Flemish primary schools, followed by qualitative analyses of student discourses as an onset for further quantitative analyses. The choice for primary education is based on the findings of Hart & Risley (1995) according to whom the quantity and quality of speech that children experience in earlier school years, predict very well how they will perform in secondary education.

The participating schools were selected on a voluntary basis from the school database of an Antwerp high school for teacher training. 60 such schools were invited to participate via email. Of these 12 schools expressed their interest in the experiment and were visited by the researchers for a more detailed briefing. Eventually, 5 schools agreed to participate. One school provided a pilot class, while the remaining schools each provided two classes to work with. All schools agreed to a written informed consent in view of data collection and all data were anonymised.

The quasi-experiment involved 11 classes, 11 teachers and, initially, 163 students. As data collection would be intense and complex, a pilot study in one class preceded the main study in the remaining 10 classes. In the pilot class, which counted 18 students, 6 triads were formed by the teacher to take part in all group activities, including pre- and post-tests. Of these 6 triads, 3 triads were followed closely: their conversations were audio and video recorded for further analysis. The main objective of this pilot study was to 'try things out' and make sure data collection would be done properly in the main study.

In the main study 5 control and 5 experimental classes of 11- and 12-year old students were included. Classes were placed in the control group or in

https://doi.org/10.59302/cj4khk48

the experimental group randomly in each school and there was no condition mixture. Within each school the control phase preceded the experimental phase. Further, the teachers involved were only briefed about (their role in) the experiment at the start of the experimental phase. This way we avoided unintended influences among teachers and students.

As in the pilot study, in each class triads were formed by the teacher: 30 in the control classes and 27 in the experiment classes (n=57). Each of these was to participate in the intervention for 14 consecutive weeks. All class teachers were asked to create heterogenous triads considering gender, cognitive ability and language skills, because this method '[…] works to reinforce a student's own learning as well as the learning of his or her fellow group members' (Coffey, 2016, p. 1). All triads remained unchanged for the duration of the experiment. In each class 3 triads (control group n=15; experimental group n=15) were randomly selected to be monitored closely. As in the pilot study their conversations were audio and video recorded throughout the intervention for further analysis.

In each class the intervention included 2 weeks for pre- and post-testing (week 1 and week 14). In the experimental group the first 4 weeks of the intervention were devoted to basic lessons introducing exploratory talk, followed by 8 weeks of twice a week group work. Control groups immediately started doing group work for a total duration of 12 weeks. In all classes, of every group assignment at least ten minutes were devoted to conversation, with minimal interruption by the teacher.

The experimental groups first (week 2-5) learnt the ground rules of exploratory talk via 5 pre-designed basic lessons. For this, we translated and adapted the basic lessons which were used in the original study and which can be found on the website of the Thinking Together Project (https://thinkingtogether.educ.cam.ac.uk). For a more extensive description of the basic lessons, see Appendix A.

The basic lessons were given by the teachers who had previously been introduced to the matter and trained for this specific purpose by the main researcher. To make sure the lesson protocols were respected in the experimental group the researchers took field notes during the lessons which were compared with one another afterwards. After each basic lesson (week 2-5) and in the 8 weeks that followed triads did regular group work for various subjects, including reflective activities about their progress in handling the ground rules. In the control group regular group work was organised from the very beginning (week 2-13), without any reference to exploratory talk or the handling of the ground rules. Flemish teachers have a high degree of freedom as how to construct their lessons and how to organise lesson activities. Much of their teaching is very student-oriented, which makes it impossible and unnatural to expect a lesson in one school to be an exact copy of a lesson in

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

another school. Consequently, the content of the regular group work was mostly determined by themes and topics answering to the regular curriculum, that is mathematics, languages, reading, world orientation (for example Children's Rights, How people build a house, Global warming…) etc. To make sure each regular group work was organised as similarly as possible, each teacher received a specially developed checklist of guidelines for the organisation of group work. For instance, every group assignment had to be problem-based so as to make sure a problem solving conversation would arise. The same checklist was used by both the control group and the experimental group teachers. Also, the teachers' lessons were systematically checked before each lesson and, if necessary, modified by the researchers. As mentioned earlier, regular group work had to be organised twice a week and conversations had to last at least ten minutes without any interruption by the teacher.

Pre- and post-testing was organised in two ways: one was a group discussion about a non-curricular topic ('Facebook is allowed to use my pictures for commercial use'), the other was the Raven's Standard Progressive Matrices test (Raven, 2003) which was taken by all triads. Paraphrasing Mercer et al. (1999) Raven's is commonly used to test the ability to reason and solve problems involving new information. It also correlates with measures of academic attainment. Conversations were transcribed verbatim and stored for qualitative and quantitative analysis. Both the non-curricular discussion and the discussions held during the problem solving test at group level were analyzed. Analysis comprised two indicators of exploratory talk: the use of key words in context and - expanding Mercer's study with analyses in Rojas-Drummond and Zapata (2004) - the use and quality of arguments.

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

3.3 Variables

For our analysis we applied Mercer's Inverted Dynamic Pyramid (Wegerif & Mercer, 1997a). This is a sociocultural discourse analysis which starts on the most abstract level, ends on the most concrete level and then returns to the abstract level. The method combines qualitative and quantitave methods of research. Qualitatively, the way language is used in context is examined in detail by means of discourse analysis. Quantitatively, concordance software is used to analyse transcripts of larger amounts of language use. This way, specific examples of language use can be generalised without any loss of the context in which the language is used.

In the method several levels of abstraction are integrated. Video and sound recordings are the most concrete data. When pulling selected features out of these data by making a transcription, the result is a first level of abstraction. The next level is pulling lists of key words in context out of the transcript. A subsequent level is doing a key word count using software (in our case NVivo 10; Bazeley & Jackson, 2013) and comparing these findings with similar data obtained from other transcripts. Whenever the key word count is finished, it can be determined qualitatively to what extent the key word context is actually contributing to exploratory talk. Based on this analysis, the quantity and quality of arguments can be determined. Iterative movements like these (from concrete to abstract and back to concrete, repeatedly if necessary) explain why the pyramid is called 'dynamic'.

Key words in context only become indicators of exploratory talk when they are used in specific speech acts (Searle, 1969) within a relevant reasoning context, such as confirmation, disagreement, explanation, justification of ideas etc. In turn, these speech acts must be generated by one of the ground rules of exploratory talk. For instance, when a student says 'I believe it's the other way round', 'I believe' is a key word in context because it introduces a disagreement, which in turn answers to ground rule 4 ('Challenges are acceptable'). When the student continues saying 'because this circle overlaps the other', 'because…' is a key word in context because it introduces an argument (that is an explanation of the statement made before), which answers to ground rule 3 ('Reasons are expected').

The quality of arguments was determined as in Rojas-Drummond and Zapata (2004). In that study, four quality levels of arguments are defined, from low to high: rudimentary (A), implicit (B), semi-explicit (C) and explicit (D). These levels of arguments represent phases of transition in the clarity and precision of arguments and highlight quality of talk. For a more extensive description of the four quality levels of arguments, see Appendix B.

In this study, the qualitative analysis of key words in context as well as the use and quality of arguments was done by the main researcher and, for reasons of interrater reliability, checked by two language master students who were not involved in the study. Also, the students were blind to whether they had to analyse control or experimental group transcriptions. This way descriptive and interpretative validity of the study were assured (Johnson, 1997). Both students analysed the transcriptions independently from one another. After that the main researcher compared his own coding results with that of the students. For key word elimination Cohen's kappa was .92 (high degree of agreement). As the quantity of arguments may indicate progress in the use of exploratory talk, determination whether an utterance is an argument or not was done by the same language master students. For this Cohen's kappa was .72 (good degree of agreement). The same protocol was used to determine the quality of arguments (Cohen's kappa = 60; reasonable reliability). Final agreement about both indicators was reached after discussion with the main researcher.

Problem solving skills were measured using Raven's Standard Progressive Matrices (Raven, 2003). Each test was divided into two parts of equal difficulty (30 puzzles): one for pre-testing, the other for post-testing. For both, the maximum attainable score was 30. Results were stored in a statistical dataset.

https://doi.org/10.59302/cj4khk48

3.4 Data analysis

The design of the study is a repeated measures design with two factors. There are two measurements of variables for each of the triads (pre-test and post-test scores), resulting in a within subjects factor. The triads are assigned to an experimental condition (exploratory talk) and a control condition (business as usual), resulting in a between subjects factor.

For the statistical analyses we used a (generalised) linear mixed effects model (Bolker et al., 2009) with the triads as a random effect. In the fixed part of the model we included both factors 'pre-test vs. post-test' and 'control group vs. experimental group' and the interaction term between both.

Not all dependent variables could be modelled with a 'normal' linear mixed effects model. Two dependent variables were log-transformed before the analyses as they are on a percentage scale ('% key words in non-curricular discussion' and '% key words in a problem solving task') and four dependent variables were modelled using a poisson model with a log-link function as they are count data: 'number of arguments in a non-curricular discussion'; 'number of arguments in a problem solving task'; 'number of high-quality arguments in a non-curricular discussion'; 'number of high-quality arguments in a problem solving task' (Fox, 2015). After estimating the models we performed post-hoc tests using a Sidak correction (Abdi, 2007) to test the statistical significance of contrasts. All statistical analyses were done in the open-source statistical language R (Core Team, 2017), making use of the following packages: lme4 (Bates et al., 2014) for the (generalised) linear mixed effects models); lmerTest (Kuznetsova et al., 2017); performance (Lüdecke et al., 2021) to check the model assumptions; emmeans (Lenth et al., 2018) for post-hoc tests; ggplot2 (Hadley, 2016) and sjPlot (Lüdecke et al., 2021) to create the plots.

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

4 Results

We will now discuss the results of our study, focusing on four sets of outcomes: keywords in context (4.2), the quantity (4.3) and quality of arguments (4.4), and students' problem solving skills (4.5). For each we will first describe the main effects and then the interaction effects (4.1). An illustrative qualitative description of a transcribed conversation is included as Appendix C.

4.1 Descriptives

In table 1 descriptive data are provided for all four analyses: the use of key words, the number of arguments and the quality of arguments in both discussions, and the scores for the problem solving test. For each indicator the first cell concerns the average at pre-test for the control group. The second cell reports the average at pre-test for the experimental group. The average for the post-test for the control group is reported in cell 3. The last cell reports the average for the post-test for the experimental group.

Table 1

Descriptive Data for the use of Key Words, the Number of Arguments and the Quality of Arguments in both Discussions, and the Scores at the Problem Solving Test. Mean Scores for the Experimental and the Control Groups at both Pre-Test and Post-Test.

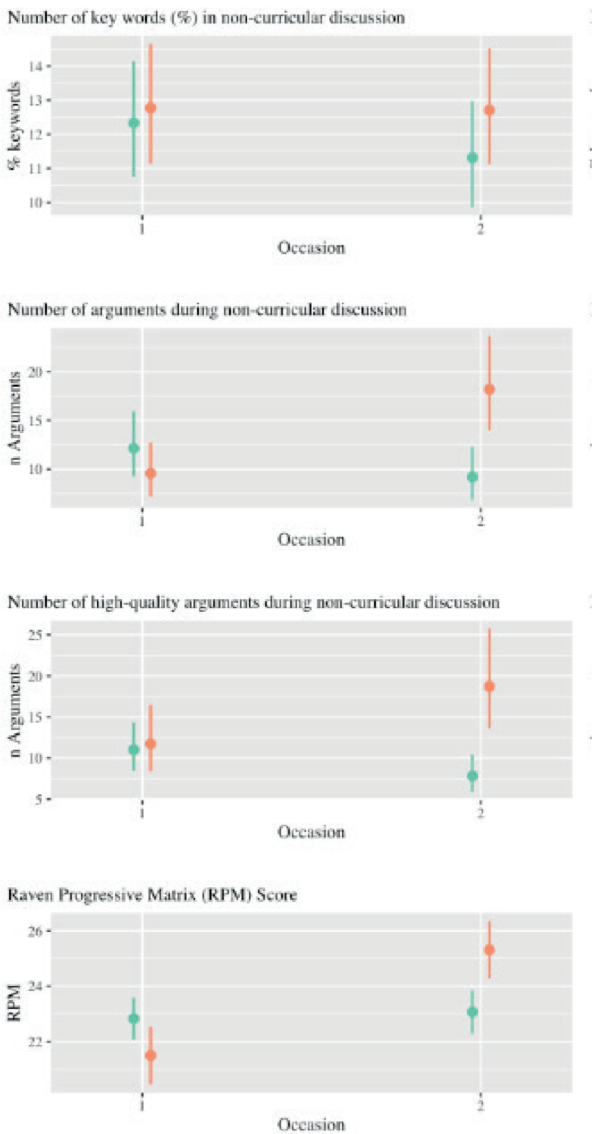
	Control condition		**Experimental condition**	
	Pre-test	Post-test	Pre-test	Post-test
Number of key words (%) in non-curricular discussion	12.4	12.6	11.9	13.7
Number of key words (%) in problem-solving task	12.4	11.1	11.8	14.2
Number of arguments in non-curricular discussion	13.9	13.4	10.8	22
Number of arguments in problem-solving task	30.4	24.2	22.1	42.6
Quality of arguments in non-curricular discussion	12.3	12.3	9	19.7
Number of arguments in problem-solving task	0.933	1.89	1.67	6.5
Scores for the problem solving test (Raven's)	22.8	21.5	23.0	25.3

A visual summary of all statistical results (predicted values and 95% Confidence Intervals based on reported models) is presented in figure 1.

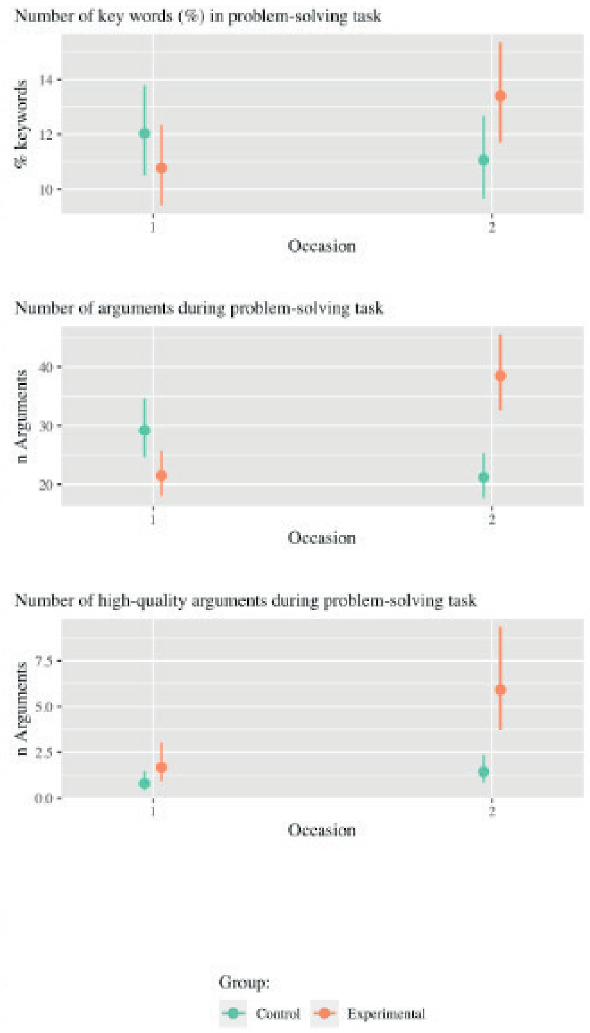
PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Figure 1

Predicted Values and 95% Confidence Intervals based on Reported Models for Percentage of Key Word use (Non-curricular & Problem-solving Discussion), number of Arguments used (Non-curricular & Problem-solving Task), number of High-quality Arguments (coded C or D) used (Non-curricular & Problem-solving tasks) and Scores for Raven's Progressive Matrices.



https://doi.org/10.59302/cj4khk48



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

In the following paragraphs all these results are described in detail.

4.2 Key words in context

Table 2 shows the use of key words in context for both the control and experimental group before and after the intervention, also distinguishing between the non-curricular task and the problem solving task.

Table 2

Parameter Estimates (Est.), 95% Confidence Intervals (C.I.) and P-value based on the Mixed Effects Model for % appearance of Key-Words (log-transformed) in the Non-curricular Task and in the Problem Solving Task.

	Non-curricular task			**Problem solving task**		
	Est.	C.I.	p-value	Est.	C.I.	p-value
Fixed part						
Intercept (*)	2.51	2.37 – 2.65	<0.001	2.49	2.35 – 2.63	<0.001
Occasion (0 = pre-test; 1 = posttest)	-0.09	-0.24 – 0.06	0.255	-0.08	-0.26 – 0.10	0.358
Group (0 = control group; 1 = experimental group)	0.04	-0.16 – 0.23	0.722	-0.11	-0.31 – 0.09	0.267
Occasion * Group	0.08	-0.13 – 0.29	0.443	0.30	0.05 – 0.56	0.022
Random part						
Variance between triads	0.03			0.01		
Residual variance	0.04			0.06		

(*) Given that the model contains both the effect of belonging to the experimental condition as the effect of being an observation of the posttest this intercept can be interpreted as the expected score for observations on the pretest for the control group

Non-curricular discussion

As table 2 shows the main effect of Occasion is not statistically significant ($p = 0.255$). This means that for the control group there is no significant increase or decrease in percentage of key words between the pre-test and the post-test. The main effect of the experimental group is not statistically significant either ($p = 0.722$). This means that for the pre-test there is no difference between the control group and the experimental group regarding the percentage of key words used. Additionally, the interaction term of Occasion with Group is not statistically significant ($p = 0.443$). So, compared with the control group, the difference between pre- and post-test is not different for the experimental group.

https://doi.org/10.59302/cj4khk48

Post-hoc tests (with a Sidak correction) show that for the experimental group there is no significant difference between the pre- and post-test (the difference being -0.086; $p = 0.723$).

Problem solving discussion
Considering the percentage of key words used in the problem solving discussion, Table 2 shows that the main effect of Occasion is not statistically significant ($p = 0.358$). So, the control group did not increase or decrease its percentage of key words during the problem solving task significantly. The main effect of experimental group is not statistically significant either ($p = 0.267$). This means that for the pre-test there is no difference between the control group and the experimental group in percentage of key words used during the problem solving task. The interaction term time between Occasion and Group, however, is statistically significant ($p = 0.022$). This means that the difference between the pre- and the post-test is different for the experimental group as compared with the control group.

Further, post-hoc tests (with a Sidak correction) show that the results of the experimental group show no significant difference between pre- and post-test (the difference is -2.328; $p = 0.101$).

4.3 Number of arguments
Table 3 shows the number of argument used by both the control and experimental group before and after the intervention during the non-curricular task and the problem solving task.

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Table 3

Parameter Estimates (Est.), 95% Confidence Intervals (C.I.) and P-value based on the Mixed Effects Model for the Appearance of Arguments in the Non-curricular (NC) and Problem-solving (RPM) Discussion.

	Non-curricular task			**Problem solving task**		
	Est.	C.I.	p-value	Est.	C.I.	p-value
Fixed part						
Intercept (*)	12.14	[9.22 – 15.99]	<0.001	29.21	[24.63 – 34.63]	<0.001
Occasion (0 = pre-test; 1 = posttest)	0.76	[0.61 – 0.93]	0.009	0.73	[0.63 – 0.84]	<0.001
Group (0 = control group; 1 = experimental group)	0.79	[0.53 – 1.17]	0.241	0.74	[0.58 – 0.94]	0.015
Occasion * Group	2.51	[1.88 – 3.35]	<0.001	2.47	[2.03 – 3.00]	<0.001
Random part						
Variance between triads	0.21			0.08		
Residual variance	0.08			0.04		

(*) Given that the model contains both the effect of belonging to the experimental condition as the effect of being an observation of the posttest this intercept can be interpreted as the expected score for observations on the pretest for the control group

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Non-curricular discussion

Table 3 shows that the main effect of Occasion is statistically significant ($p = 0.009$). The control group triads show a significant decrease in number of arguments during the non-curricular discussion: the parameter estimate is lower than 1, which means that the number of arguments used at the post-test is 0.76 times higher (so actually lower) than the number used at the pre-test. The main effect of the experimental group is not statistically significant ($p = 0.241$). So, for the pre-test there is no difference between the control group and the experimental group considering the number of arguments used. The interaction term of Occasion with Group is statistically significant ($p < 0.001$), which means that the difference between pre- and post-test differs for the experimental group compared with the control group. For the control group the incidence rate is 0.76 times greater (so actually smaller) for the post-test compared with the pre-test. For the experimental group this difference between pre- and post-test is different: the post-test incidence rate is 1.908 (= 0.76 * 2.51) times higher than the pre-test incidence rate.

Post-hoc tests (with a Sidak correction) show a significant difference for the experimental group between pre- and post-test (the difference is 1.900; $p < 0.001$). This means that triads in the experimental group use 1.9 times more arguments during the post-test than during the pre-test.

Problem solving discussion
As table 3 also shows, the main effect of Occasion is statistically significant ($p < 0.001$). In other words, the control group triads show a significant decrease in number of arguments during the problem solving task: the parameter estimate is lower than 1, which means that the number of arguments used during the post-test is 0.73 times higher (so actually lower) than the number used during the pre-test. The main effect of experimental group is statistically significant ($p = 0.015$), so for the pre-test, there is a difference between the control group and the experimental group considering the number of arguments used (the incidence ratio is 0.74 times greater for the experimental group than for the control group). The interaction term of Occasion with Group is statistically significant ($p < 0.001$). So, compared with the control group, the difference between pre- and post-test is different for the experimental group. For the control group the incidence rate is 0.73 times greater (so actually smaller) for the post-test than for the pre-test. For the experimental group this difference between the pre- and post-test is different: the post-test incidence rate is 1.803 (= 0.73 * 2.47) times higher than the pre-test incidence rate.

Post-hoc tests (with a Sidak correction) reveal that the experimental group shows a significant difference between the pre- and the post-test (difference is 1.791; $p < 0.001$). This means that triads in the experimental group use almost 1.8 times more arguments during the post-test than during the pre-test.

4.4 Quality of arguments
Table 4 provides an overview of the quality of arguments used by both the control and experimental group before and after the intervention. Results comprise the non-curricular task and the problem solving task. For statistical reasons, arguments A and B were taken together and presented as lower quality level arguments, while the combination of arguments C and D represents a higher quality level argumentation.

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Table 4

Parameter Estimates (Est.), 95% Confidence Intervals (C.I.) and P-value based on the Mixed Effects Model for the Appearance of Higher and Lower Quality Arguments in the Non-curricular (NC) and Problem-solving Task (RPM).

	Non-curricular task			**Problem solving task**		
	Est.	C.I.	p-value	Est.	C.I.	p-value
Fixed part						
Intercept (*)	11.01	[8.43 – 14.38]	<0.001	0.80	[0.44 – 1.47]	0.477
Occasion (0 = pre-test; 1 = posttest)	0.71	[0.57 – 0.89]	0.003	1.79	[0.93 – 3.44]	0.082
Group (0 = control group; 1 = experimental group)	1.07	[0.69 – 1.64]	0.767	2.10	[0.90 – 4.85]	0.084
Occasion * Group	2.24	[1.62 – 3.11]	<0.001	1.97	[0.83 – 4.67]	0.122
Random part						
Variance between triads	0.18			0.27		
Residual variance	0.08			0.46		

(*) Given that the model contains both the effect of belonging to the experimental condition as the effect of being an observation of the posttest this intercept can be interpreted as the expected score for observations on the pretest for the control group

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Non-curricular discussion

Table 4 tells us that the main effect of Occasion is statistically significant ($p < 0.001$): the control group triads show a significant decrease in number of high-quality arguments (C or D) during the non-curricular discussion: the parameter estimate is lower than 1, which means that the number of arguments used during the post-test is 0.71 times higher (so actually lower) than the number of arguments used during the pre-test.

The main effect of experimental group is not statistically significant ($p = 0.767$). This means that, considering the number of arguments used in the pre-test, there is no difference between the control group and the experimental group (the incidence ratio is 1.07 times greater for the experimental group compared to the control group). The interaction term of Occasion with Experimental_Group is statistically significant ($p < 0.001$), which means that the difference between the pre- and post-test is different for the experimental group compared with the control group. In the control group the incidence rate is 0.71 times greater (so actually smaller) in the post-test than for in pre-test. For the experimental group this difference between the pre- and post-test is different: the post-test incidence rate is 1.590 (= 0.71 * 2.24) times higher than the pre-test incidence rate.

Based on the post-hoc tests (with a Sidak correction) we conclude that the experimental group EG shows a significant difference between the pre- and post-test (the difference being 1.595; $p < 0.001$). The triads in the experimental group use almost 1.6 times more arguments in the post-test than in the pre-test.

Problem solving discussion

Another finding shown in table 4 is that the main effect of Occasion is not statistically significant ($p = 0.082$): the control group triads show no significant increase in number of high-quality arguments (C or D) during the problem solving task. The main effect of Experimental_Group is not statistically significant ($p = 0.084$). So, in the pre-test there is no difference between the control group and the experimental group considering the number of high quality arguments used (the incidence ratio is 2.10 times greater for the experimental group compared with the control group, but this is not statistically significant). The interaction term of Occasion with Group is not statistically significant ($p = 0.122$), which means that the difference between the pre- and post-test is not different for the experimental group compared with the control group.

Based on the post-hoc tests (with a Sidak correction) we conclude that for the experimental group there is a significant difference between the pre- and post-test (difference is 3.52; $p < 0.001$). This means that triads in the experimental group use 3,5 times more high quality arguments during the post-test compared with the pre-test.

https://doi.org/10.59302/cj4khk48

4.5 Problem solving skills at group level

Table 5 shows the scores for the problem solving test of both the control and experimental group before and after the intervention.

Table 5

Parameter Estimates (Est.), 95% Confidence Intervals (C.I.) and P-value based on the Mixed Effects Model for Scores for Raven's Standard Progressive Matrices (RPM) at Triad Level.

	RPM		
	Est.	C.I.	p-value
Fixed part			
Intercept (*)	22.83	[22.06 – 23.61]	<0.001
Occasion (0 = pre-test; 1 = posttest)	0.24	[-0.65 – 1.13]	0.595
Group (0 = control group; 1 = experimental group)	-1.33	[-2.64 – -0.02]	0.046
Occasion * Group	3.57	[2.07 – 5.07]	<0.001
Random part			
Variance between triads	1.58		
Residual variance	2.94		

(*) Given that the model contains both the effect of belonging to the experimental condition as the effect of being an observation of the posttest this intercept can be interpreted as the expected score for observations on the pretest for the control group

https://doi.org/10.59302/cj4khk48

From this model we learn that the main effect of Occasion is not statistically significant (p = 0.595): the control group shows no significant score difference for Raven's on the pre- and post-test. The main effect of the experimental group is statistically significant (p = 0.046). This means that the pre-test scores show a difference between the control group and the experimental group (the experimental group scores are 1.33 points lower). The interaction term of Occasion with Group is statistically significant ($p < 0.001$), which implies that the score difference between the pre- and post-test is different for the experimental group compared with the control group. The control group's score for the post-test is 0.243 higher compared with the pre-test. For the experimental group this difference between the pre- and post-test is different: the post-test score is 3.81 (= 0.24 + 3.57) points higher than the pre-test score.

Based on the post-hoc tests (with a Sidak correction) we draw two conclusions. First, for the post-test, there is a significant difference between both groups (the difference is 2.240; p = 0.005): on the post-test the experimental group scores 2.24 points higher than the control group. Second, the experimental group scores on the pre-test differ significantly from those on the post-test (the difference is 3.812; $p < 0.001$). Triads in the experimental group score 3.8 points higher for Raven's on the post-test than on the pre-test.

5 Discussion

In general, all main effects were in favour of the experimental group and can be explained by the intervention. Students of the experimental group improved their reasoning skills after the intervention, by using more key words in context and more arguments which were also of a higher quality. Also, the average experimental group score on the problem solving test improved significantly when comparing the pre- and post-test. For all variables, the control group made little or no progress at all. Some results need further discussion, though.

Surprisingly, progress of the experimental group was not significant in the non-curricular discussion considering the use of key words in context, whereas in the problem solving discussion it was. A similar outcome is described and explained by Rojas-Drummond et al. (2006, p.92) who found that 'engaging in explicit and accountable reasoning […] was useful for success in joint solving of the reasoning test where the aim was to find the single correct underlying essentially mathematical pattern that united a series of pictures'. The researchers further found that explicit reasoning did not serve more divergent tasks as good as convergent tasks (like the problem solving test in our study), an explanation which is also offered by Tin (2003). As a more open discussion has more divergent features than a problem solving task and consensus is not always required, students may lose the feeling of necessity to use exploratory talk.

Likewise, progress of the experimental group was not significant considering the quality of arguments in the problem-solving discussion, whereas in the non-curricular discussion it was. Here, the explanation is more straight-forward. During the non-curricular discussion the students were not working with learning materials they could refer to (non-verbally). Contrastively, when solving Raven's Progressive Matrices the students were inclined to talk more implicitly as they 'replaced' talk by a mere pointing at the puzzles and using the test sheets to demonstrate what they meant. Hence, even though the quantity of their arguments increased significantly, the quality remained on a less explicit level.

5.1 Limitations and the importance of replication

Based on our findings it is our belief that exploratory talk can be taught and learned in the Flemish educational context with similar positive effects as in the Mercer study. Direct observed effects are that students increase their reasoning and argumentative skills and their problem solving ability at group level. The importance of our findings must also be seen in the broader picture of replication. Mercer's (1995) study revealed exploratory talk to have a considerable learning potential in classroom practice. Consecutive studies have confirmed this in several ways (T'Sas, 2018) and have added extra positive effects to the ones already known, but so far, a direct replication of Mercer's

https://doi.org/10.59302/cj4khk48

study in a different context and without author or research team overlap was lacking. The current study went back to the basics and helps to generalise the results described in the Mercer study. It also helps to overcome some of the limitations mentioned by Wegerif et al. (1999). One of these concerns sample size. Though, in general, the experimental groups made more progress in acquiring exploratory talk, some control triads also did, for certain indicators. Qualitative analyses in our study suggest that some students had already acquired some exploratory conversation skills before the experiment started (see also Rojas-Drummond & Zapata, 2004). By working with 5 experimental classes and 5 control classes we attempted to attain more generalising results than a purely qualitative study would yield, but experiments with a larger study population are still required, as was suggested by Wegerif et al. (1999) when faced with the same phenomenon.

5.2 Next steps

We believe there remains some 'black box' information to be disclosed which now fell outside the scope of this study. Qualitative as well as quantitative analysis may shed more light on the way the mastery of exploratory talk evolves between pre- and post- testing. Additional analysis of video recordings may provide information about the improvement of the 'collaborative atmosphere' and corresponding attitudes intervening aspects such as lesson materials and role shifting in groups. We also believe future research should address the way exploratory talk may help students to cross cultural divides which are rapidly becoming a dominant phenomenon in our classrooms (Geldof, 2013). A number of studies already suggest that the traditional way of teaching, in which dialogue largely remains absent, excludes students from ethnic-cultural minorities as well as low SES students (Reay, 2006; García-Carrión et al., 2020) and marginalises them (Sutherland, 2013). In accordance with Herrlitz-Biro et al. (2013) we also believe that teaching exploratory talk may benefit learners in a more demanding context, like schools with many language weak students. In this respect, Mercer (2010a) noted that significant progress in mastering exploratory talk was seen in schools in deprived areas, with lots of students from a problematic socioeconomic background. More research is necessary to confirm this.

Finally, we believe exploratory talk has the potential to become an integrated part of language-oriented education. Already, there is a lot of scientific evidence to support this. Classroom experiments on exploratory talk have been performed in a wide variety of school subjects, like science (Cervetti et al., 2014; Dawes et al., 2010; Enghag et al., 2007; Haglund & Jeppsson, 2012; Polo et al., 2015), mathematics (Kassoti & Kliapis, 2009; Mercer & Sams, 2006; Murphy, 2015; Rojas-Drummond et al., 2001; Schmitz & Winskel, 2008; Webb, 2015), geography (Bullen et al., 2002), reading comprehension and writing skills (Boyd & Kong, 2017; Brevig, 2006; Maloch, 2002; Soter et al., 2008), music education

https://doi.org/10.59302/cj4khk48

(Nikolaidou, 2012); reasoning skills in specific contexts such as philosophy for children (Topping & Trickey, 2014) etc. Whenever a reasoning test was used to measure cognitive effects, like Raven's Progressive Matrices, those effects are to be considered as cross-curricular, meant to improve reasoning and problem solving.

6 Conclusion

This study reports of the replication in five Flemish primary schools of a quasi-experiment in the UK, described in Mercer et al. (1999), during which students of the experimental group were taught the type of conversational skill called exploratory talk. Pre- and post-tests in experiment and control classes comprised a non-curricular discussion and a problem solving test. Qualitative and quantitative analysis of the students' conversations and correlation with their scores on the test showed significant progress in the experimental group: the students developed better reasoning and argumentative skills and improved their score on the problem solving test.

The reported effects imply that exploratory talk can be taught/learned and if students make use of it consistently during group work, they use language in such a way that they learn from one another, because the quality of their interaction, the exchange of ideas and insights improves: they will use more key words in context which show better reasoning skills and they will use more and better arguments to support any claim they make or to challenge another student's claim. Consequently, group talk becomes more rational and consensus driven. In that respect, the findings of this study confirm earlier research (Mercer, 1996; Mercer & Sams, 2006; Mercer et al., 1999; Rojas-Drummond & Zapata, 2004; Wegerif, 1996; Wegerif & Mercer, 1997b). Direct replication of the Mercer (1995) study confirms that exploratory talk reflects an educationally effective intellectual activity, a social mode of thinking (Mercer, 1996; 2004).

Up till now, despite its learning potential exploratory talk was uncharted territory in Flemish education, leaving a possible quality gap in the teaching of oral skills. Interventional studies like ours help to close this gap.

Sources of support

University of Antwerp and Jan T'Sas (personally) facilitated primary and subsidiary research activities in view of a PhD study. Karel de Grote Hogeschool (Antwerp) funded the intervention and all necessary equipment (50% FTE for 2 years - PWO 1402, cameras and microphones), as well as human resources.

Acknowledgment

Special thanks to Ria Van den Eynde (Karel de Grote Hogeschool) for her substantial contribution to this study.

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics, 3*, 103-107.

Alexander, R. (2008). Culture, dialogue and learning: Notes on an emerging pedagogy. *Exploring talk in school*, 91-114. https://doi.org/10.4135/9781446279526

Asterhan, C. S. (2013). Epistemic and interpersonal dimensions of peer argumentation. *Affective learning together*, 251-271. https://doi.org/10.1111/j.1551-6709.2009.01017.x

Barnes, D. (1976). From communication to curriculum. *Harmondsworth: Penguin Books*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. https://doi.org/10.18637/jss.v067.i01

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo*. Sage Publications Limited.

Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly (1982-)*, 399-410.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135. https://doi.org/10.1016/j.tree.2008.10.008

Bonset, H., & Braaksma, M. (2008). Het schoolvak Nederlands opnieuw onderzocht: Een inventarisatie van onderzoek van 1997 tot en met 2007.

Bonset, H., & Hoogeveen, M. (2011). Mondelinge taalvaardigheid in het basisonderwijs. *Een inventarisatie van empirisch onderzoek*.

Boyd, M., & Kong, Y. (2017). Reasoning Words as Linguistic Features of Exploratory Talk: Classroom Use and What It Can Tell Us. *Discourse Processes, 54*(1), 62-81. https://doi/org/10.1080/0163853X.2015.1095596

Brevig, L. (2006). Engaging in Retrospective Reflection [Reports - Descriptive]. *Reading Teacher, 59*(6), 522-630. http://www.reading.org/publications/journals/rt/v59/i6/

Bruner, J. S. (1960). *The process of education:[a searching discussion of school education opening new paths to learning and teaching]*. Vintage Books.

Bullen, K., Moore, K., & Trollope, J. (2002). The Influence of Pupil-Generated Ground Rules on Collaborative Learning in the Classroom: A Pilot Study [Reports - Research]. *Evaluation and Research in Education, 16*(4), 202-217. http://www.multilingual-matters.net/erie/016/erie0160202.htm

Burns, C., & Myhill, D. (2004). Interactive or inactive? A consideration of the nature of interaction in whole class teaching. *Cambridge Journal of Education, 34*(1), 35-49. https://doi.org/10.1080/0305764042000183115

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018).

https://doi.org/10.59302/cj4khk48

The role of replication studies in educational research. *Journal for Research in Mathematics Education, 49*(1), 2-8. https://doi.org/10.5951/jresematheduc.49.1.0002
Cazden, C. B. (2001). The language of teaching and learning. *The language of teaching and learning*, 348-369.
Cervetti, G. N., DiPardo, A. L., & Staley, S. J. (2014). Entering the Conversation: Exploratory Talk in Middle School Science [Reports - Research]. *Elementary School Journal, 114*(4), 547-572. http://dx.doi.org//10.1086/675638
Coffey, H. (2016). *Cooperative learning*. Retrieved 25 September from **www.learnnc.org/lp/pages/4653#note1**
Core Team, R. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria: URL https://www.R-project.org/.[Google Scholar]*.
Coultas, V. (2012). Classroom talk: Are we listening to teachers' voices? *English in Education, 46*(2), 175-189. https://doi.org/10.1111/j.1754-8845.2012.01125.x
Damhuis, R., & Van der Zalm, E. (2017). A provocative approach for communication with low-proficient children: Examining how the interactive role of teachers and children changes. *Language and Literacy, 19*(2), 51-71. https://doi.org/10.20360/g2837n
Dawes, L., Dore, B., Loxley, P., & Nicholls, L. (2010). A talk focus for promoting enjoyment and developing understanding in science. *English Teaching-Practice and Critique, 9*(2), 99-110.
Dawes, L., Fisher, E., & Mercer, N. (1992). The quality of talk at the computer. *Language and Learning*, 22-22.
De Corte, E. (1996). Actief leren binnen krachtige onderwijsleeromgevingen. *Impuls, 26*(4), 145-156.
De Schrijver, J., Cornelissen, E., Verschoren, V., & Sermeus, J. (2021). Controverse in de wetenschapsklas: een dialogische en cultuursensitieve aanpak om student-leerkrachten voor te bereiden op gevoelige thema's. *Tijdschrift voor lerarenopleiders/Vereniging Lerarenopleiders Nederland; Vereniging Lerarenopleiders Vlaanderen.-Eindhoven, 2007, currens, 42*(4), 95-102.
Dourneen, J. (2013). It ain't what you say, it's the way that you say it: The role of pragmatics in A Level English Language students' learning conversations during collaborative writing using a computer. *English in Education, 47*(1), 33-48. https://doi.org/10.1111/eie.12003
Eke, R., & Lee, J. (2008). *Using talk effectively in the primary classroom*. Routledge.
Enghag, M., Gustafsson, P., & Jonsson, G. (2007). From everyday life experiences to physics understanding occurring in small group work with context rich problems during introductory physics work at university. *Research in Science Education, 37*(4), 449-467. http://dx.doi.org/10.1007/s11165-006-9035-4
English, E., Hargreaves, L., & Hislam, J. (2002). Pedagogical dilemmas in the

National Literacy Strategy: primary teachers' perceptions, reflections and classroom behaviour. *Cambridge Journal of Education, 32*(1), 9-26. https://doi.org/10.1080/03057640220116409

Fernandez, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2002). Re-Conceptualizing "Scaffolding" and the Zone of Proximal Development in the Context of Symmetrical Collaborative Learning [Reports - Research Tests/Questionnaires]. *Journal of Classroom Interaction, 50*(1), 54-72. http://www.jciuh.org/issues/vol50no1.htm

Flanders, N. (1970). *Analyzing Teaching Behavior*. Addison-Wesly Pub. Co.

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

García-Carrión, R., López de Aguileta, G., Padrós, M., & Ramis-Salas, M. (2020). Implications for social impact of dialogic teaching and learning. *Frontiers in psychology, 11*, 140. https://doi.org/10.3389/fpsyg.2020.00140

Geldof, D. S. (2013). Hoe migratie onze samenleving verandert. *Acco, Leuven*.

Golanics, J. D., & Nussbaum, E. M. (2008). Enhancing online collaborative argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning, 24*(3), 167-180. https://doi.org/10.1111/j.1365-2729.2007.00251.x

Hadley, W. (2016). *Ggplot2: Elegrant graphics for data analysis*. Springer.

Haglund, J., & Jeppsson, F. (2012). Using self-generated analogies in teaching of thermodynamics. *Journal of Research in Science Teaching, 49*(7), 898-921. **https://doi.org/10.1002/tea.21025**

Haneda, M., & Wells, G. (2008). Learning an additional language through dialogic inquiry. *Language and Education, 22*(2), 114-136. https://doi.org/10.2167/le730.0

Harris, M. (1995, Jan). Talking in the Middle - Why Writers Need Writing Tutors. *College English, 57*(1), 27-42. **https://doi.org/Doi**/10.2307/378348

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in The Everyday Experience of Young American Children*. Brookes.

Herrlitz-Biro, L., Elbers, E., & de Haan, M. (2013). Key Words and the Analysis of Exploratory Talk [Reports - Research]. *European Journal of Psychology of Education, 28*(4), 1397-1415. http://dx.doi.org/10.1007/s10212-013-0172-7

Hoetker, J. (1968). Teacher questioning behavior in nine junior high school English classes. *Research in the Teaching of English, 2*(2), 99-106.

Hoogeveen, M., & Bonset, H. (1998). Het schoolvak Nederlands onderzocht. *Een inventarisa*.

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational psychology review, 22*(2), 139-154. https://doi.org/10.1007/s10648-010-9131-x

Johnson, R. B. (1997). Examining the validity structure of qualitative research.

Education, 118(2), 282-293.
Kagan, S. (2014). *Coöperatieve leerstrategieën: research, principes en de praktische uitwerking: beknopte uitgave*. Bazalt.
Kassoti, O., & Kliapis, P. (2009). Exploratory Talk for Problem Solving. *Pme 33: Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education, Vol 3, 3*, 289-296.
Keefer, M. W., Zeitz, C. M., & Resnick, L. B. (2000). Judging the quality of peer-led student dialogues. *Cognition and instruction, 18*(1), 53-81. https://doi.org/10.1207/S1532690XCI1801_03
Kruger, A. C. (1993). Peer collaboration: Conflict, cooperation, or both? *Social Development, 2*(3), 165-182. https://doi.org/10.1111/j.1467-9507.1993.tb00012.x
Kumpulainen, K. (1996). The nature of peer interaction in the social context created by the use of word processors. *Learning and Instruction, 6*(3), 243-261. https://doi.org/10.1016/0959-4752(96)00005-9
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*(1), 1-26. https://doi.org/10.18637/jss.v082.i13
Lammers, H. (1993). De impasse in het onderwijs in de mondelinge vaardigheden. *Spiegel, 3*, 9-35.
Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version, 1*(1), 3. https://doi.org/10.32614/cran.package.emmeans
Light, P. (1991). Peers, problem solving, and computers. *Golem, 3*(1), 2-6.
Lippens, N. (2023). *Ontwikkelen en ondersteunen van dialogische en reflectieve interacties binnen professionele leergemeenschappen met student-leraren* Aeres Hogeschool].
Lofgren, R., Schoultz, J., Hultman, G., & Bjorklund, L. (2013). Exploratory Talk in Science Education: Inquiry-Based Learning and Communicative Approach in Primary School. *Journal of Baltic Science Education, 12*(4), 482-496. https://doi.org/10.33225/jbse/13.12.482
Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60). https://doi.org/10.21105/joss.03139
Main, P. (2024, 26 January 2024). Classroom practice: exploratory talk. 1. https://www.structural-learning.com/post/exploratory-talk
Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304-316. https://doi.org/10.3102/0013189x14545513
Maloch, B. (2002, Jan-Mar). Scaffolding student talk: One teacher's role in literature discussion groups. *Reading Research Quarterly, 37*(1), 94-112. https://doi.org/Doi/10.1598/Rrq.37.1.4

https://doi.org/10.59302/cj4khk48

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Multilingual matters.
Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359-377. https://doi.org/10.4324/9780429400759-4
Mercer, N. (2010a). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology, 80*(1), 1-14. https://doi.org/10.1348/000709909X479853
Mercer, N. (2010b, 25 November 2010). Language and education: improving pupils' skills in communication, learning and reasoning. Leerstoel Onderwijsvernieuwing en -samenwerking, Antwerp University.
Mercer, N., & Dawes, L. (2014). The study of talk between teachers and students, from the 1970s until the 2010s. *Oxford Review of Education, 40*(4), 430-445. https://doi.org/10.1080/03054985.2014.934087
Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge.
Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education, 20*(6), 507-528. https://doi.org/10.2167/le678.0
Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's Talk and the Development of Reasoning in the Classroom [Reports - Research]. *British Educational Research Journal, 25*(1), 95-111. https://doi.org/10.1080/0141192990250107

https://doi.org/10.59302/cj4khk48

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in philosophy and education, 27*(4), 283-297. https://doi.org/10.1007/s11217-007-9071-1
Morrison, K. (2019). Realizing the promises of replication studies in education. *Educational Research and Evaluation, 25*(7-8), 412-441. https://doi.org/10.1080/13803611.2020.1838300
Murphy, C. (2015). Changing Teachers' Practices through Exploratory Talk in Mathematics: A Discursive Pedagogical Perspective. *Australian Journal of Teacher Education, 40*(5). http://dx.doi.org/10.14221/ajte.2015v40n5.4
Nikolaidou, G. N. (2012, Feb). ComPLuS model: A new insight in pupils' collaborative talk, actions and balance during a computer-mediated music task. *Computers & Education, 58*(2), 740-765. **https://doi.org/10.1016/j.compedu.2011.09.027**
Nussbaum, E. M. (2005, Jul). The Effect of Goal Instructions and Need for Cognition on Interactive Argumentation. *Contemporary Educational Psychology, 30*(3), 286-313. http://dx.doi.org/10.1016/j.cedpsych.2004.11.002
Palincsar, A. S., & Brown, A. L. (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher, 39*(8), 771-777.
Piaget, J. (1970). Science of education and the psychology of the child. Trans. D.

Coltman.
Polo, C., Lund, K., Plantin, C., & Niccolai, G. (2015). Analyzing Exploratory Talk as a Socio-Cognitive Practice: Identity, Group Argumentation, and Class Debate Quality. Hal-01208319
Rajala, A., Hilppo, J., & Lipponen, L. (2012, 2012). The Emergence of Inclusive Exploratory Talk in Primary Students' Peer Interaction [Reports - Research]. *International Journal of Educational Research, 53*, 55-67. http://dx.doi.org/10.1016/j.ijer.2011.12.011
Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223-237). Springer.
Reay, D. (2006). The zombie stalking English schools: Social class and educational inequality. *British journal of educational studies, 54*(3), 288-307. https://doi.org/10.1111/j.1467-8527.2006.00351.x
Riley, N. R. (2006). Methods for Evaluating Critical Learning Using Online Discussion Forums [Reports - Research]. *Technology, Pedagogy and Education, 15*(1), 63-78. http://www.informaworld.com/openurl?genre=article&id=doi:10.1080/14759390500435770
Rojas-Drummond, S., Mazon, N., Fernandez, M., & Wegerif, R. (2006). Explicit Reasoning, Creativity and Co-Construction in Primary School Children's Collaborative Activities [Reports - Research]. *Thinking Skills and Creativity, 1*(2), 84-94. **http://dx.doi.org/10.1016/j.tsc.2006.06.001**
Rojas-Drummond, S., & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research, 39*(1), 99-111. https://doi.org/10.1016/s0883-0355(03)00075-2
Rojas-Drummond, S., Mercer, N., & Dabrowski, E. (2001). Collaboration, scaffolding and the promotion of problem solving strategies in Mexican pre-schoolers. *European Journal of Psychology of Education, 16*(2), 179-196. <Go to ISI>://WOS:000172436400004
Rojas-Drummond, S., Torreblanca, O., Pedraza, H., Velez, M., & Guzman, K. (2013). 'Dialogic scaffolding': Enhancing learning and understanding in collaborative contexts. *Learning Culture and Social Interaction, 2*(1), 11-21. **https://doi.org/10.1016/j.lcsi.2012.12.003**
Rojas-Drummond, S., & Zapata, M. P. (2004). Exploratory Talk, Argumentation and Reasoning in Mexican Primary School Children [Reports - Evaluative]. *Language and Education, 18*(6), 539-557. http://www.multilingual-matters.net/le/018/6/default.htm
Schmidt, S. (2016). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of general psychology, 13*(2), 90-100. https://doi.org/10.1037/14805-036
Schmitz, M. J., & Winskel, H. (2008). Towards effective partnerships in a collaborative problem-solving task. *British Journal of Educational Psychology, 78*, 581-596. **https://doi.org/10.1348/000709908X281619**

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.
Setati, M., Adler, J., Reed, Y., & Bapoo, A. (2002). Incomplete Journeys: Code-Switching and Other Language Practices in Mathematics, Science, and English Language Classrooms in South Africa. *Language and Education, 16*(2), 128-149. https://doi.org/10.1080/09500780208666824
Simon, H. A. (1980). Cognitive science: The newest science of the artificial. *Cognitive science, 4*(1), 33-46. https://doi.org/10.1207/s15516709cog0401_2
Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford Univ Pr.
Slavin, R. E. (1996). Cooperative learning in middle and secondary schools. *The Clearing House, 69*(4), 200-204.
Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research, 47*(6), 372-391. https://doi.org/10.1016/j.ijer.2009.01.001
Sutherland, J. (2006). Promoting Group Talk and Higher-Order Thinking in Pupils by "Coaching" Secondary English Trainee Teachers [Reports - Research]. *Literacy, 40*(2), 106-114. http://dx.doi.org/10.1111/j.1467-9345.2006.00436.x
Sutherland, J. (2013). Going "Meta": Using a Metadiscoursal Approach to Develop Secondary Students' Dialogic Talk in Small Groups [Reports - Research]. *Research Papers in Education, 30*(1), 44-69. http://dx.doi.org/10.1080/02671522.2013.850528
Tin, T. B. (2003). Does talking with peers help learning? The role of expertise and talk in convergent group discussion tasks. *Journal of English for Academic Purposes, 2*(1), 53-66. https://doi.org/10.1016/S1475-1585(02)00033-4
Topping, K. J., & Trickey, S. (2014). The role role of dialog in philosophy for children. *International Journal of Educational Research, 63*, 69-78. https://doi.org/10.1016/j.ijer.2013.01.002
T'Sas, J. (2013). *Sprekend leren. Wanneer groepswerk echt rendeert in het lager onderwijs.* Brugge: Die Keure.
T'Sas, J. (2018). *Learning outcomes of exploratory talk in collaborative activities.* Doctoral thesis. Universiteit Antwerpen.
T'Sas, J., & Daems, F. (2024). Exploratory talk in times of globalisation and digitalisation: a narrative review. *L1-Educational Studies in Language and Literature, 24*(2), p. 1-38.
Van den Branden, K. (2015). *Onderwijs voor de 21ste eeuw. Een boek voor leerkrachten en ouders*. Acco.
Van den Branden, K., & Van Gorp, K. (2000). How to evaluate CLIM in terms of intercultural education. *Intercultural education, 11*(sup1), 42-51.
Van der Veen, C., Dobber, M., & van Oers, B. (2017). Engaging children in dialogic classroom talk: Does it contribute to a dialogical self? In *The dialogical self*

theory in education: A multicultural perspessctive. Springer Dordrecht, the Netherlands. https://doi.org/10.1007/978-3-319-62861-5_4
Van Gorp, K. (2010). # Meester, wat is genen?# De effectiviteit van kennisconstructie en schooltaalverwerving bij NT2-leerlingen aan het einde van het basisonderwijs.
Vlaamse_overheid. (2024). Onderwijsdoelen. Retrieved on 2 September 2024 from https://www.onderwijs.vlaanderen.be/nl/onderwijspersoneel/van-basis-tot-volwassenenonderwijs/lespraktijk/onderwijsdoelen-en-leerplannen/onderwijsdoelen
Voss, J. F. (2012). Problem solving and the educational process. In *Foundations for a psychology of education,*. 251-294. Routledge. https://doi.org/10.4324/9780203052273-10
Webb, L. (2015). The Use of Cartoons as a Tool to Support Teacher Ownership of Mathematics Curriculum Change. *African Journal of Research in Mathematics Science and Technology Education, 19*(1), 57-68. https://doi.org/10.1080/10288457.2015.1012905
Webb, P., Whitlow, J., & Venter, D. (2016). From Exploratory Talk to Abstract Reasoning: a Case for Far Transfer? *Educational psychology review*, 1-17. https://doi.org/10.1007/s10648-016-9369-z
Wegerif, Linares, J. P., & Rojas-Drummond, S. (2005). Thinking Together in the UK and Mexico: Transfer of an Educational Innovation [Reports - Evaluative]. *Journal of Classroom Interaction, 40*(1), 40-48. http://cmcd.coe.uh.edu/coejci/issues/vol40no1.htm
Wegerif, Mercer, N., & Dawes, L. (1999). From Social Interaction to Individual Reasoning: An Empirical Investigation of a Possible Socio-Cultural Model of Cognitive Development [Reports - Research]. *Learning and Instruction, 9*(6), 493-516. https://doi.org/10.1016/S0959-4752(99)00013-4
Wegerif, R. (1996,). Using computers to help coach exploratory talk across the curriculum. *Computers & Education, 26*(1-3), 51-60. **https://doi.org/Doi** 10.1016/0360-1315(95)00090-9
Wegerif, R., & Mercer, N. (1997a). A dialogical framework for researching peer talk. *Language and Education Library, 12*, 49-64.
Wegerif, R., & Mercer, N. (1997b). Using computer-based text analysis to integrate qualitative and quantitative methods in research on collaborative learning. *Language and Education, 11*(4), 271-286. https://doi.org/10.1080/09500789708666733

PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/cj4khk48

APPENDIX A

In this study, the experimental groups first (week 1-4) learnt the ground rules of exploratory talk via 5 pre-designed basic lessons. These lessons were given by the teachers who had previously been introduced to the matter and trained for this specific purpose by the main researcher. Every basic lesson consists of practical assignments for which students have to work together in groups, of whole-class moments, and of feedback and feed forward reflections on talking and listening skills. In order to guarantee adherence to the intervention protocol all participating teachers were provided with an identical description of the basic lessons, including the appropriate lesson materials.

The basic lessons each have a typical structure consisting of specific lesson targets, success criteria, a description of lesson activities, a list of materials to be prepared by the teacher and self and peer assessment cards for the students. This structure is and remained identical to the ones developed by Mercer et al. (1999), but certain contents and materials were adapted to fit in with the Flemish context. For example, if a UK basic lesson included a short discussion about the British monarchy, this was rather irrelevant to Flemish students and had to be altered.

In each lesson the students are given prompts which stimulate the use of key words in context. For instance, in lesson 1 students learn what talking is, what its function is, how talk is used to exchange information and to ask questions, when talk can be annoying or improper, and how it relates to listening. The success criterium for this lesson is that, at the end, students can explain to one another how and why people talk. During several short activities the students discuss all these issues in triads, using three prompts: ‘listen’, ‘listen’ and ‘talk’. These prompts are distributed within the triads and function as instructions, that is only the student who has the ‘talk’-prompt is allowed to say something, while the other students who have a ‘listen’-prompt, have to listen and are not allowed to interrupt the speaker. In order to make sure every student has the opportunity to talk or listen, prompts are exchanged on a regular basis. After each activity, during a whole-class discussion, the teacher asks questions like: ‘How did the discussion by means of the prompts go?’, ’Which differences did you notice if you compare your discussion with talk on the playground?’, ‘Who was a good listener and how can you tell he was?’ Etcetera. After a final lesson activity the students self and peer assess the conversation within their triad. This gives the teacher ample opportunities to provide their students with feedback. For example, after group work and self/peer assessment the teacher asks: ‘In which group were Why-questions asked?’ If a triad has failed to do so, the teacher asks: ‘Why did no one in your group ask Why-questions?’ and ‘How would you avoid this tomorrow, in your next group assignment?’ Next day, the teacher starts her instructions by addressing the same group: ‘How were you

https://doi.org/10.59302/cj4khk48

going to make sure you asked Why-questions?' etc. Such moments of feedback and feed forward are essential in the students' learning process.

All basic lessons build upon one another. Lesson 2 focuses on listening, turn-taking, interrupting someone, inviting every member of the triad to participate, and asking for clarification. Here prompts like 'What do you think?' are used. Lesson 3 is all about asking for and expressing arguments in order to come to (partial) conclusions. (Prompt: 'Why do you think that?'). Lesson 4 focuses on expressing personal preferences and decision making as a group, based on arguments and counter-arguments, giving each other constructive feedback, and learning how important it can be to work together (Prompts: 'I agree with you, because…', 'I don't agree with you, because…'). Finally, in Lesson 5 a round-up of the previous lesson leads to the formulation of the ground rules for exploratory talk. From then on, this set of rules will be visible in the classroom at all times for reference. It is very important to realise that the basic lessons are cumulative, that is not only do they build on one another, they are regularly being referred to and the prompts of every 'old' lesson are added to those of the new lesson. After the basic lessons, the prompts are not used any longer, but students are free to ask for them at any time, for example whenever they feel they need them.

https://doi.org/10.59302/cj4khk48

APPENDIX B

In this study, the quantity and quality of arguments was analysed as an expansion of the original Mercer et al (1999) study. For this analytical focus support is found in several studies which have demonstrated 'positive effects of discussing ideas upon learning' (Rojas-Drummond & Zapata, 2004, p. 543) and Habermas' claim that 'learning processes by which we acquire knowledge of the world, by which we overcome our difficulties in comprehension and by which we renovate and extend our language are all supported in argumentation' (*ibid*.).

Rojas-Drummond and Zapata (2004) define argumentation as 'the act of providings reasons to make admissible a certain position, opinion or conclusion, or to confront other's positions, opinions or conclusions' (Rojas-Drummond et al., 2004, p. 540). An argument is then described as the combination of an assertion followed by one or more supporting utterances which can be several kinds of reasons. Both researchers worked out these concepts in a complete framework for argument building, which they used and which was also used in this study as a tool to help analyse exploratory talk (table 1).

Table 1

Categories and characteristics of arguments (Rojas-Drummond & Zapata, 2004)

Category	Characteristics	Example
Rudimentary arguments	Signalling and deixis	Wait, yes, it's the one because look, it's like this. No, because of this. It would be this one, because look, it goes like this. Ah, no, this one, yes, because look.
Implicit arguments	All referents are implicit	It's this one, right? Yes, because look, it's in, it's in the middle and this one does not have it. No, because look, it's separating and when it does, it goes together, yes? Together, yes? Yes, then it goes like this and it keeps moving until it touches, and it's on that side at the bottom, it's number 5
Semi-explicit arguments	Congruent but unfi-nished statements. There is at least one non-explicit element (underlined)	No, the square / / the stripped square. There's one circle missing with this one. It's the cross / / the second figure has the dots removed. Let's look at the sequence, here it has like this, they remove the X and the dots. It remains the cross. We assume this figure has the dots removed and here it has this removed, then here an X would follow
Explicit arguments	Congruent and finished statements. All ele-ments are explicit	No, wait / / In the middle figure they remove only the little circles, it doesn't have dots, it would remain just the cross. No, I say it's a white square, a square without the circle. Look, in the figure on the left they remove the circle, in the one at the bottom the diamond, and the square remains.

For reasons of interrater reliability, this analysis was checked by two researchers who were not involved in the study. In cases of doubt, mutual agreement was reached through discussion. For argument recognition Cohen's kappa was 72, which is fairly reliable. Table 1 illustrates how the analysis was done. We labelled the arguments with letters: A (rudimentary argument), B (implicit argument), C (semi-explicit argument) and D (explicit argument).

https://doi.org/10.59302/cj4khk48

APPENDIX C

Table 2 and 3 are illustrative descriptions of two transcripts of the same triad which was observed in the experimental group. The first extract is taken from a conversation which took place during the pre-test, while the students were solving Raven's Progressive Matrices. The second took place during the post-test (idem). Both conversations were translated literally from Dutch to English.

Table 2

Pre-test: problem solving activity (Raven's Progressive Matrices)

Kim:	Leave it. I will take *(points)* that square.
Teresa:	*(points)* No, this one.
Kim:	*(points)* It's this or this.
Teresa:	*(points)* I think it's this.
Jef:	*(points)* No, that one
Teresa:	*(points)* No, it's this.
Jef:	(leaning backwards on his chair) OK, Kim
Teresa:	We take six.
Jef:	Oh come one, Kim, we have to agree, don't we.
Teresa:	It's six, really.
Kim:	OK, leave it. But if it's wrong, then it's your fault.
Jef:	You are in our group, Kim.
Kim:	But I am usually right.
Jef:	Usually. And so am I. In mathematics. This is a kind of mathematics.
Kim:	(pointing at the next puzzle) This, three.
Jef:	That's three alright.

This extract shows a lot of characteristics of disputational talk. Pupils are hardly using any arguments, reasoning is very short, there are no 'what do you think?' questions, let alone why-questions, and no joint agreement is reached. From the start Kim believes he knows the correct answer. As soon as he experiences objections he turns away from the conversation. When Jef tries to draw him back in, Kim only warns him and Teresa that he is not responsible should the answer be wrong. His argument is all about maintaining self-identity (Polo et al., 2015), e.g. he says that he often has the answers right, probably in other contexts. Jef then reacts in the same way, saying he also has many answers right, especially in mathematics (with which he compares the solving of the puzzle). Teresa makes her point in the first half of the conversation and even tries to force a decision, but she withdraws from the conversation as soon as Kim and Jef start convincing one another how good they are. Eventually Kim initiates the next puzzle and proposes a solution (to which Jef immediately agrees). Exploratory

https://doi.org/10.59302/cj4khk48

talk is mostly absent in this conversation.

Table 3

Post-test: problem solving activity (Raven's Progressive Matrices)

Teresa:	I think it's two. Kim?
Kim:	I think it's five.
Teresa:	Why do you think that?
Jef:	Wait, no…
Kim:	Look… *(points)* You have this pattern
Jef:	*(points)* A round square, a square with a cross inside it, a square with small circles. Here's one round, one with a cross going through it, one with circles. So this should be, because … Like that and there is none with a circle, don't you think?
Kim:	But with …
Teresa:	Do you see anywhere… *(points)* like this, wait, take your finger from it. Do you see one like this?
Teresa:	You mean with a different drawing?
Jef:	No, I don't
Kim:	No, I don't
Teresa:	No, so that is not part of it.
Jef:	*(points)* In that case it must be this, don't you think?
Kim:	Yes, yes
Teresa:	So, do we agree?
Jef:	Yes.

This dialogue is more exploratory than the first one. Pupils ask for each other's opinion, ask for arguments and counterchallenge claims. Overall, there is less pointing and more explicit conversation and the group reaches a joint agreement. Teresa initiates the conversation by expressing her solution but she immediately asks if Kim agrees. Kim then suggests another solution. Jef, who was very competitive and quickly bad-tempered in the pre-test still, reacts impulsively but now he does so by formulating an elaborate argument for his claim. Teresa counterchallenges him, not so much by giving a counterargument but by questioning his logic. Then Jef understands and changes his claim himself, to which all agree.

Both fragments show how the use key word in context and the use of arguments can change the type of conversation from disputational to exploratory (and vice versa). In the first transcript key words like 'why', 'because', 'so', 'agree', etc are lacking and consequently, so are (elaborate) arguments. In the second transcript a why-question generates an elaborate argument ('Look…', 'So….', 'Because…', 'Don't you think?') which is also a very good example of a

https://doi.org/10.59302/cj4khk48

long functional utterance. Additional questions (‘Do you see…?’, ‘You mean with …?’) refine this argument and eventually a conclusion is drawn (‘In that case…’) after which the group’s consent is asked (‘So, do we agree?’). All the time, all group members remain involved and turn-taking is more symmetrical.

https://doi.org/10.59302/cj4khk48

Auteurs

Jan T'Sas is gastprofessor aan de Antwerp School of Education, Faculteit Sociale Wetenschappen Universiteit Antwerpen, Antwerpen, België. ORCID 0000-0002-8835-2574.
Sven De Maeyer is als hoogleraar verbonden aan het Department of Training and Education Sciences, Universiteit Antwerpen, Antwerpen, België, ORCID 0000-0003-2888-1631.

Correspondentieadres: Jan T'Sas, Antwerp School of Education, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, België.
E-Mail: jan.tsas@uantwerpen.be

Samenvatting

Leereffecten van exploratieve gesprekken tijdens coöperatief leren

Hoe effectief zijn coöperatieve activiteiten, zoals groepswerk, voor het leren van leerlingen? Van de vele onderzoeken die op deze vraag ingaan, stelt een aanzienlijk aantal één belangrijke voorwaarde voor: het gebruik van exploratieve gesprekken. De voorbije decennia zijn belangrijke Britse experimenten op dit gebied herhaald, verfijnd of uitgewerkt in andere landen. Omdat directe replicatorstudies een belangrijk middel zijn om resultaten te veralgemenen in verschillende contexten, werd een gelijkaardige studie opgezet in Vlaanderen, het Nederlandstalige deel van België. Tijdens een twee jaar durend quasi-experiment werden aan groepen van elf- en twaalfjarige lagereschoolleerlingen de basisprincipes van exploratief spreken bijgebracht en gedurende acht opeenvolgende weken in de klaspraktijk toegepast via groepsopdrachten. Pre- en posttests omvatten het meten van redeneer- en probleemoplossingsvaardigheden. De resultaten bevestigen het oorspronkelijke Britse onderzoek zoals beschreven in Mercer et al. (1999), evenals andere, vergelijkbare studies: leerlingen van de experimentele groep verbeterden hun redeneer- en probleemoplossende vaardigheden op groepsniveau aanzienlijk, terwijl controlegroepen deze vooruitgang niet boekten.

Kernwoorden: exploratieve gesprekken, gespreksvaardigheden, coöperatief leren, dialogisch onderwijs

https://doi.org/10.59302/cj4khk48