The inextricable intertwining of communication and higher order mathematical thinking

Moritz Seibold, Klaus-Peter Eichler, Daniela Bímová, Sandra Gleissberg

The inextricable intertwining of communication and higher order mathematical thinking

Moritz Seibold[1], Klaus-Peter Eichler[2], Daniela Bímová[3] and Sandra Gleißberg[1]

[1]University of Education Schwäbisch Gmünd, Germany; seibold@mms-gmuend.de

[2]NORD University, Faculty of Education and Arts, Bodø, Norway

[3]Technical University of Liberec, Faculty of Science, Humanities and Education, Czech Republic

We investigate the effect of unsupervised communicative collaboration among students in an introductory university mathematics course on their ability to solve unfamiliar problems. We examined how the study habits of 39 teacher students throughout one semester relate to their performance on unfamiliar mathematical problems during an exam. Our findings suggest that a direct relationship between increased effort and improved outcomes is not always present. Increased individual study time only positively impacts performance in tasks focused on reproducing mathematical concepts and applying routines. In contrast, increased communicative collaboration between the students during the semester seems to improve the ability to solve tasks that require higher order thinking skills. We discuss potential consequences for university mathematics education and highlight emerging research questions from our results.

Keywords: Mathematics education, mathematical communication, argumentation, abstract reasoning, logical thinking.

Introduction

The integration of the concept of higher order thinking in educational research is related to the idea that certain activities and learning tasks demand more cognitive effort than others. This perspective gained wider recognition with the advent of learning taxonomies, such as Bloom's taxonomy and its revised version (Krathwohl, 2002), but it is not strictly tied to these frameworks. The fundamental premise of this concept is that acquiring higher order thinking skills nurtures more generalizable skills than mere rote learning mathematical routines. Thus, the ability to tackle unfamiliar problems through reasoning and using known mathematical tools is a primary goal emphasized across school mathematics curricula that focus on process-related competencies associated with higher order thinking. The standards for mathematics education set in recent decades worldwide echo the significance of these process-related competencies. For instance, the Common Core State Standards underscore competencies such as problem solving, communication, reasoning and proof (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Equivalent guidelines exist in other states' educational standards. Despite efforts to create frameworks addressing the ambiguities surrounding the definition of higher order mathematical thinking (Staples & Truxaw, 2012), the inconsistent use of terms like critical thinking, problem solving, and higher order thinking after their introduction in educational research poses challenges. We adopt the term based on Lewis and Smith's (1993) definition, which posits that higher order thinking involves synthesizing and manipulating both new and stored information to achieve a goal or address perplexing problems. As Lewis and Smith, we take into account that whether a task necessitates higher order thinking is contingent upon the learning level of the students and cannot be exclusively determined by the specific attributes of the task (1993).

To develop the mathematical understanding required for tackling higher order thinking tasks, natural language comprehension and expression play a critical role, as communication facilitates the reinforcement of mental representations of mathematical concepts (O'Halloran, 2015; Wilkinson et al., 2018). As a result, it appears reasonable to hypothesize that increased communication will lead to improved mathematical performance. Discursive approaches to mathematics education verify this association: Prediger et al. (2022) recently showed that language-based mathematics instruction led to significant gains in conceptual understanding of fractions compared to a control group. We aim to investigate whether a similar outcome presents itself in situations where collaboration is increased outside of the educational setting, independent of the quality of communication. We specifically aim to determine whether this effect can be clearly distinguished from the effect of communication on tasks that are familiar to the students in a way that they can solve them straightforwardly. We commence by briefly outlining a prior experimental design involving a mock exam that we abandoned based on its outcomes. This led us to improve the mock exam by focusing on the influence of collaboration on higher order thinking tasks in mathematics students. We describe in detail the design choices that informed the second mock exam. Notably, the results presented in this paper are exclusively based on the second mock exam. Finally, we discuss the potential limitations of our investigation and introduce research questions arising from this study.

Methods

This study was conducted within an introductory university mathematics course designed for teacher students. The course includes a lecture and a seminar, and covers topics such as mathematical logic, number theory, and basic properties of relations and operations in $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. Students are expected to submit weekly assignments, which are subsequently discussed in seminars. The assignments typically require five to ten hours of work per week outside of regular university courses, and students are given the freedom to determine their approach. This free choice of working methods enables us to examine the effectiveness of various student-selected approaches. We employed students' exam performance as a measure of their mathematical task performance, seeking to assess the influence of their study time and method (individual vs. collaborative) on their performance in an unsupervised, organic context.

First mock exam

Our initial goal was to probe the influence of various videotelephony or online chat services that students utilize for exam preparation on their actual performance in an exam. We used students' results in this math exam as an indicator of the overall performance and to discern the impact of selected online collaboration services on their unsupervised, organic collaboration. We predicated this design choice on the premise that exams stand as the primary metric for evaluating students' performance in universities, and unsupervised, voluntary collaboration among students therefore predominantly targets achieving favorable exam outcomes. To initiate our investigation, we created a mock exam two months prior to the end of the semester. We tailored the tasks in this initial mock exam to mirror a genuine representation of the final exam in terms of its intricacy.

We constructed a questionnaire to collect data on students' collaboration methods and the usage of different videotelephony or online chat services. We subjected this data to analysis, comparing it with the students' exam outcomes. The scope of this investigation encompassed analyzing students' self-reported time spent learning, both individually and collaboratively. Contrary to our expectations, the analysis indicated that the amount of collaborative learning, either face-to-face or via digital platforms, did not seem to influence the students' performance on the first mock exam. Remarkably, for a subset of the exam's subtasks, the correlation deviated substantially. These subtasks required providing justification or proof for a phenomenon never encountered by the students, or an unaccustomed approach to addressing a problem that was familiar to them. Despite not being inherently challenging, these subtasks could require a higher level of thinking from individuals who had not previously encountered the subject matter. Conversely, straightforward tasks, i.e. tasks that entail providing proofs or justifications that were already addressed in a comparable manner within the course, do not need the same level of cognitive effort to solve.

This observed reversed correlation for some subtasks attracted our attention, prompting us to classify tasks into these two distinct categories. We then sought to assess the effect of the time invested in individual and collaborative learning on each of these two task types. Our characterization of these tasks is elaborated in the subsequent section. Recognizing that the actual exams of the first-semester mathematics module predominantly involve tasks centered on the execution of well-practiced routines or minor alterations to familiar tasks, our mock exam contained a limited set of tasks demanding the synthesis and manipulation of information, which we term as higher order thinking tasks. Consequently, we revised our methodological approach and undertook data collection once again with a subsequent mock exam for the same course. The ample interval between the two mock exams afforded us the opportunity to introduce multiple questions pertaining to content that was yet untaught at the time of the initial mock exam. Considering this, we projected a negligible, if any, carryover effect from the performance in the first mock exam to the second.

Second mock exam

As indicated earlier, the initial mock exam had a limited set of tasks categorized as higher order thinking tasks, necessitating an augmentation in their count for our study. Students undertook this practice exam in the concluding week of the course, just prior to the official examination phase. In our efforts to prevent potential misconceptions, we refrained from adopting an infeasible 50/50 split between higher order and straightforward tasks. Our emphasis was on ensuring the mock exam tasks mirrored, to a reasonable extent, what students might expect in the final exam. Out of a total score of 60 points spanning 14 tasks, 20 points were apportioned to six higher order thinking tasks and the remaining 40 to eight tasks of a more direct nature. This arrangement was deliberate to reduce the likelihood of adequately prepared students facing unexpectedly challenging tasks in the mock exam, which could inadvertently dampen their confidence ahead of the final examination. Simultaneously, our aim was to ensure a discernible variation in the achieved scores in the higher order thinking tasks, allowing for clear detection of possible effects.

In our statistical analysis, we selected the aggregated score in the corresponding tasks as an indicator of students' performance in these two different types of tasks: straightforward tasks (1) requiring the retrieval of knowledge from memory or the execution of a mathematical procedure, and higher order thinking tasks (2) necessitating more complex evaluations or generalizations of the covered mathematical concepts to solve a novel, yet related problem.

To clarify the differences between these outcome variables, we will illustrate our classification of tasks into straightforward and higher order thinking tasks. As depicted in Figure 1, the task on the left asks students to prove the transitivity of the congruence relation in $\mathbb{Z}$. Given the course's exploration of relations, including the proof of transitivity in the relation "… less than …" in $\mathbb{N}$, this type of task is classified as straightforward task. On the other hand, the logic underpinning the phenomenon explored in the right-side task of Figure 1 isn't readily derived from any routine covered in the module. While not inherently difficult, the task mandates a synthesis of student insights about the potential generalizations of the examples provided in the course. As this required synthesis introduces a unique challenge within the module's context, it is classified as higher order thinking task. We find categorizing tasks as straightforward or higher order based exclusively on inherent task attributes lacking. This is because continuous engagement with specific concepts could transition tasks previously branded as higher order thinking into the straightforward category. Therefore, we crafted our higher order thinking tasks to be novel experiences for students, untouched either in course discussions or the related academic literature on the subject.

What belongs in the place marked with **coffee stain**?

To show: $\forall a, b, c \in \mathbb{Z} : a \equiv b\ (m) \wedge b \equiv c\ (m) \Rightarrow a \equiv c\ (m)$

Let $a, b, c \in \mathbb{Z}$ be arbitrary.

$a \equiv b\ (m) \Rightarrow$

(coffee stain)

$\Rightarrow m|(a - c)$

$\Rightarrow a \equiv c\ (m)$

Q.E.D.

Amelie has to determine the following expression $0,8^2 + 0,2$.

By mistake she calculates $0,8 + 0,2^2$, but obtains the correct solution.

Her friend Ida is puzzled by this and tries it with the following expression: $0,7^2 + 0,3$.

Again, they get the same result when they "accidentally" square the other summand.

Elaborate on the situations in which this is possible and provide a justification for this phenomenon.

Figure 1: Example of a straightforward task (left) and a higher order thinking task (right)

The examination was conducted using the Moodle platform, which is commonly utilized for conducting exams in the relevant module of our university. Although the platform is utilized, examinations are carried out in-person and under strict supervision. The examination tasks were randomized to prevent data inaccuracies resulting from cheating attempts, and students were restricted from switching back and forth between questions. In our prior methodological design, based on the first mock exam, we utilized a pen-and-paper questionnaire and encountered numerous instances of implausible data. Upon investigation, we determined that this phenomenon arose from students sitting adjacent to one another and collaboratively responding to questions about individual study habits. Consequently, we refined our modus operandi to incorporate Moodle not merely for the examination but also for the survey delineating students' academic tendencies, presenting the questionnaire items in a randomized order.

To quantify the communicative collaboration among students, we collected data on the average amount of time they spent collaborating each week while learning the content of the course modules.

Along with this predictor variable, we collected data on the average weekly time spent on individual study for module content to evaluate the impact of two different studying methods (i.e., individual versus collaborative) on two distinct task types (i.e., straightforward versus higher order thinking tasks). Additionally, we investigated the breakdown of average collaborative study hours between face-to-face and digital collaboration among the students.

We employed a thorough data collection approach to ensure data accuracy and minimize the likelihood of potential misinterpretation of the questionnaire items. To this end, participants were presented with two slightly differing questionnaire items for each predictor variable, and the questions were not posed consecutively. These paired items, albeit with subtle variations in content, aimed to gauge the same underlying construct. Prior to aggregating the means of these paired items for analysis, we scrutinized the degree of variation between them via scatterplots. Since the data were plausible and no outliers deviated excessively, we did not exclude any participants from the analysis. Utilizing these means, we considered the following predictor variables, in addition to the previously mentioned outcome variables (aggregated scores in straightforward tasks, higher order thinking tasks, and their sum): (1) average individual study hours per week, (2) average collaborative study hours per week, and (3) percentage shares of face-to-face versus virtual collaboration. To assess the impact of collaboration on the achieved scores, irrespective of its modality (either face-to-face or digital), we examined the aggregate of the latter two variables.

For the data analysis, we used the statistical programming environment R (v4.2.2; R Core Team, 2022). Correlations, both Pearson's product-moment and Kendall's rank, were computed between predictor and outcome variables. Due to the strong correlation observed between predictor and outcome variables, it was not possible to use linear regression analysis to differentiate the variation explained by different predictors. This is expected, given that the total sum score incorporates both straightforward and higher order thinking task scores.

Findings

The second mock exam, as described earlier, consisted of 14 tasks with 60 achievable points. 39 students participated in the exam (M=30.27, SD=6.56). The total score outcomes were consistent with the expected values based on the results of prior years' exams. Within the scope of the 60 potential points, six higher order thinking tasks accounted for 20 points, while the remaining 40 points could be obtained through eight straightforward tasks. Performance in higher order thinking tasks was significantly lower, averaging around 3 out of 20 points (M=3.23, SD=2.05), compared to the straightforward tasks, where students scored an average of 27 out of 40 points (M=27.04, SD=5.65). Figure 2 displays the scores for straightforward and higher order thinking tasks based on average studying time and method (individual versus collaborative studying).

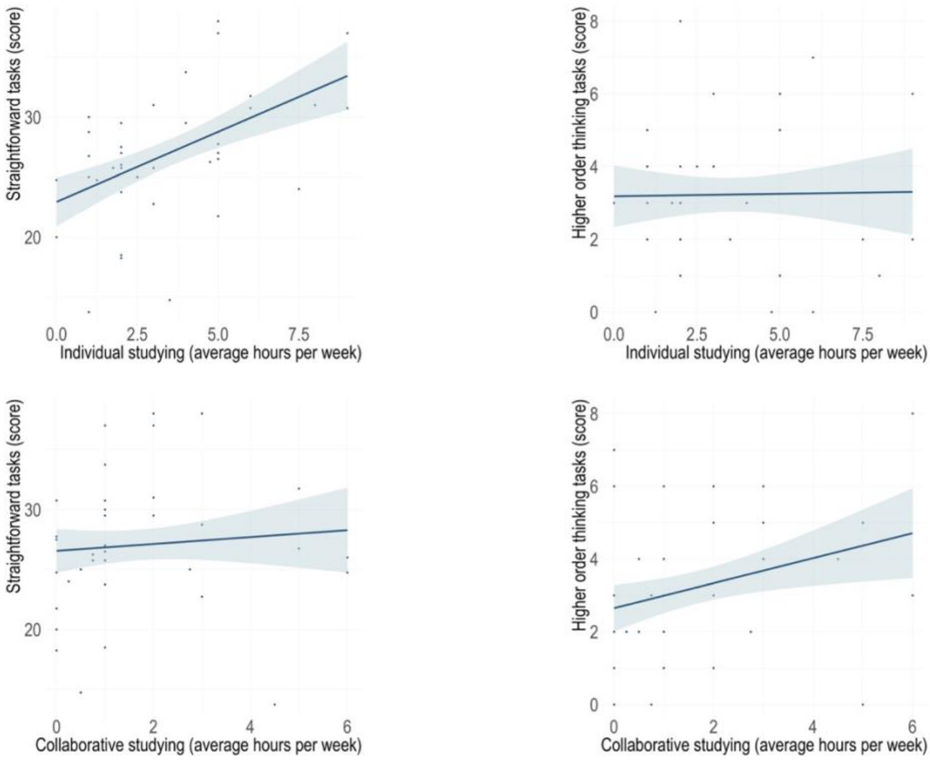

Figure 2: Scores in straightforward and higher order thinking tasks versus studying time and method

Initially, the association between average weekly collaboration and higher order thinking task scores might seem nuanced. However, when differentiating between face-to-face and digital collaboration methods, the connection becomes more salient, as illustrated in Figure 3. The plots were created using the ggplot2 package (v3.4; Wickham, 2016).

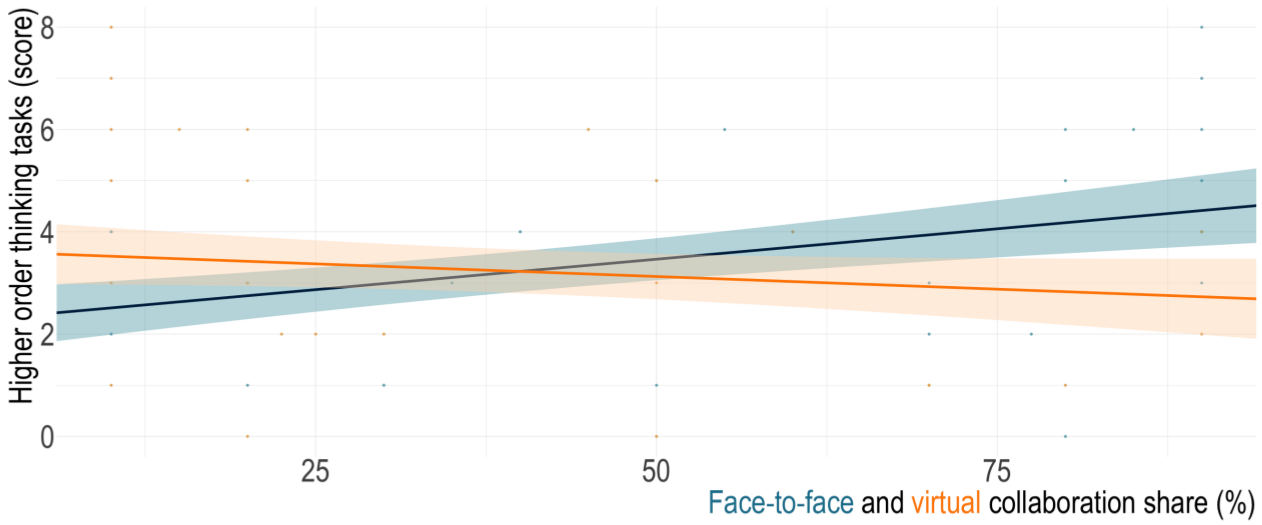

Figure 3: Scores in higher order thinking tasks versus shares of face-to-face and virtual collaboration

On average, students indicated they spent more weekly hours engaged in individual study of the course material (*M*=3.52, *SD*=2.37) than in collaborative study (*M*=1.69, *SD*=1.68). We calculated

Pearson product-moment correlation coefficients to examine the potential influence of communicative collaboration on the successful application of mathematical concepts in solving novel tasks we classified as higher order thinking. We also computed Kendall rank correlation coefficients to investigate how the proportional share of collaboration may affect attained scores, considering the violated normality of the distribution of these proportions. Additionally, this analysis was performed to determine if this effect was distinct from its impact on the tasks we classified as straightforward. These results are summarized in Table 1.

Table 1: Correlation coefficients between achieved scores and average weekly preparation

	Average weekly hours		Proportional share of collaboration	
Type of task	Alone	Collaboration	Face-to-face	Virtual
Straightforward tasks	.489**	.086	.325**	-.013
Higher order thinking tasks	.015	.283	.294*	-.096
Sum	.426**	.162	.378**	-.163

$* p < .05., ** p < .01., *** p < .001.$

Discussion

Our findings provide initial evidence suggesting that even unsupervised, natural communication regarding mathematical content might be a viable strategy for improving proficiency in mathematical tasks associated with higher order thinking. In contrast, our results imply that individual learning may prove more advantageous for tasks that have been extensively practiced or are otherwise familiar to the students. However, several questions remain unexplored. Three pressing inquiries include: (1) Discerning the factors behind improved performance in higher order tasks when students increase collaborative study hours through qualitative studies, encompassing observations of individual students and pairs of students engaging in learning and problem solving. (2) Evaluating whether integrating higher order thinking tasks into exams or emphasizing them during the semester boosts students' development, given the prevalent tilt towards individual study as reported by the students. (3) Investigating why virtual collaboration tools seemed less effective than face-to-face interactions. This would allow to investigate whether specific virtual platforms were responsible and aims to identify strategies to optimize the use of digital collaboration tools for mathematical cooperative learning. Additionally, it remains critical to delve deeper into the modalities of how students collaborated — be it online or offline, the extent to which these collaborations engaged with higher order thinking tasks, and the core subject matter of these discussions, whether they centered on mathematical concepts or primarily served motivational purposes.

We acknowledge our study's limitations and recognize the inherent complexities associated with the nature of collaboration within mathematical disciplines. While collaboration is integral, it's essential to note that individual preferences and prior exposure to certain types of tasks can significantly influence outcomes. The differentiation between straightforward and higher order tasks is not static;

it is influenced by a student's previous encounters with similar challenges. Therefore, the classification of tasks into these two categories is best determined by the instructor of a specific course, who, with their in-depth understanding of the student's learning progression, can discern if a task is a higher order thinking task for one of their students or not. Our observations suggest a concerning trend: students may be underprepared for higher order thinking tasks due to their limited inclusion in standard university math examinations. As a result, there is a propensity for students to primarily hone routine skills. If future research aligns with these observations, there's a compelling case for integrating more communicative collaboration tasks in university-level mathematics courses. The notably poor performance in higher order thinking tasks observed in both mock exams might indicate a curriculum issue, as one cannot fault students for efficiently allocating their limited time to perform particularly well in examinations.

Acknowledgment

This contribution was created as one of the outcomes of the project iTEM (Improve Teacher Education in Mathematics), EHP-CZ-ICP-2-018.

References

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218. http://dx.doi.org/10.1207/s15430421tip4104_2

Lewis, A. B., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, *32*(3), 131–137. http://dx.doi.org/10.1080/00405849309543588

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for Mathematics*. http://corestandards.org/

O'Halloran, K. L. (2015). The language of learning mathematics: A multimodal perspective. *The Journal of Mathematical Behavior*, *40*, 63–74. http://dx.doi.org/10.1016/j.jmathb.2014.09.002

Prediger, S., Erath, K., Weinert, H., & Quabeck, K. (2022). Only for multilingual students at risk? Cluster-randomized trial on language-responsive Mathematics instruction. *Journal for Research in Mathematics Education*, *53*(4), 255–276. http://dx.doi.org/10.5951/jresematheduc-2020-0193

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Staples, M. E., & Truxaw, M. P. (2012). An initial framework for the language of higher-order thinking mathematics practices. *Mathematics Education Research Journal*, *24*(3), 257–281. http://dx.doi.org/10.1007/s13394-012-0038-3

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer International Publishing.

Wilkinson, L. C., Bailey, A. L., & Maher, C. A. (2018). Students' mathematical reasoning, communication, and language representations: A video-narrative analysis. *ECNU Review of Education*, *1*(3), 1–22. http://dx.doi.org/10.30926/ecnuroe2018010301