Learning and Instruction 63 (2019) 101217

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

# Do those who talk more learn more? The relationship between student classroom talk and student achievement

Klara Sedova*, Martin Sedlacek, Roman Svaricek, Martin Majcik, Jana Navratilova, Anna Drexlerova, Jakub Kychler, Zuzana Salamounova

*Department of Educational Sciences, Faculty of Arts, Masaryk University, Arna Novaka 1, 602 00, Brno, Czech Republic*

A B S T R A C T

There have been efforts to investigate the link between classroom talk and student achievement for some time. However, studies considering individual student participation in classroom talk have thus far been rare. The research reported in this study was carried out on 639 ninth grade students at Czech middle schools. Observations took place in language arts lessons; talk time and the number of utterances with reasoning were recorded for each student. Achievement was measured using a standardized reading literacy test.

The results confirmed a strong link between a given student's talk time and number of utterances featuring reasoning and that student's achievement. As for student talk time, a connection at the classroom level was also identified – students in talkative classrooms had better results. However, there was not a connection between utterances with reasoning and better results at the classroom level. A positive link between individual participation and achievement was observed in all students regardless of socio-economic background or gender.

## 1. Introduction

In recent decades, classroom talk and its contribution to learning have become a key topic in the educational sciences, as well as in philosophy (Kennedy, 2014), psychology (Marková, 2003), sociology (Gurevitch, 2001), and linguistics (Skidmore & Murakami, 2012). There is relatively wide agreement in the scientific community that there is a relationship among talking, thinking, and learning (Resnick, Asterhan, & Clarke, 2017). This assumption has given rise to a range of concepts collectively referred to as dialogue-intensive or talk-intensive pedagogies (Snell & Lefstein, 2018; Wilkinson, Murphy, & Binici, 2015). These concepts share the fundamental premise that good teaching encourages students to actively participate in classroom discourse.

In talk-intensive pedagogies, students speak often, and their utterances are long and elaborate and include reasoning words; the utterances are clearly directed to others; teachers ask open-ended questions that encourage thinking; students do not repeat memorized facts but argue and engage in thoughtful activity; teachers are responsive to students, paying attention to students' thoughts, building on them, and developing them; and students listen to each other, asking questions and having discussions (compare, for example, Alexander, 2006; Lyle, 2008; Lefstein & Snell, 2014).

There have been efforts to collect valid research evidence of any changes in student achievement relating to the quality of classroom dialogue (see van der Veen & van Oers, 2017). However, it has been repeatedly pointed out that more studies of this kind are needed (Howe & Abedin, 2013; Resnick et al., 2017). The aim of this study is therefore to examine, in natural classroom environments of Czech middle schools, the hypothesis that active student participation in classroom talk may be connected with better student achievement.

### 1.1. Theoretical background

The interest in how students participate in classroom talk is grounded in sociocultural theory, especially as it was presented by Vygotsky. Vygotsky (1978) postulated that each psychological function appears twice in a child's development – first on the social level (i.e., as the child interacts with other people) and only later on an individual level (in the form of internalized psychological processes). He believed that speech and thought are closely interlinked and that children can internalize and integrate what they have been able to talk about (Vygotsky, 1978). It follows from this thesis that as more opportunities for communication are created, children's internalization of knowledge could become both faster and of a higher quality. The assumption that communicating and thinking are closely related is widely accepted by the scientific community. Sfard (2008) used the term *commognition*, coined as a blend of communication and cognition, to emphasize the indivisibility of these phenomena. Sfard (2008), with reference to Vygotsky, understood learning as a way of participating in a certain discourse. The Vygotskian perspective implies that language can be used

* Corresponding author.
*E-mail address:* ksedova@phil.muni.cz (K. Sedova).

https://doi.org/10.1016/j.learninstruc.2019.101217
Received 28 May 2018; Received in revised form 23 May 2019; Accepted 8 June 2019

as a tool in service of intellectual development. This notion necessarily raises the question of how to handle talk in schooling (Resnick et al., 2017) – how to implement talk-intensive pedagogies in real classrooms. It is simultaneously necessary to find research evidence that could support the theory on which talk-intensive pedagogies are based. The aim of this study is to provide some such evidence.

### 1.2. Research on the relationship between classroom talk and student outcomes

Generally, there are three kinds of research studies designed to present evidence concerning the relationship between classroom talk and student outcomes: case studies of a particular teacher in a particular class, intervention studies, and correlational large-scale studies. All of them have tremendous potential to extend our knowledge of the studied phenomena.

A number of case studies have been conducted investigating the effects of talk-intensive discourse facilitated by a particular teacher in a particular class (for example Billings & Fitzgerald, 2002; Kutnick & Colwell, 2010; Mercer & Littleton, 2007; Scott, Ametller, Mortimer, & Emberton, 2010; Forman, Ramirez-DelToro, Brown, & Passmore, 2017). These studies have provided valuable insights into the potential of talk and language to intensify learning. Yet, their results cannot simply be taken as a reflection of how teaching in accordance with talk-intensive pedagogies works at school in general, as these studies usually focus on exceptionally competent and motivated teachers.

Intervention studies have been undertaken to implement a change towards talk-intensive teaching and monitor its effects on student learning (see, e.g., Reznitskaya et al., 2001; Andreassen & Bråten, 2011; Kiemer, Gröschner, Pehmer, & Seidel, 2015; O'Connor et al., 2015; Furtak et al., 2016; van der Veen, de Mey, van Kruistum, & van Oers, 2017; Alexander, 2018). It is quite difficult to summarize the findings from these studies as they varied considerably – students were of different ages, interventions were done in different courses, and different outcomes were measured.[1] These studies have indicated that classroom discourse characteristics changed as a result of intervention and student outcomes also subsequently increased (at least partially).

Other studies failed to confirm these links – there were no convincing changes in classroom talk and/or in student outcomes. Murphy, Wilkinson, Soter, and Hennessey (2009) carried out a meta-analysis of 42 intervention studies conducted in literacy classrooms and noted that there was typically an increase in the proportion of student talk and a decrease in the proportion of teacher talk, which is in accordance with talk-intensive pedagogies. However, this change in classroom talk was not always related to improved reading comprehension test results. This implies that findings from intervention studies have not yet provided clear answers to the question of whether talk-intensive pedagogies are connected with better student achievement.

The third type of research conducted to explore the link between classroom talk and student achievement is represented by correlational large-scale studies. These studies have extraordinary potential to advance the understanding of the workings of educational dialogue in authentic classroom situations, as they are able to capture all kinds of naturally occurring forms of classroom talk as well as variability in learning outcomes. We will discuss this type of study in more detail below.

Nystrand, Gamoran, Kachur, and Prendergast (1997) tested more than 1100 eighth and ninth graders in language arts lessons. The students' understanding of literature was tested with a written examination and essay. In addition, classes were observed to determine the time devoted to discussion, authentic questions, uptake, and high-level evaluation. The results of the study were not entirely unambiguous: they identified a positive effect from the observed characteristics on student test scores for the eighth graders but not for the ninth graders.

A similar study was later repeated by Applebee, Langer, Nystrand, and Gamoran (2003). It involved 974 students of different ages in language arts lessons. The same classroom discourse variables were observed as in the previous research. Testing took place through productive written assignments. The analysis indicated a clear positive link between classroom talk features and student performance.

McElhone (2012) carried out a study on 495 students in fourth and fifth grade in reading lessons. Achievement was measured through a reading comprehension test. The study identified high frequency conceptual press—a pattern involving responses to student contributions with requests for clarification and elaboration—to be a key feature of classroom discourse. The study also investigated the opposite pattern of reducing conceptual press, in which the teacher simplifies and narrows the original question, prompts the student, answers the question themselves, or asks for a response from a student other than the one previously called on. The results demonstrated that the pattern of reduced conceptual press was negatively related to results on reading comprehension tests; the results for the high conceptual press pattern were not statistically significant.

Michener (2014) examined a sample of 236 students in third to fifth grade. Teaching was observed and students were given reading comprehension tests. Multilevel modeling indicated a positive impact stemming from teacher explanations (teacher talk longer than two lines of transcript) and, in contrast, a negative impact stemming from elaborated student utterances with explanations (student utterances of more than two lines of transcript that expressed a coherent idea).

Muhonen, Pakarinen, Poikkeus, Lerkkanen, and Rasku-Puttonen (2018) conducted research in language arts and physics/chemistry lessons. The sample consisted of 608 sixth grade students. Researchers recorded video of 158 lessons which were subsequently coded in terms of quality of instructional dialogue (i.e., the degree of engagement in deep and meaningful conversations with clear learning content) and quality of feedback (i.e., the extent of the teacher's extension and expansion of student talk). Grades were used as an indicator of student achievement. The results indicated a positive relationship between the quality of classroom talk and student achievement: the higher the quality of classroom talk was, the higher the grades received by the students in the given class and subject were.

The findings from the research cited above are varied. Some confirm the assumptions of researchers about the positive relation between talk-intensive discourse and student achievement (Applebee et al., 2003; Muhonen et al., 2018), while others are less conclusive (McElhone, 2012; Nystrand et al., 1997) or even contradictory (Michener, 2014). To a certain extent, the variability of the results may be explained by the variation in observed classroom talk characteristics (for example, teacher's authentic questions, open discussion time, conceptual press, student explanation, quality of feedback) as well as by the different ways used to measure student outcomes (e.g., writing assignments, comprehension tests, grades).

However, we understand that there may be another important source of fluctuation in the results of such investigations. The measurement of classroom discourse characteristics in these studies was related to the whole class: for example, the total time of student talk, the total amount of open discussion in a class, or the total number of authentic questions asked by the teacher. These indicators were then linked to the aggregate performance of the whole class on tests. This obscures the distinctions among individual students in the class. There

[1] Alexander (2018) measured the outcomes of 4th grade students in English, mathematics, and science tests; Andreassen and Bråten (2011) evaluated the reading comprehension performances of 5th grade students; Kiemer et al. (2015) perceived autonomy, competence, and intrinsic learning motivation in mathematics results of 9th grade students; Furtak et al. (2016) evaluated achievement in biology of high school students; O'Connor et al. (2015) charted the outcomes of 6th grade students in mathematics tests; Reznitskaya et al. (2001) measured the argumentation of 4th and 5th grade students in persuasive essays; van der Veen et al. (2017) evaluated communicative competence and subject matter knowledge in early childhood education.

are almost certainly differences in their participation in classroom talk and in their test results. Different students, for example, speak for different amounts of time, differ in the degree to which they are engaged in open discussion, or answer different types of questions. The failure to consider the differences among individual students can be viewed as a limitation of the studies cited above. This study is designed to address the existing knowledge gap, as we consider individual student characteristics both in terms of their participation in classroom talk and in terms of their achievement.

### 1.3. Differences among students

Empirical research shows that not all students engage in classroom talk to the same extent. A classic study by Brophy and Good (1970) demonstrated that both the quantity and quality of verbal interactions with the teacher was considerably different for individual students in a class. Similar results have been produced by many more recent studies, for example Black (2004), Jurik, Gröschner, and Seidel (2013), Clarke (2015), and Helgevold (2016). Black (2004) examined different patterns of participation and noted that individual students tended to engage in habitual and recurring ways. Kovalainen and Kumpulainen (2007) reported that there are different types of students in classrooms. Their study describes vocal participants who are very active and take the initiative in all circumstances, responsive participants who are not active in themselves but answer questions from the teacher, and silent participants who rarely take part in communication. The frequency with which students engage in classroom talk and how long they speak are identifiable differences among students. According to Black (2004), students also differ in the level of productivity of their responses. In non-productive participation, student utterances are short, unsurprising, and do not require great cognitive effort; the student's primary goal is to meet the teacher's requirements. In contrast, productive participation describes student interactions that contain verbal actions that appear to create and maintain the shared understandings underpinning the learning process. In such cases, student utterances are elaborate and contain argumentation and reasoning. Black (2004) argues that there are differences among students not only in the frequency of their participation but also in the quality of their contributions.

These findings indicate that there are different participation degrees and quality in classroom discourse. These differences presumably create differing learning opportunities and, ultimately, could lead to differences in student achievement. This assumption was recently verified by Webb et al. (2014) and Ing et al. (2015). They conducted a study in mathematics lessons on 111 students. All student talk during whole-class and small-group discussions was recorded on videos that were then coded according to whether the students referred to their own thoughts or commented on the ideas of others and according to the degree of conceptual elaboration. Student achievement was measured using standardized mathematics tests. The study found a positive link between student participation and achievement.

On the other hand, there were contrasting findings from an experimental study by O'Connor et al. (2017) of two sixth grade classes with 44 students total. Both classes had been previously included in an intervention focused on implementing academically productive talk. A culture of active engagement had been established in both classrooms. In both classes, the same subjects were discussed, but the conditions were manipulated. Each class underwent two teaching units – one unit taught through talk-intensive instruction and the other unit taught through traditional direct instruction. Mathematics tests (pre-tests and post-tests) were used to measure student learning. At the same time, the total number of words uttered during a lesson was counted for each student in both classes. Generally, the analysis indicated that students in the talk-intensive conditions talked many times more and at the same time scored better on the test. But the results did not indicate an association between the number of words spoken by a student and that student's test score in either condition. Hence, the authors claimed that both silent and vocal students benefited from being in a classroom where talk-intensive instruction had been established, and it did not matter whether they verbally participated in the given lesson. This perspective contradicts the findings of Webb et al. (2014) and Ing et al. (2015), who stated that individual participation predicts individual student achievement.

Throughout chapters 1.2. and 1.3., we have referred to studies with differing results and varied underlying assumptions. Nystrand (1996; 2006) argued that classroom discourse shapes student skills through the establishment of classroom epistemology. Basically, Nystrand assumed a homogeneous impact of classroom discourse on all students in the class. O'Connor et al. (2017) agreed with this assumption and added that even students who do not speak in class benefit from talk-intensive instruction. Webb et al. (2014) and Ing et al. (2015) provided a complementary perspective based on the knowledge that different students participate in different ways and that therefore the impact of classroom discourse on different students in a single class may differ. Being aware of these differences in both assumptions and results, we decided to carry out our own study. In doing so, we bore in mind both possibilities: that students may be affected by the discourse established in their classroom and they may also be affected by the scope and quality of their own participation in this discourse. This double perspective constitutes the unique contribution of our study.

## 2. The present study

The aim of this study is to contribute to the understanding of the relationship between student participation in classroom talk and their achievement in the ninth grade. The research was conducted under natural conditions on a relatively large sample (32 classes, 639 students). Unlike previous studies of this type, we distinguished two levels. First, we were interested in determining whether student achievement might be related to the fact that a given student was in a talk-intensive classroom characterized by a high degree and quality of student participation. Second, we examined whether student achievement might be related to the fact that a particular student participated frequently and productively in classroom talk. We used student talk time and the number of student utterances with reasoning made during a lesson as indicators of student participation. We measured student achievement by means of the results from standardized reading literacy tests conducted by the Czech School Inspectorate (CSI).

In line with our research intent, we formulated two main hypotheses. We expected to identify a statistically significant positive relation between student participation and student achievement (H1). At the same time, we hypothesized that a given student's test result would be significantly accounted for by their individual involvement in classroom talk and by the overall character of classroom talk (H2).

## 3. Method

The research design was based on a combination of observational data obtained by the authors and from standardized reading literacy tests conducted by the CSI. The CSI is a key central institution in the evaluation of the education system in the Czech Republic that distributes and evaluates standardized tests focused on different areas of student learning to measure student achievement in Czech schools. Murphy et al. (2009) reported that the use of non-standardized tests designed by researchers is one of the weaknesses of a number of studies investigating the impact of classroom discourse on student achievement. The decision to use tests conducted and evaluated by the CSI in this study addresses this shortcoming and should help to ensure the validity and reliability of the assessment.

### 3.1. Measures

In this study, we measured student participation in classroom talk

and student achievement. We also included socio-economic background and gender as contextual variables since these characteristics are believed to affect student achievement (e.g., Bodovski, Jeon, & Byun, 2017; Pfeffer, 2008; Parker et al., 2018).

#### 3.1.1. Student participation in classroom talk

We operationalized the quantity of student participation in classroom talk as the amount of time a student talked during a lesson. In each class, we observed two lessons. We believe that data from these two lessons represent the nature of classroom talk in the given class and also participation patterns of individual students.[2] Teachers were instructed to teach as they normally would. They did not receive any instruction as to the content of the observed lessons or instructional methods. We therefore assume that both teachers and students behaved in a way similar to their usual conduct. We only counted student utterances that were part of whole-class teaching that involved interactions between the teacher and students and among students. We excluded such types of talk as reading a text, individual work, and group work. We also excluded any talk that did not relate to the subject matter at hand (for example, organizational matters).

We operationalized the quality of student participation in classroom talk as the amount of utterances with reasoning made by a student during a lesson. We based our method on the classification of student utterances proposed by Pimentel and McNeill (2013): (1) no response, (2) word/phrase, (3) complete thought (resembles a sentence but no explanation of thinking is included), and (4) thought and reasoning (resembles a sentence and includes explanation). We counted those utterances that corresponded to the fourth type of this classification.

An example of an utterance with reasoning:

**Teacher:** What does the sentence that the universe has no timeless geography mean? It is quite complex. Could anyone try to explain?

**Student**: It is not really possible to show what the universe is like when it is constantly moving so we cannot even know in a timeless way what shape it will take. When we picture a map, the stars and the galaxies cannot be placed exactly because they are constantly moving and no one knows how.

The quantity and quality of student participation were measured using a tablet application developed by the study authors. The application is based on a class diagram with places marked where each student sits. There were always two observers present in the classroom. The first recorded talk time. The moment a student began to speak, the observer marked his or her place in the diagram and activated the measurement. When the student finished, the observer deactivated the measurement. The second observer recorded the occurrence of utterances with reasoning. As soon as a student made this type of utterance, the observer marked his or her place in the diagram and the utterance was included in the count. While measuring talk time (first observer) is relatively non-problematic, identifying utterances with reasoning (second observer) is a difficult task. The reliability of coding could have been enhanced by the presence of a higher number of coders in the class or by making video recordings. We did not, however, opt for either of these options, in order to minimize disruption to the normal course of teaching. We know from previous field studies (e.g., Sedova et al., 2014; Sedova et al., 2016; Sedova et al., 2017) that teachers perceive the presence of a video camera in a lesson very negatively and there is a significant increase in the likelihood that they will refuse to participate in the research. We anticipated a similar effect would occur if there were to be a larger number of observers in the classroom. We did not want our sample to be significantly affected by negative self-selection on the part of the participants, so we decided to limit the intervention in the normal course of action to the presence of two observers in the classroom. Moreover, any interference with normal conditions raises the risk that the observed actors will change their behavior. By minimizing invasive stimuli, we also sought to reduce the possibility that teachers and students would change their interaction behavior in response to our presence.

In total, there were four pairs of coders performing the observation. All coders were thoroughly trained. Four of them counted exclusively the quantity of student participation and the other four exclusively the quality – the occurrence of utterances with reasoning. These four coders, who carried the greatest responsibility during the observation, are co-authors of this study (the fourth, fifth, sixth, and seventh author). Prior to the start of the research, each coder practiced individually and in a group under supervision (by the first and third author of this study) identification of utterances with reasoning using older video recordings available to the research team. During the training, coders at first practiced recognizing utterances with reasoning in transcripts of lessons. They then watched the video and entered information into the tablet application in real time regarding which student made an utterance with reasoning. This was done individually and then checked for agreement among individual coders. The training was finished when agreement among all coders reached 90%; this took about 43 h per one coder responsible for coding utterances with reasoning. The other four observers were responsible for the talk time measurement. Their training was much shorter. We presented the tablet application to them and instructed them how to work with it. Then they practiced real-time coding of one video recording until we were certain that the differences among observers in talk time measured for individual students were minimal. This took about 9 h per one coder responsible for measurement of talk time.

Observations took place in November and December 2017. Only those students who attended both lessons were included in the analysis. This procedure did not result in a significant data loss (up to 4%). For each student, we considered their average talk time for the two lessons. We also worked with the average number of arguments that each student articulated in the two lessons.[3] For data at the class level, our calculations included total student talk time, which was calculated as the sum of all individual student talk times per one lesson. We also calculated the total number of student utterances with reasoning per one lesson.

#### 3.1.2. Student achievement

We used the results of individual students on reading literacy tests designed and evaluated by the CSI as an indicator of student achievement. In Czech curriculum, developing reading literacy is mostly concentrated in the language arts, which is the subject where we carried out the observation. Testing took place in November 2017. The test consisted of 16 tasks. An example of one of the tasks is attached in Appendix. Students were given several longer extracts from texts written in Czech and the level of their understanding was assessed with several open- and closed-ended questions. The tasks covered all areas of reading literacy that were operationalized in accordance with the curriculum for the grades concerned. In this study, we use the student success rate on the main section of the test (100% is a perfect score) as the indication of student achievement.

[2] We subsequently gathered additional data intended for close qualitative analysis from four classrooms. We conducted six video recordings of lessons in each of these four classrooms. This allowed us to compare values of talk time and number of utterances with an argument for individual students from measurement 1 (two lessons) and measurement 2 (six lessons). The aim was to test whether there were significant differences between the two measurements. We used the Wilcoxon signed ranks T-test and paired samples test. Based on these tests, we confirm the null hypothesis ($p = 0.11$). There was no significant distinction between the mean of the measure 1 and measure 2.

[3] We paid careful attention to the consistency of student seating so as not to cause confusion in the data due to different seating positions of students in the two lessons.

#### 3.1.3. Students' socio-economic background

Students' socio-economic background was measured through the index of the highest occupational status of parents (HISEI). Occupational data for the student's father and mother were obtained through open-ended questions. The responses were coded as four-digit International Standard Classification of Occupations codes and then recoded to the International Socio-Economic Index of Occupational Status (ISEI) (Ganzeboom, De Graaf, & Treiman, 1992). The HISEI for a given student corresponded to the higher ISEI score of the two parents or to the only available parent's ISEI score. Higher ISEI scores indicate higher levels of occupational status.

### 3.2. Sample

The sample for this study consisted of ninth grade students (ISCED 2A) who were involved in the CSI's selective reading literary testing. In this national sample survey, CSI included altogether 163 schools from the total of 4221 Czech lower secondary schools. From this sample, the researchers chose schools from three Czech districts.[4] In total, 23 schools were approached; two of these schools refused to take part in the research. The study included a total of 21 schools and 32 classes (all ninth grade classes). The sample consisted of 639 students. There were approximately 21 students on average in the class; 52% boys and 48% girls. The age of the students at the time of the measurement ranged between 14 and 15 years.

In all of these classes, researchers conducted structured classroom talk observations in two different language arts lessons (2 × 45 min).[5] Teachers were instructed to teach the lessons in their usual way. In total, we recorded data from 64 lessons.

Students completed questionnaires to determine a number of contextual variables, such as socio-economic background, school grades, academic self-esteem, and attitudes toward the subject and the teacher. Table 1 summarizes the descriptive results of all variables relevant to this text.

Students in our sample had average test scores of approximately 44% (Standard Dev. = 15.36).[6] The classroom discourse indicators revealed that they spoke relatively little in the monitored lessons – on average, each student spoke for 11 s in one lesson and made 0.5 utterances with reasoning. In our sample, 46 students spoke for 0 s. Those students had average test scores of approximately 41% (Standard Dev. = 15.12). These general data indicate that the classroom talk in the sample has relatively traditional features – it is dominated by teacher talk, with only a small number of elaborate student utterances containing arguments. This finding is in line with our previous studies carried out in Czech schools (Sedova et al., 2014). However, our results show a great variability in student participation. In almost all classes, some students spoke for significantly longer amounts of time than others and some students remained completely silent. The number of utterances with reasoning for individual students also varied substantially. Of all the students observed, 91% made some utterance. At least one utterance with reasoning was produced by only 42% of the students.

### 3.3. Research ethics

We first made an arrangement to work with the CSI. They agreed that we could use the results of their reading literacy tests in our research if the schools involved and the students' parents agreed. We then sought oral consent from the school principals and all the teachers to allow us to conduct the research in their schools and classrooms. In the next step, we sought written consent from all of the parents of the students participating in the observed classes. Participants were assured of confidentiality and of the ability to withdraw at any time. No one withdrew during the study. All participants were assigned numbers, and any personally identifying information was removed from the data prior to processing.

### 3.4. Analytical strategy

In this study, our main focus was on investigating the effect that student participation in classroom discourse might have on student achievement. We also considered student gender and socio-economic background in the analysis.

We constructed multilevel models. Multilevel models, also referred to as random coefficient models, enable testing whether the constants and slopes of the function symbolizing influence (the regression coefficient) differ in different groups (see Hox, 2002). The purpose of multilevel models is to test this proposition using group characteristics, which in our study were represented by classroom characteristics.[7] The purpose of the analyzed models was to determine, at both the class level and the individual student level, whether classroom discourse might affect student achievement. To do so, we worked in the first step with a zero model. This model enables the variance of the dependent variable (student academic achievement) to be broken down into a component caused by differences among individual students and a component caused by differences among classes. Intra-class correlation (ICC) was determined as the ratio of variance caused by differences among classes to total variability (Soukup, 2006). To test the other hypothesis,[8] we used fixed-slope models[9] that involve factors at both the student and classroom levels; in all cases, we looked for models that contained only factors with an effect on the explained variable at a statistical significance level of 0.05. The equation of the models is as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 A_{ij} + \beta_3 C_i + \beta_4 (X_i * X_j) + \beta_5 (X_i * C_i) + \beta_6 (X_i * A_i) + u_{0j} + e_{ij}$$

where Y represents the academic achievement level of student *i* in class *j*, $X_{ij}$ is the amount of talk time by student *i* and the total talk time for all students in class *j*,[10] $A_{ij}$ identifies the socio-economic status of student *i* and the group status of class *j*, and $C_i$ represents the gender of student *i*. It follows from the equation that we intend to estimate any impact for the variables of talk time and student socio-economic status at both possible levels, i.e. at the individual and whole-class levels. Furthermore, the model indicates that we also intended to estimate the impact of three interaction effects (the interaction of the talk time for a student and for the whole class, the interaction between time and gender, and the interaction between time and socio-economic status). Parameters *u* and *e* are random errors at the class ($u_{0j}$) and student level ($e_{ij}$).

[4] The Czech Republic is divided into 14 districts. The data were collected in three districts (South Moravian Region, Olomouc Region, and Vysocina Region). We addressed all schools in these three districts that were involved in CSI testing of reading literacy.

[5] There were one or two weeks between the two observations.

[6] This seems to be a high standard deviation; however, we should note that similar results occurred with Czech students in an international assessment (see Blažek and Příhodová, 2017).

[7] Another advantage of multilevel models is that the nested structure of the data can be taken into account.

[8] We expected to identify a statistically significant positive effect from student participation on student achievement. At the same time, we hypothesized that a given student's test result would be significantly accounted for by both their individual involvement in classroom talk and the overall character of classroom talk.

[9] The randomness of the first-level variables in all cases proved to be statistically insignificant.

[10] Our second variable conceptualizing classroom talk, i.e. the number of utterances with an argument, can be substituted for X.

**Table 1**
Descriptive indicators for the variables tracked.

| Variable | Mean | Stand. deviat. | Median | Min | Max | Percentiles 25 | Percentiles 75 | N |
|---|---|---|---|---|---|---|---|---|
| Reading literacy (success in %) | 43.86 | 15.36 | 43.47 | 6.52 | 82.61 | 32.61 | 54.35 | 602 |
| *Students* | | | | | | | | |
| Time (s) | 11.45 | 18.02 | 5.35 | 0 | 180.06 | 1.74 | 13.90 | 534 |
| Boy time (s) | 10.29 | 14.66 | 4.89 | 0 | 98.50 | 1.68 | 12.04 | 274 |
| Girl time (s) | 12.44 | 20.46 | 5.76 | 0 | 180.06 | 1.74 | 15.66 | 257 |
| Reason[a] | 0.57 | 1.31 | 0.01 | 0 | 15.50 | 0 | 0.50 | 535 |
| Boy reason[a] | 0.51 | 1.29 | 0.01 | 0 | 15.50 | 0 | 0.50 | 278 |
| Girl reason[a] | 0.63 | 1.34 | 0.01 | 0 | 13.00 | 0 | 0.50 | 254 |
| HISEI | 47.52 | 18.56 | 45.00 | 12.00 | 88.00 | 34.00 | 54.00 | 582 |
| *Class* | | | | | | | | |
| Total time (s) | 189.55 | | | | | | | 32 |
| Total reason (sum)[a] | 9.67 | | | | | | | 32 |
| HISEI | 47.59 | | | | | | | 32 |

[a] Amount of utterances with reasoning made by a student during a lesson.

## 4. Results

We used the following strategy to construct the regression models. First, we focused on the aforementioned zero model. We then gradually added individual independent variables. For each predictor,[11] we tested both options as random or fixed components. Due to space constraints, here we only present the final results. The introduction of each variable in the form of a random effect was shown to be insignificant for all of the variables considered at the student level. This phase resulted in the elimination of any factors not involved in explaining the variability in students. In the last step, we estimated the effect of some interactions between significant predictors. We present only significant interactions.

Table 2 summarizes the parameters of the main estimated models. Because we hypothesized that differences among classes might play a role in explaining variability in student achievement, we first estimated the zero model. Using this model, the average success rate on the reading literacy test when eliminating all factors was estimated at approximately 43%. Intra-class correlation (ICC) equaled 0.18. This result can be interpreted as meaning that 18% of the differences in student success rate on the test could be attributed to the class they come from.

In the next step, we introduced the average time a student actively spoke during a lesson (measured in seconds). The model estimated that this was a significant predictor ($p < 0.01$). The calculated model parameters also indicated that talk time had a positive impact on student achievement. Its strength can be interpreted as meaning that an increase in average talk time of approximately 100 s (1.7 min) correlates with an increase in test scores of 13.7 percentage points. This is undoubtedly a high figure if we take into account the average score in the test (43.7%). It seems to us that this confirms our main hypothesis. We also assessed how to fit the estimated model to the data. We used the Bayesian information criterion (BIC), which compares the estimated model with a zero or saturated model. Raftery (1995) suggested the following key for the evaluation of the credibility of compared models: a 0–2 point decrease means the evidence was weak, 2–6 positive, 6–10 strong, and more than 10 very strong. A comparison of Model 1 with the zero model reveals a difference (decrease) of more than 800 points. At the student level, it seems that talk time was therefore a highly significant predictor of student achievement.

In Model 2, we added information about the student's socio-economic background as indicated by the occupational status of the family (HISEI). This model better matched the obtained data (there was a decline in the BIC). The impact of talk time at the student level remained virtually the same. At the same time, the HISEI also seemed to be a significant predictor. Here we can see a positive link. Since the HISEI is centralized, the calculated estimate can be interpreted as follows: an increase in parent status by one unit appeared to correspond to an average increase in test scores of nearly 2 percentage points.

The next explanatory variable at the first (student) level added to the model was student gender (Model 3). Although the decrease in BIC was not great (13 points), this model fits the data better. The prediction strength of talk time and family status replicated the estimates from

**Table 2**
Results from multilevel models of the impact of talk time on academic achievement.

| | | Estimate | Std. Error | df | T | *p* |
|---|---|---|---|---|---|---|
| Model 0 | Intercept | 43.758 | 1.258 | 28.090 | 34.773 | 0.000 |
| | BIC[12] | 4963.224 | | | | |
| | ICC[13] | 18% | | | | |
| Model 1 | Intercept | 42.441 | 1.332 | 34.077 | 31.868 | 0.000 |
| | Time | 0.137 | 0.038 | 495.135 | 3.576 | 0.000 |
| | BIC | 4161.216 | | | | |
| | ICC | 14% | | | | |
| Model 2 | Intercept | 42.856 | 1.276 | 33.191 | 33.593 | 0.000 |
| | Time | 0.135 | 0.038 | 456.203 | 3.529 | 0.000 |
| | HISEI | 1.721 | 0.744 | 467.652 | 2.314 | 0.021 |
| | BIC | 3883.940 | | | | |
| | ICC | 13% | | | | |
| Model 3 | Intercept | 44.334 | 1.452 | 56.646 | 30.523 | 0.000 |
| | Time | 0.132 | 0.039 | 450.680 | 3.372 | 0.001 |
| | HISEI | 1.822 | 0.751 | 465.377 | 2.427 | 0.016 |
| | Boy | −2.722 | 1.330 | 452.624 | −2.047 | 0.041 |
| | BIC | 3870.053 | | | | |
| | ICC | 12% | | | | |
| Model 4 | Intercept | 39.080 | 8.940 | 29.115 | 4.371 | 0.000 |
| | Time | 0.112 | 0.041 | 453.371 | 2.724 | 0.007 |
| | HISEI | 1.940 | 0.778 | 443.375 | 2.494 | 0.013 |
| | Boy | −2.844 | 1.332 | 450.207 | −2.135 | 0.033 |
| | Time_Class | 0.018 | 0.011 | 31.939 | 1.571 | 0.106 |
| | HISEI_Class | 0.044 | 0.165 | 29.953 | 0.266 | 0.792 |
| | BIC | 3876.483 | | | | |
| | ICC | 12% | | | | |
| Model 5 | Intercept | 38.277 | 2.793 | 49.667 | 13.704 | 0.000 |
| | Time | 0.323 | 0.104 | 416.903 | 3.103 | 0.002 |
| | HISEI | 2.036 | 0.752 | 463.968 | 2.707 | 0.007 |
| | Boy | −3.584 | 1.576 | 449.838 | −2.275 | 0.023 |
| | Time_Class | 0.032 | 0.012 | 49.860 | 2.594 | 0.012 |
| | Time*Time_Class | −0.002 | 0.000 | 377.551 | −2.604 | 0.010 |
| | Time*Boy | 0.061 | 0.077 | 463.434 | 0.789 | 0.430 |
| | Time*HISEI | 0.018 | 0.034 | 462.966 | 0.539 | 0.590 |
| | BIC | 3848.742 | | | | |
| | ICC | 11% | | | | |

[11] With the exception of gender, where it was not appropriate given the exhaustive categories.

[12] Bayesian Information Criterion.

[13] Intra-class correlation.

**Table 3**
Results from multilevel models of the impact of the number of arguments on academic achievement.

| | | Estimate | Std. Error | df | T | *p* |
|---|---|---|---|---|---|---|
| Model 0 | Intercept | 43.758 | 1.258 | 28.090 | 34.773 | 0.000 |
| | BIC | 4963.224 | | | | |
| | ICC | 16% | | | | |
| Model 1 | Intercept | 43.197 | 1.300 | 29.898 | 33.238 | 0.000 |
| | Reason | 1.598 | 0.574 | 497.457 | 2.785 | 0.006 |
| | BIC | 4198.613 | | | | |
| | ICC | 15% | | | | |
| Model 2 | Intercept | 43.527 | 1.256 | 29.430 | 34.665 | 0.000 |
| | Reason | 1.645 | 0.577 | 462.477 | 2.853 | 0.005 |
| | HISEI | 1.732 | 0.741 | 473.500 | 2.337 | 0.020 |
| | BIC | 3927.571 | | | | |
| | ICC | 13% | | | | |
| Model 3 | Intercept | 45.396 | 1.415 | 51.255 | 32.089 | 0.000 |
| | Reason | 1.579 | 0.574 | 457.337 | 2.751 | 0.006 |
| | HISEI | 1.820 | 0.747 | 470.682 | 2.436 | 0.015 |
| | Boy | −3.475 | 1.335 | 458.749 | −2.603 | 0.010 |
| | BIC | 3911.072 | | | | |
| | ICC | 12% | | | | |
| Model 4 | Intercept | 50.582 | 7.759 | 30.624 | 6.519 | 0.000 |
| | Reason | 1.474 | 0.607 | 449.272 | 2.429 | 0.016 |
| | HISEI | 1.989 | 0.775 | 445.111 | 2.567 | 0.011 |
| | Boy | −3.443 | 1.337 | 456.508 | −2.575 | 0.010 |
| | Reason_Class | 0.051 | 0.136 | 35.897 | 0.374 | 0.710 |
| | HISEI_Class | −0.120 | 0.157 | 30.443 | −0.762 | 0.452 |
| | BIC | 3914.326 | | | | |
| | ICC | 12% | | | | |
| Model 5 | Intercept | 49.857 | 8.050 | 31.455 | 6.193 | 0.000 |
| | Reason | 3.245 | 1.221 | 466.511 | 2.658 | 0.008 |
| | HISEI | 1.564 | 0.854 | 447.160 | 1.830 | 0.068 |
| | Boy | −3.803 | 1.520 | 454.183 | −2.502 | 0.013 |
| | Reason_Class | 0.128 | 0.145 | 38.898 | 0.881 | 0.384 |
| | Reason*Reason_Class | −0.080 | 0.040 | 463.808 | −2.031 | 0.043 |
| | Reason*Boy | 0.986 | 1.430 | 449.184 | 0.689 | 0.491 |
| | Reason*HISEI | 0.748 | 0.712 | 460.769 | 1.050 | 0.294 |
| | BIC | 3911.327 | | | | |
| | ICC | 12% | | | | |

previous models, and there were also significant differences in success rate between boys and girls. Boys' scores on the reading literacy tests were an average of nearly 3 percentage points lower than girls' scores.

In the subsequent models, we included class-level variables. In Model 4, we estimated test success relative to the talk characteristics of the entire class. This second level is represented by the total amount of time that students in the class actively participated in talk. We also introduced the average occupational status of the parents of the entire class. The estimated model indicated that aggregate talk time was also a significant predictor of student achievement. The significance of this variable was 0.1, which is higher than what is common in the social sciences (0.05). In this case, however, it is necessary to take into account that multilevel models require a relatively high number of groups (ideally 100 and more). Our number of classes did not meet this requirement. Hox (2002) showed that in some situations multilevel models can be estimated with even fewer groups (even less than 10). He noted that in such cases the fact that the result underestimates the standard error needs to be taken into account. Therefore, he recommended avoiding working with standard statistical significance, i.e. an alpha level of 0.05. If the number of groups is around 20 to 30, the critical threshold should be higher (0.09–0.1). Such is the case for Model 4. Taking these criteria into account, the hypothesis predicting a significant influence from student talk time at class level was confirmed. The second aggregate variable—the average occupational status of the class—turned out to be statistically insignificant. It therefore did not contribute to explaining the variance in academic achievement. This is a surprising finding to some extent. Parental occupational status was a significant predictor at the individual level. Data from the Programme for International Student Assessment have repeatedly demonstrated in the Czech context that socio-economic status plays an important role at not only the individual level but also the school level (Straková, 2016). Our data at the class level did not confirm this finding. The relatively small number of classes for the second-level models could have been an influencing factor. Looking at the descriptive data, however, it seems to us that the relative similarity of the classes included in our sample in terms of average occupational status is a more prominent reason. In our model, this was not a characteristic that explained student achievement on the literacy test. Overall, it appears that Model 4 did not provide a more accurate estimate than the previous model (the BIC slightly increased). We therefore eliminated the aggregate occupational status in the steps that followed.

The last model presented, Model 5, estimated student achievement with the contribution of selected interactions between several variables that had proved in previous models to be significant at the first or second level. The information criteria indicated that this estimate best suited the data obtained. New slopes (regression coefficients) were estimated for the inclusion of all variables. The strength of the influence of individual talk time increased substantially. The new estimate indicated that the average test success rate was about 30 percentage points higher for the individual students who spoke for about 100 s in a lesson. The significance of the impact of total talk time in a class was therefore confirmed; the significance of this predictor was now at the usual level ($p < 0.05$). The predictors of student gender and occupational status remained at approximately the level of Model 3. Interesting conclusions also arose from the tested interactions. The first tested interaction, focused on the relationship between the talk time of a particular student and the total time of all students in the class, was significant. Although the influence is negligible, it indicates a negative effect of the interaction in question. This means that the positive link between a student's result in the test and their overall speaking time is

weaker in more talkative classes (i.e., in classes that are above the sample mean). In other words, a student from a very talkative class must also be very talkative for the individual-time effect described above to apply. It is nonetheless important to repeat that despite being statistically significant, this effect is negligible. A segment of individual student talking time that amounts to 10 s may be connected to a 3.2% increase in the student's academic achievement. The same amount of individual student talking time may be connected to a 3% increase in more talkative classes. The remaining interactions for the influence of talk time were insignificant. This indicates that the talk time was neither strengthened nor weakened in any group of student characteristics. This is again a remarkable finding, and it appears to confirm our assumptions. The impact of individual talk time on student achievement was the same for girls and boys and especially for groups of students with different socio-economic backgrounds.

We used the same analytical procedure in the next step, in which we changed the key explanatory variable. Talk time was replaced by the number of utterances with reasoning in the lesson. We included in the estimates the number of utterances on both the individual and whole-class levels (total utterances with reasoning made by all students in a lesson). The other predictors were identical. Table 3 presents the results of these estimates.

A comparison of the results (Tables 2 and 3) indicates that the two factors of classroom discourse were very similar in terms of their impact on student academic achievement. In other words, the number of utterances with reasoning made by a student during a lesson appears to be a significant explanatory factor for their reading literacy test results. When a student articulated on average an extra 1 argument during a lesson, the average test success rate was about 1.5 percentage points higher (Model 1). When all predictors were included, there was an increase in the impact by as much as 3 percentage points (Model 5). The control variables (gender and occupational status) were used in essentially the same way as in the models based on talk time (see models 1 to 5).

There was a substantial change in the effect from the number of utterances with reasoning at the whole-class level. It turned out that this was not a significant influence (Model 4). The interaction between utterances at the first and second levels was statistically significant (Model 5) but materially negligible.[14] Our data in this respect supported the conclusion that the quality of talk was a significant predictor of student achievement in reading literacy; its impact seemed to be connected to the individual performance of each student. At the whole-class level, the students' talk did not contribute to explaining the variance in student performance in the dependent variable of reading literacy.

## 5. Discussion, limitations, and implications

The aim of this study was to test, under natural conditions in ordinary classes, the existence of a connection between the frequency and quality of student talk during a lesson and student achievement on a reading literacy test. Our results indicated that this relationship existed and was strong. Students who talked and argued more in language arts lessons had better performance results on reading literacy tests.

At present, the central role of talk in relation to learning and knowledge construction in the classroom has come to the forefront of scholarly attention. It is assumed that students learn through talk and their learning outcomes can be largely attributed to the quality of classroom discourse (for example Mercer & Littleton, 2007; Resnick et al., 2017). These hypotheses are widely shared, but robust empirical evidence obtained in whole-class interactions is still scarce. This study can be therefore viewed as a contribution toward supporting the stance that talk-intensive pedagogies are based on correct premises.

This study contributes to the theory development in the field with the notion that it is useful to pay attention to general characteristics of classroom talk and also to individual participation patterns of particular students. Unlike previous studies of a similar type, we examined the relationship between participation and achievement in parallel at the individual and class levels. In other words, we asked whether it is sufficient to place a student in a talkative classroom or if they have to actively speak, express their ideas, and reason. The assumption of the influence of the nature of student talk at the class level is based on the notion of a learning community in which students work together as a group to acquire new knowledge (Lyle, 2008), while the capacities of more advanced students are available to other members of the group who may observe, try out, and gradually internalize new ways of communication and thinking (Reznitskaya & Gregory, 2013). It has been suggested that students benefit from listening to their actively participating classmates (O'Connor et al., 2017). Our findings support this idea only to a limited extent. We found that students had better test results when they were in a class with an overall high amount of student talk. At the same time, there appeared to be a stronger link between individual student participation and their individual achievement. As for utterances with reasoning, it is not possible to say that to be located in a classroom with a high frequency of such utterances is itself linked to better student results. It does seem true that there is a link at the individual level between the number of utterances with reasoning spoken by a particular student and that student's achievements on a reading literacy test.

Our findings are thus more in line with the perspective, previously outlined by the studies of Webb et al. (2014) and Ing et al. (2015), that classroom discourse affects different students in a class in different ways, depending on how actively they participate. Our analysis indicated that those individuals who spoke and argued in class had better results on a reading test. In theory, this finding can be explained through Vygotsky's concept of internalization (Vygotsky, 1978). Talk is essential in this process; the more opportunities for talk there are, the faster and better the internalization of the newly learned knowledge is. If we imagine the classroom as a community of practice (Lave & Wenger, 1991), it is possible to distinguish between students who participate centrally (talk and argue) and those who participate peripherally (listen to the talk and argumentation of their peers). Our findings may indicate that peripheral participation, witnessing the activities of more skillful and experienced members of the community, is not connected with learning outcomes comparable to the outcomes for those who participate centrally.

It is possible that the differences between our finding that vocal students have better learning outcomes than silent students and the findings of O'Connor et al. (2017) that silent students have the same outcomes as vocal students may be, to some degree, explained with reference to different study designs. O'Connor et al. (2017) collected data in classes in which students underwent an intervention focused on the implementation of talk-intensive instruction and had been socialized for a long time in a culture of active participation, including active listening. The authors concluded that even the students who did not speak in the observed lessons were engaged and therefore were able to benefit from the whole-class discussion. This was probably not the case for the classrooms in our sample. Our data indicate that teaching in these classrooms was rather traditional, with a small proportion of student talk. Consequently, our study does not contradict the study by O'Connor et al. (2017). It was implemented in different conditions and shows the relationship between participation in classroom talk and student achievement in an environment that does not appear to be dialogue-rich and does not seem to have an established culture of active participation.

If there is a connection between how students talk in classes and their learning outcomes, it seems to make sense to give students room

[14] This is an inverse relationship; the weakening of the significance of individual arguments along with an increase in the sum of these arguments in a class cannot be considered materially relevant with regard to the scale.

for elaborate talk and also to strive to ensure that their talk is of good quality and contains arguments. This is an important message for schools and teachers. Teaching in Czech schools is still very traditional and not much student-oriented (Sedova et al., 2014). The results from this study may help persuade Czech teachers and their educators that a change toward talk-intensive pedagogies could be desirable as it might help improve student learning. Training in the use of dialogic procedures is not yet a common part of Czech teacher education. This study provides some evidence for their inclusion in teacher education or professional development. A number of studies have reported that students participate unevenly in class (for example Jurik et al., 2013; Clarke, 2015; Helgevold, 2016). In light of our findings, we can anticipate that such a situation creates unequal learning opportunities in terms of the potential link between participation in classroom talk and student learning gains. We believe that there will always be students who will talk more, because they know more about the topic (Clarke, 2015) or are more extroverted (Young, 2014) or more motivated (Jurik et al., 2013). Despite this, we think that our analysis implies a clear recommendation for teachers – to be aware of the possible link between classroom talk and achievement and to strive to invite all students in the classroom to participate in classroom conversations. However, teachers often perceive this task as very difficult (Snell & Lefstein, 2018). In the future, it may therefore be necessary to devote energy to a focused examination of the circumstances in which teachers are able to get students to participate more evenly. In addition to the need to devote research attention to these issues, it is also vital to shift the focus of teacher education or professional development in this direction. This means that teachers need to learn about methods of talk-intensive teaching and they also need to be guided in creating conditions for a more proportionate participation of all students, regardless of their achievement, skills, or other characteristics.

Another important implication emerges from our analyses. The link between participation and achievement appeared to apply equally for all students, regardless of socio-economic background and gender. The existence of educational inequalities according to socio-economic background has been confirmed repeatedly (for example Bodovski et al., 2017; Pfeffer, 2008). The effort to remedy unequal conditions and minimize educational inequalities is one of the most important missions of contemporary educational sciences. One can find in the literature a suspicion that instructional strategies that emphasize student responsibility and activity, including talk-intensive pedagogies, create more of a disadvantage for already disadvantaged students, especially students from families with low socio-economic backgrounds (Andersen & Andersen, 2017). However, our findings indicated that when students, including disadvantaged ones, are successfully involved and prompted to participate and reason, the positive link was universal. Given how few widely applicable solutions capable of optimizing learning opportunities for all students are available, we consider our conclusions to be extremely important.

### 5.1. Limitations of the study

In this study, we strove to explore the link between student participation in classroom talk and student achievement on a reading literacy test. We conducted the study under natural conditions in order to capture the processes that are typical of the environment and represent the common educational experience of Czech students. We also strove to have a sufficiently large sample of classes that would not be burdened by self-selection. This intention led to some methodological choices that make up the limitations of this study.

The first limitation is presented by the fact that we conducted a correlational study. Hence, we do not have control over the direction of causality. Our argument that participation in classroom talk affects achievement is derived from our theoretical premises. The identified connection might also indicate that student achievement predicts involvement in classroom talk. In other words, high achievers, as measured by reading literacy tests, speak more and argue more. The causal direction could be controlled if we conducted an intervention study including a pre-test and a post-test, as well as intervention and control groups of students. In conducting such a study, however, we would not capture common and typical processes and phenomena.

Another limitation of our study is the fact that the quality of student contributions was coded by a single coder, directly during the lessons. This kind of real-time coding can be prone to error. This decision was driven by our attempt to get a sample representing common teachers and classes. We were concerned that the use of recording technology or a larger number of observers in the class would lead to the reluctance of schools and teachers to participate in the research and to the tendency to change the interaction behavior. We tried to minimize the risks of this choice by thoroughly training the coders, who co-authored the study.

Further limitations—having to do with the possibility of interpreting our findings—are related to the data we gathered in this data collection setting. Above all, we observed a small amount of talk and reasoning on the part of students in Czech classes. Therefore, our findings cannot illustrate how the relationship between participation and achievement works in a dialogue-rich environment. We can only assert that in a situation involving the limited participation of students in classroom talk, the degree of their individual involvement appears to play a highly important role.

Another fact that undermines our interpretation is that the observed classes were relatively homogeneous in terms of the students' socio-economic backgrounds. In other words, there were not considerable differences among the students in the sample in this respect. Our finding that a positive relationship between participation and achievement appears to apply regardless of socio-economic background should therefore be accepted with caution. It is possible that in the case of a more heterogeneous sample this conclusion would not be confirmed.

### 5.2. Next steps

This study has brought important, but in many ways limited, knowledge. One advantage is that it is part of a wider project in which we have additional sources of data to analyze.[15] Not all of the data that we collected from this sample of students have been utilized in this study. In addition, the present quantitative study was followed by a qualitative study (February to May 2018) in which we included four classes from the original sample. In these classes, we made a series of video recordings of the lessons and conducted a number of interviews with both teachers and students.

This material should allow us to shed more light on the predictors of student participation in the near future. In other words, it may enable us to discover what leads some students to be vocal and others to remain silent. In addition, thanks to qualitative data, we should be able to identify with greater specificity the reasons for non-participation: are silent students simply disaffected, or rather do they not feel the need to speak even though they are fully engaged? Are there any barriers that discourage them from participation? In addition to the focus on the characteristics of individual students, the wider context of classroom culture needs to be explored in greater detail because the conditions of participation are necessarily socially construed. The question remains: In what ways do the teachers' behavior and expectations influence whether the student engages? And further, what is the role of peer relationships and interactions in the classroom in this respect?

**Funded**

This article is an output of the project *On the Relationship between*

[15] Such as questionnaires on student engagement and other additional data about students (Sedova et al., 2016).

*Characteristics of Classroom Discourse and Student Achievement* (GA17-03643S) funded by the Czech Science Foundation.

## Appendix. An example of one of the tasks from reading literacy test

Read the text and choose the right answer.

Towards the end of the 19th century, the disease known as beriberi (in Singhalese "I cannot") with characteristic symptoms ranging from weakness to paralysis spread dramatically in Dutch East Indies (present-day Indonesia). The Dutch government, concerned that there is hardly anybody left to slave on the plantations, set up a special committee. Christian Eijkman (1858–1930), a physician of the state prison in Batavia (present-day Jakarta) was among its members. On one beautiful day, Doctor Eijkman was enjoying a view out of the window of his official apartment overlooking the prison courtyard. His head full of the cursed beriberi, he was watching the hens pecking around the yard. He was intrigued by their strange movements and postures; somehow they reminded him of the sick inmates… It turned out that the hens were, through kitchen garbage, eating basically the same diet as the inmates—mostly rice. Specifically, it was husked rice. The husks were removed because a product that had been processed in this way looks better and as a result can cost more.

The already suspecting Eijkman only needed to ask the "competitors" a few questions and everything was clear. Jails where the inmates consumed only husked rice had high numbers of beriberi patients, while in the jails where the managers economized and fed their inmates with cheap rice that had not been husked, the disease almost did not occur.

In 1897, Eijkman published his discovery. Despite its significance, there was almost no response. At that time, nutrition science was dominated by caloric assessment of nutrients, while the causes of diseases were decided by young and ambitious bacteriology. The idea that a mere deficiency in some trace element in diet could cause a serious disease or even death simply seemed ridiculous.

However, in 1911 Eijkman's work was discovered by a young biochemist of Polish origin Kazimierz Funk (1884–1967) who was living in London. He first tested Eijkman's conclusions on pigeons and then got a kilo of rice husks and used them to laboriously prepare six grams of white powder. This powder, even in milligram amount, reliably cured beriberi. Funk called it vitamin B; *vita* means "life" in Latin and *amin* was for the amin group that the powder contained. He used letter "B" to avoid confusion with a substance of a similar category that had been shortly before discovered in milk by Funk's colleague Frederick Hopkins (1861–1947) who called it growth factor A (today known as vitamin A). It is evident that Funk's name was widely accepted, even though the amin group after which it was named is out of all the thirteen vitamins that are known contained only in "his" vitamin (it is at present known as vitamin $B_1$). Eijkman and Hopkins were awarded the Nobel Prize in 1929.

(Houdek F., Tůma J.: Objevy a vynálezy tisíciletí [Discoveries and Inventions of the Millennium], Nakladatelství Lidové noviny 2002, p. 233)

| Which of the following statements is directly contradicted in the text above? | [ID477754] |
|---|---|
| ○ | Christian Eijkman died in prison in Batavia. |
| ○ | The beriberi disease can be cured with six grams of a special white powder. |
| ○ | Kazimierz Funk was later awarded the Nobel prize. |
| ○ | The beriberi disease was widespread in all prisons of the Dutch East Indies. |
| ○ | Vitamin $B_1$ can be found in particular in husked rice; rice that has not been husked does not contain it at all. |

## References


Alexander, R. (2006). *Towards dialogic teaching: Rethinking classroom talk.* Cambridge: Dialogos.

Alexander, R. (2018). Developing dialogue: Genesis, process, trial. *Research Papers in Education, 33*, 561–598.

Andersen, I. G., & Andersen, S.,C. (2017). Student-centered instruction and academic achievement: Linking mechanisms of educational inequality to schools' instructional strategy. *British Journal of Sociology of Education, 38*, 533–550.

Andreassen, R., & Bråten, I. (2011). Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms. *Learning and Instruction, 21*, 520–537.

Applebee, A. N., Langer, J., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*, 685–730.

Sedova, K., Salamounova, Z., & Svaricek, R. (2014). Troubles with dialogic teaching. *Learning, Culture and Social Interaction, 3*(4), 274–285.

Sedova, K., Salamounova, Z., Svaricek, R., & Sedlacek, M. (2017). Teachers' emotions in teacher development: Do they matter? *Studia Paedagogica, 22*(4), 77–110.

Billings, L., & Fitzgerald, J. (2002). Dialogic discussion and the Paideia seminar. *American Educational Research Journal, 4*, 907–941.

Black, L. (2004). Differential participation in whole-class discussions and the construction of marginalised identities. *Journal of Educational Enquiry, 5*, 34–54.

Blažek, R., & Příhodová, S. (2017). *Mezinárodní šetření pisa (PISA)2015.* Praha: ČŠI.

Bodovski, K., Jeon, H., & Byun, S.-Y. (2017). Cultural capital and academic achievement in post-socialist Eastern Europe. *British Journal of Sociology of Education, 38*, 887–907.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*, 365–374.

Clarke, S. N. (2015). The right to speak. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.). *Socializing intelligence through academic talk and dialogue* (pp. 167–180). Washington, DC: American Educational Research Association.

Forman, E. A., Ramirez-DelToro, V., Brown, L., & Passmore, C. (2017). Discursive strategies that foster an epistemic community for argument in a biology classroom. *Learning and Instruction, 55*, 32–39.

Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., et al. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*, 267–291.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status'. *Social Science Research, 21*, 1–56.

Gurevitch, Z. (2001). Dialectical dialogue: The struggle for speech, repressive silence, and the shift to multiplicity. *British Journal of Sociology, 52*, 87–104.

Helgevold, N. (2016). Teaching as creating space for participation – establishing a learning community in diverse classrooms. *Teachers and teaching, 22*, 315–328.

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education, 43*, 325–356.

Hox, J. (2002). *Multilevel analysis. Techniques and applications.* London: Lawrence Erlbaum Associates.

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., et al. (2015). Student participation in elementary mathematics classrooms: The missing link between teacher practices and student achievement? *Educational Studies in Mathematics, 90*, 341–356.

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23*, 33–42.

Kennedy, D. (2014). Neoteny, dialogic education and an emergent psychoculture: Notes on theory and practice. *Journal of Philosophy of Education, 48*, 100–117.

Kiemer, K., Gröschner, A., Pehmer, A.-K., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction, 35*, 94–103.

Kovalainen, M., & Kumpulainen, K. (2007). The social construction of participation in an elementary classroom community. *International Journal of Educational Research, 46*,

141–158.
Kutnick, P., & Colwell, J. (2010). Dialogue enhancement in classrooms. Towards a relations approach for group working. In K. Littleton, & C. Howe (Eds.). *Educational dialogues: Understanding and promoting productive interaction* (pp. 192–215). London, New York: Routledge.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University: Press.
Lefstein, A., & Snell, J. (2014). *Better than best practice: Developing teaching and learning through dialogue.* London: Routledge.
Lyle, S. (2008). Dialogic teaching: Discussing theoretical context and reviewing evidence from classroom practice. *Language and Education, 22*, 222–240.
Marková, I. (2003). Constitution of the self: Intersubjectivity and dialogicality. *Culture & Psychology, 9*, 249–260.
McElhone, D. (2012). Tell us more: Reading comprehension, engagement, and conceptual press discourse. *Reading Psychology, 33*, 525–561.
Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* London & New York: Routledge.
*Michener, C.. Features of dialogic instruction in upper elementary classrooms and their relationships to student reading comprehension. (2014).* Retrieved from http://hdl.handle.net/2345/bc-ir 103554.
Muhonen, H., Pakarinen, E., Poikkeus, A.-M., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2018). Quality of educational dialogue and association with students' academic performance. *Learning and Instruction, 55*, 67–79.
Murphy, K., Wilkinson, I. A. G., Soter, A. O., & Hennessey, M. N. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*, 740–764.
Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English, 40*, 392–412.
Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue. Understanding the dynamics of language and learning in the English classroom.* New York, London: Teachers College Press.
O'Connor, C., Michaels, S., & Chapin, S. (2015). "Scaling down" to explore the role of talk in learning: From district intervention to controlled classroom study. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.). *Socializing intelligence through academic talk and dialogue* (pp. 111–126). Washington: AERA.
O'Connor, C., Michaels, S., Chapin, S., & Harbaugh, A. G. (2017). The silent and the vocal: Participation and learning in whole-class discussion. *Learning and Instruction, 48*, 5–13.
Parker, P. D., Van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction, 54*, 125–137.
Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review, 24*, 543–565.
Pimentel, S. D., & McNeill, K. L. (2013). Conducting talk in secondary science classrooms: Investigating instructional moves and teachers' beliefs. *Science Education, 97*, 367–394.
Raftery, A. (1995). Bayesian model selection in social research. In P. Marsden (Ed.). *Sociological methodology* (pp. 111–163). Oxford: Blackwell Publishers.
Resnick, L. B., Asterhan, C. S. C., & Clarke, S. N. (2017). Student discourse for learning. In G. E. Hall, D. M. Gollnick, & L. F. Qzinn (Eds.). *Handbook of teaching and learning.* Wiley: Wiley-Blackwell.
Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*, 155–175.
Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist, 48*, 114–133.
Scott, P., Ametller, J., Mortimer, E., & Emberton, J. (2010). Teaching and learning disciplinary knowledge. In K. Littleton, & C. Howe (Eds.). *Educational dialogues. Understanding and promoting productive interaction* (pp. 289–303). London, New York: Routledge.
Sedova, K., Sedlacek, M., & Svaricek, R. (2016). Teacher professional development as a means of transforming student classroom talk. *Teaching and Teacher Education, 57*, 14–25.
Sfard, A. (2008). *Thinking as communicating.* New York: Cambridge University Press.
Skidmore, D., & Murakami, K. (2012). Claiming our own space: Polyphony in teacher–-student dialogue. *Linguistics and Education, 23*, 200–210.
Snell, J., & Lefstein, A. (2018). "Low ability," participation, and identity in dialogic pedagogy. *American Educational Research Journal, 55*, 40–78.
Soukup, P. (2006). Proč užívat hierarchické lineární modely? *Czech Sociological Review, 42*, 987–1012.
Straková, J. (2016). *Mezinárodní výzkumy výsledků vzdělávání. Metodologie, přínosy, rizika a příležitosti.* Praha: Univerzita Karlova.
van der Veen, C., de Mey, L., van Kruistum, C., & van Oers, B. (2017). The effect of productive classroom talk and metacommunication on young children's oral communicative competence and subject matter knowledge: An intervention study in early childhood education. *Learning and Instruction, 48*, 14–22.
van der Veen, C., & van Oers, B. (2017). Advances in research on classroom dialogue: Learning outcomes and assessments. *Learning and Instruction, 48*, 1–4.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.
Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., et al. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers instructional practices, and learning. *International Journal of Educational Research, 63*, 79–93.
Wilkinson, I. A. G., Murphy, P. K., & Binici, S. (2015). Dialogue-intensive pedagogies for promoting reading comprehension: What we know, what we need to know. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.). *Socializing intelligence through academic talk and dialogue* (pp. 37–50). Washington, DC: American Educational Research Association.