ZDM - Mathematics Education (2025) 57:351–364
https://doi.org/10.1007/s11858-025-01680-5

ORIGINAL PAPER

Fostering Mathematical Modelling Competency in Different Learning Environments and Educational Contexts – an Exploratory Comparative Analysis of Four Intervention Studies

Alfred Lindl[1] · Rina Durandt[2] · Werner Blum[3]

Accepted: 22 March 2025 / Published online: 6 May 2025

Abstract
The project "Comparative Studies into Teaching Approaches for Mathematical Modelling" (CoSTAMM) has so far involved a total of 453 participants in four studies, all focused on fostering students' mathematical modelling competency through different learning environments, consisting of the same five-lesson modelling unit and two different teaching styles ("method-integrative" versus "teacher-directive") which represent certain elements of quality teaching. All studies followed the same pre-post-test intervention design with the same instruments but differing in their educational contexts. Three studies were implemented at the tertiary level in South Africa, and one study at the secondary level in Germany. Against this background, this paper examines the question to what extent differential effects of the interventions may be explained by the implemented specific learning environment or by the specific educational context. For this purpose, all studies were integrated into a comparative linear mixed model and analysed at once. An additional qualitative evaluation of the participants' responses to the test items was carried out to supply deeper insights. Overall, the modelling unit proved to be effective independently of the educational context, and more effective for the groups that followed the method-integrative teaching style. This paper thus provides empirical evidence that an educational intervention can effectively be transferred from one context to another while yielding simultaneously positive learning gains.

Keywords Comparative studies · Educational context · Mathematical modelling competency · Method-integrative teaching style · Modelling unit · Teacher-directive teaching style

1 Theoretical Background and Research Question

Mathematical modelling competency is the ability to construct and to use mathematical models for solving real-world problems, as well as to analyse or to compare such models (Niss & Højgaard, 2019). Fostering this competency, related to models which are accessible at certain educational levels, is an important goal of mathematics teaching around the world, in school as well as in university studies where mathematics plays a role. We know from a multitude of empirical studies that this goal is not at all easy to achieve (for an overview on empirical results see Niss & Blum, 2020, Chap. 6, and for recent research, in particular, Carreira & Blum, 2021a; Kaiser & Schukajlow, 2022; Schukajlow et al., 2023; Siller et al., 2024). Mathematical knowledge, skills and abilities are necessary but not sufficient for acquiring modelling competency. Much more is needed, in particular well-planned learning environments which fulfil essential criteria of quality teaching.

There is dense empirical evidence that the well-known *quality criteria* for effective mathematics teaching (see Brunner & Star, 2024; König et al., 2021; Kunter et al., 2013; Praetorius et al., 2018) are equally relevant for teaching mathematical modelling in particular (see Blum, 2015, for details). These criteria comprise an effective classroom management (such as structuring lessons clearly or using time effectively), orientation towards the learners (such as linking new content with individual students' pre-knowledge or using students' mistakes as learning opportunities), cognitive and meta-cognitive activation of the learn-

✉ A. Lindl
alfred.lindl@ur.de

1 University of Regensburg, Regensburg, Germany

2 University of the Witwatersrand, Johannesburg, South Africa

3 University of Kassel, Kassel, Germany

ers (that is stimulating students' mental activities, and fostering learning and working strategies), and a demanding orchestration of topics. What is often underestimated in teaching is the crucial role of meta-cognition for learning (see, for example, meta-cognition as a facet of cognitive activation, Praetorius et al., 2018), also and especially for the learning of mathematical modelling (for an overview see Vorhölter, 2024). Several studies have shown the benefits of meta-cognitive strategies, also at the tertiary level (e.g., Wedelin et al., 2015). The teaching of mathematical modelling seems to be particularly effective if *instructional* elements (that is well-designed modelling units, minimal-adaptive teacher guidance, and meta-cognitive support of students) and *constructional* elements (that is independent modelling activities of students, and their use of suitable strategies) are appropriately intertwined (Schukajlow & Blum, 2023; for similar findings from science teaching see De Jong et al., 2023). Especially important here are *strategic* teacher interventions which help students to overcome cognitive barriers without losing their independence (Leiss et al., 2010; Stender et al., 2017). A recent study at the tertiary level (Spooner, 2024) confirms how important it is for students' learning of mathematical modelling that teachers find a balance between letting students work independently and giving non-prescriptive guidance when necessary, including a lesson where the teacher demonstrates as a "model" (Brown et al., 1989) how modelling tasks can be solved. Similarly, Hidayat et al. (2025) also identify a need for instructional strategies to enhance the effectiveness of mathematical modelling instruction.

However, a caveat is always necessary when it comes to general statements about teaching effects because any learning is always situated (Watson, 1998) and thus depends on the specific learning context (the content, the tasks, the teaching methods, the students and teachers, the culture, etc.). What works under certain conditions, in a sample or in a study, does not necessarily work equally well in different contexts, be it because cultural specificities come into play (the importance of which has been emphasised already by Bishop, 1988), be it due to certain specifics of a study which, if possible, need to be identified and clearly named, or be it simply due to chance. This poses a challenge not only for social science research in general to ensure the replicability of results (Cai et al., 2018; Wiggins & Christopherson, 2019), but also for research on mathematical modelling in particular. Accordingly, Cevikbas et al. (2022) and Schukajlow et al. (2018) call for comparative studies on the teaching and learning of mathematical modelling where the same research design is transferred to a different educational context (as an example where the same Fermi tasks were given to students in different countries and cultures see Ärlebäck et al., 2024) or where certain elements such as a sequence of tasks are the same and certain elements such as the teaching style are varied.

To our knowledge, no study in mathematical modelling to date has used the same intervention, the same study design, and the same evaluation approaches in different educational contexts (e.g., study programmes, countries, or school levels) in order to test and to compare their respective effectiveness or to identify optimal conditions. This is where this paper comes in by taking advantage of the opportunity to simultaneously analyse data from four individual studies of the CoSTAMM project ("Comparative Studies into Teaching Approaches for Mathematical Modelling", see Sect. 2.1), in each of which the same study design with the same modelling unit and the same two teaching styles ('method-integrative' and 'teacher-directed'), hence with two different learning environments (described in Sect. 2.3), was used, but the target populations were exploratively varied due to different ad hoc survey opportunities, so that four different educational contexts (three with South African tertiary students and one with German secondary students, see Sect. 2.2) were involved. Aspects resulting from cultural differences between these two countries are discussed in Sects. 2 and 4 of the paper. The focus is therefore on the following *research questions*:

RQ 1: How does the teaching intervention work in each of the four studies, and how do the effects differ?

RQ 2: Can differences in the effects be attributed to specifics of the learning environments (in particular the teaching styles) and/or the educational contexts?

We assumed that all studies show progress in students' mathematical modelling competency, with a bigger progress in the intervention groups with the more elaborate teaching style (the 'method-integrative' style). We further assumed that the educational context has no significant additional influence on this, and its effects can only be seen in the way the participants work on and solve individual tasks. Correspondingly, the second research question will be investigated both from a quantitative and from a qualitative perspective.

2 Design and Methodology

2.1 The CoSTAMM Project

In the *CoSTAMM project*, launched in 2018 and directed by the authors of this paper, four studies have been carried out, three at the tertiary education level in South Africa (one in 2019, see Durandt et al., 2022a, and two in 2022, see Durandt et al. 2023, 2024) and one at the secondary level in Germany (in 2022, see Durandt et al., 2025). The aim of all four studies was to investigate the effects of two different learning environments, consisting of the same modelling unit and two different teaching styles called "teacher-directive" and "method-integrative" on the development of students' mathematical modelling competency. For this purpose, an identical longitudinal intervention was used in all four studies. This comprised an implementation of the two learning environments (Sect. 2.3) and the same instruments in the pre- and post-test (Sect. 2.4), which were realised in the same methodological way but in different educational contexts (Sect. 2.5) and evaluated identically (Sect. 2.6). This suggests that a comparative analysis of these four parallel studies is possible to examine the effectiveness of the intervention against the background of different samples and educational contexts. However, as this joint comparative analysis was not initially the core intention of the CoSTAMM project, the selection of the samples and the collection of background information were not systematically oriented towards this, which is why certain limitations are inherent in this explorative approach (see Sect. 4.2).

2.2 Sample Characteristics

Altogether a total of 453 participants was involved in those four studies. 235 students were exposed to the method-integrative (MI) style, and 218 students to the teacher-directive (TD) style. The first study, in 2019, involved 139 participants (46 MI, 93 TD) from an extended curriculum programme in engineering at a large public university in Johannesburg (for the results see Durandt et al., 2022a). The study was imbedded in the topic area of calculus suitable for first-year students at the start of their tertiary education. An *extended degree programme* is one year longer than the normal mainstream programme and designed to enable students with academic potential, that did not meet the normal admission requirements of the university, with an opportunity to obtain a degree in engineering. Typically, the first academic year is divided into two study years that contain additional developmental modules and support the students through a student-centred teaching and learning experience.

The second study, in 2022, again had a sample of 201 first-year engineering students (112 MI, 89 TD) from the same university, but this time from the mainstream programme whose entrance requirements are generally higher than for the extended curriculum programme. Although the focus of the course content is on calculus, it starts off with content on pre-calculus (e.g., introduction to functions).

The third study, also in 2022, involved 42 second-year analytical chemistry students (all MI) again from the same university, now from a science diploma programme whose admission requirements are generally lower than for science degree programmes. The mathematical content courses at the university are specifically designed for the professional development of students (for example, the mathematics course in the second-year science diploma programme includes pre-calculus content).

The fourth study was implemented in Germany at the end of the school year in 2022 on a sample of 71 grade 9 students (35 MI, 36 TD) from four different schools in the state of Hesse. The students of all four classes were taught according to the official syllabus for grade 9, and thus had a few experiences with smaller modelling tasks in the context of linear functions.

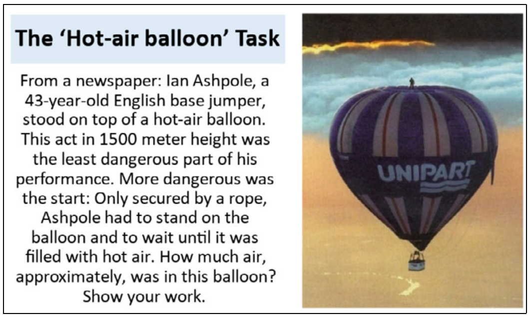
The tertiary studies are comparable due the age group (most students were between 19 and 21 years), the importance of the mathematical topic area of functions and calculus for the tertiary curricula, and similar contextual factors. For example, the language of instruction was English (like the language of instruction at school), although most students' home language is a native African language (e.g., Zulu). The three South-African samples differ in their mathematical entrance qualifications according to their programmes, as the results of a diagnostic test (see, for example, Durandt et al., 2021) and the pre-test (cf. Table 1) also suggest. Further, the individual participants in the South African samples may differ rather significantly in their school experiences (although no data is available to show these differences) because SA is a country with huge educational inequalities, and although students are exposed to the same mathematics curriculum, they can have very different learning experiences at school (see TIMSS reports, Reddy et al., 2022). In contrast, the German grade 9 students were mostly between 14 and 15 years old and the language of instruction was German, which was also most students' and the teachers' home language. An analogous use of the diagnostic test in the German sample was not possible for curricular, administrative, and organizational reasons.

In all studies the participation was voluntary (for both students and teachers), standard ethical procedures were followed, and students were not disadvantaged in their formal course results by the results of the tests.

2.3 Learning Environments: Teaching Styles and Modelling Unit

The core content of the intervention are modelling learning environments with two different *teaching styles*, a teacher-directed style called "teacher-directive" (TD) and an independence-oriented style with certain directive elements called "method-integrative" (MI). Both styles go back to the DISUM project (for further details on DISUM see Blum, 2011; Schukajlow et al., 2012) and their effectiveness was to be investigated not only at the secondary level (as in DISUM) but also at the tertiary level. Essential elements of the *teacher-directive* style (which is widely used in a similar way in many countries) are

- a systematic change between whole-class teaching (oriented towards the "average student"), and students' individual work in exercises with similar tasks, and
- the development of common solutions guided by the teacher.



The time required to evacuate an aircraft after an emergency landing at an airport needs to be known by the emergency and safety services. There are conflicting needs of aircraft construction, safety, access and ease of exit.

In a simple mathematical model, an aircraft fuselage wide enough for two seats either side of a central aisle is considered with passengers exiting singly at the front and the rear of the aircraft. Which one of the following options contains parameters, variables or constants, each of which should be included in the model?

A. Time elapsed after the emergency landing; Number of people evacuated at time *t*; Time of day at which the landing occurred.

B. Speed of people leaving their seats; Initial delay in unbuckling seatbelts before the first person can leave; Amount of personal items carried out.

C. Number of people evacuated at time *t*; Time of day at which the landing occurred; Width of the emergency exits.

D. Total time to evacuate everyone; Space between passengers leaving; Width of the emergency exits.

E. Number of people in the aircraft; Time elapsed after the emergency landing; Number of people evacuated at time *t*.

Fig. 1 (**a**) 'Hot-air Balloon' task, and (**b**) 'Aircraft Evacuation' task

Essential elements of the *method-integrative* style are

- a balance between students' independent work and teacher's guidance by minimal-adaptive (in particular strategic) interventions, aiming at individual student solutions,
- a systematic change between independent work in groups and whole-class activities, and (as a specific directive element)
- a demonstration by the teacher how to solve modelling tasks guided by a meta-cognitive instrument called "Solution Plan" which students could subsequently use in the lessons as a meta-cognitive aid.

The *Solution Plan* (see Schukajlow et al., 2015), also developed in DISUM, is a four-step modelling cycle with the steps "Understanding task", "Searching mathematics", "Using mathematics", "Explaining result". Though in its core independence-oriented, the method-integrative style should not be confused with what is often called "inquiry-based learning" because in such conceptions, teachers are often in a more reserved role, whereas in the method-integrative style teachers play an absolutely vital role, not only for explaining and conveying knowledge but also and above all for supporting students and initiating their thought processes. We decided to expose both South African and German students to this style, independent of their previous experiences from school, because what ultimately counts for successful learning is that students are cognitively activated, and more generally that basic principles of quality teaching are met (Hattie, 2023).

In addition to those elements mentioned above, essential criteria for quality teaching (see Sect. 1), especially an effective classroom management (such as a clear lesson structure and a recognisable separation of learning and assessment), a supporting and mistake-friendly classroom climate, and challenging tasks, were explicitly considered in both teaching styles.

These two styles were implemented in all four studies in the same *modelling unit*, which was developed specifically for this intervention. This unit comprises five lessons (45 minutes each) with the following contents (for details see Durandt et al., 2022b): In *lesson 1* the problem of how much air is in a balloon shown on a picture (cf. figure 1a), and the analysis of two real-life graphs, in *lesson 2* the modelling of the growth of the diameter of trees and the evacuation of an aircraft (as a multiple-choice task, taken from Haines et al., 2001, cf. figure 1b), in *lessons 3 and 4* the problem of the optimal speed for vehicles in dense traffic, and in *lesson 5* the problems of how tall a giant would be to fit the world's biggest shoes and how high a statue is whose foot is shown, as well as two tasks with proportions. All students received identical tasks in the same order during the implementation. All tasks in the unit are on the pre-calculus level and contain no culture-specific contexts, which made it possible to implement the studies in both countries and at both the secondary and the tertiary level (for examples see Figure 1a and 1b).

2.4 Measuring Instruments

As instruments for assessing performance progress, the *pre-* and *post-tests* were developed specifically for the project and aligned with the unit, and each contained three sections A, B, and C: *Section A* with open modelling tasks, aligned with lessons 1, 2, and 5 from the unit, such as estimating the weight of a container (pre-test) or the volume of a straw roll (post-test) shown on a photo (cf. Figure 2a/b), *section B* with mathematical tasks, with topics relevant for lessons 2, 3, and 4, such as identifying direct or inverse proportionality or handling a rational function, and *section C* with multiple-choice modelling tasks selected from the test of Haines et al. (2001), aligned with lesson 2. So, the tasks in sections A and C were to assess students' mathematical modelling competency, while the tasks in section B were to assess students' mathematical competency. We decided to have two separate sections for open and multiple-choice modelling tasks because the respective tasks are of a different nature insofar all tasks in section A require to construct appropriate models whereas all tasks in section C require to choose among given models. For each test section, two versions with parallel items were constructed (for sections A and B) or selected (for section C), whose real-world contexts differ only

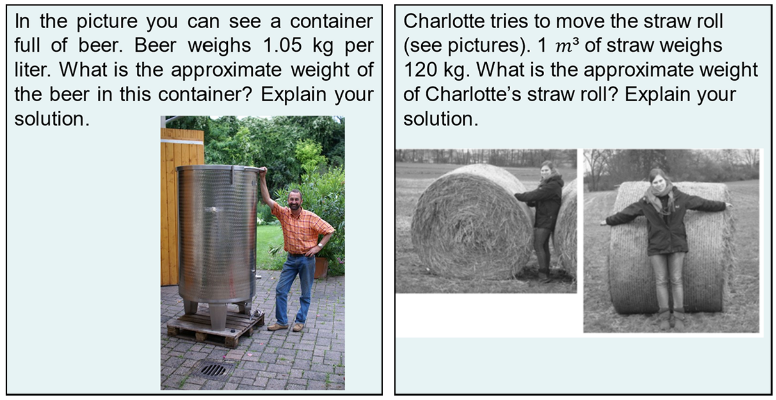


Fig. 2 (**a**) Pre-test task 'Weight of Container', and **(b)** corresponding post-test task 'Volume of Straw Roll'

slightly, allowing a rotation design and thus a comparison of the results.

The same *coding scheme* in the pre- and post-tests was carried out for all studies. For the open modelling test items (section A), student answers were marked according to a 3-mark scale. For example, in the 'Volume of a Straw Roll' task (cf. Figure 2b) the solutions were marked as: 0 marks for no or completely wrong work, 1 mark for making reasonable assumptions related to the height and radius of the roll, 2 marks for applying, in addition, a correct mathematical formula to calculate the volume of a cylinder, and 3 marks for making, in addition, an accurate estimation of the weight and presenting a final answer. For the mathematical test items (section B), student answers were marked according to a fixed memorandum; limited variation in correct answers was only possible in item B3. For the multiple-choice modelling items (section C), similar to the task presented in Figure 1b, student answers were marked according to a fixed memorandum, allowing full and partial credit, and no variation in marking was possible. This coding scheme was initially applied independently to the open-ended responses by at least two subject specialists, postdoctoral researchers in mathematics education, and a satisfactory Cohen's Kappa value of $\kappa = .737$ was obtained. Discrepancies between raters or tricky cases were discussed and finally coded by consensus.

The codings were summarized per test section and the overall test into scales and their internal consistencies were estimated using the McDonald's Omega reliability indicator (McDonald, 1999; see Table 1). As in previous reports (e.g., Durandt et al., 2022a), one item of the multiple-choice modelling tasks in section C had to be removed from the corresponding scale for psychometric reasons (negative part-whole corrected item-total correlation). Given the intended broad but time-efficient assessment of modelling competency with a small number of content-valid, heterogeneous items, the internal consistencies achieved are low, but quite usual. This should be considered when interpreting the results (Schmitt, 1996; Stadler et al., 2021; Taber, 2018) and suggests a differentiated analysis at item level (Sect. 3.2).

2.5 Implementation

In the three South African studies, the modelling unit was implemented during official mathematics tutorial class times by the second author of this paper. She taught all MI and TD groups except for one TD group in the first study which was taught by another experienced university teacher. This was done in close coordination, providing the material, and following a predefined implementation protocol. All of these resources were also made available to the German teachers (in a verified translation) who conducted the study in their own classes during regular teaching time. For this purpose, they also received detailed instruction in several preliminary meetings with the second and third authors of this paper in order to implement both styles in an optimized form.

According to the teachers' lesson protocols and their individual feedback, the implementation of the modelling unit went according to plan in all four studies. All pre- and post-tests were administered by the teachers themselves directly before and after the modelling unit, taking into account the implementation guidelines (e.g., available processing time was 45 minutes per test).

2.6 Data Analysis and Availability

Similarities and differences in initial level competency (pre-test) and in competency development (pre-post-test comparison) between the studies are analysed in linear mixed models that consider the longitudinal structure of the data at level 2 hierarchised by person (Hilbert et al., 2019; Lindl et al., 2020). At level 1, these differ in their dependent variable, that is the study-by-study z-standardised sum scores for each test section. To achieve a direct comparison of the effects of the learning environment and the educational context in a single model, the following 0/1 dummy-coded predictors are used: "time" (pre- vs. post-test), "group" (MI vs. TD), and "context" (Eng. ext., Eng. reg., Chem.; for this methodological approach, see Lindl et al., 2020). The MI group of the German study (Ger.) was chosen, for pragmatic reasons,

Table 1 Internal consistencies (McDonald's ω) per test (section), means (M), and standard deviations (SD) per group and test (section)

Test (section)	Method-integrative groups				Teacher-directive groups		
	Germany	South Africa			Germany	South Africa	
	School grade 9 $N=35$	Anal. chem. $N=42$	Engin. reg. $N=112$	Engin. ext. $N=46$	School grade 9 $N=36$	Engin. reg. $N=89$	Engin. ext. $N=93$
Pre-Test: 12 items, $\omega=.47$	8.43 (3.51)	7.50 (3.24)	8.73 (2.64)	8.20 (2.85)	8.38 (2.91)	9.10 (2.86)	8.39 (2.69)
A: 2 items, $\omega=.14$	2.17 (1.36)	0.50 (0.89)	0.62 (0.95)	1.07 (1.31)	1.35 (1.25)	0.57 (1.12)	0.75 (1.05)
B: 6 items, $\omega=.55$	3.23 (1.63)	4.10 (1.74)	4.25 (1.48)	3.67 (1.52)	4.12 (1.20)	4.37 (1.63)	3.52 (1.40)
C: 4 items, $\omega=.18$	3.03 (1.64)	2.90 (1.76)	3.85 (1.87)	3.46 (1.72)	2.91 (1.76)	4.16 (1.62)	4.12 (1.81)
Post-Test: 12 items, $\omega=.48$	11.43 (4.30)	9.50 (2.47)	10.76 (3.48)	11.79 (2.92)	8.96 (3.58)	10.17 (3.18)	10.33 (2.38)
A: 2 items, $\omega=.21$	3.21 (1.60)	1.36 (1.36)	1.48 (1.77)	2.64 (1.63)	2.32 (1.56)	1.23 (1.49)	1.86 (1.72)
B: 6 items, $\omega=.57$	3.64 (1.87)	4.67 (1.36)	5.27 (1.60)	5.17 (1.42)	3.21 (1.83)	5.02 (1.36)	4.67 (1.33)
C: 4 items, $\omega=.28$	4.57 (1.99)	3.48 (1.74)	4.01 (1.85)	3.98 (2.02)	3.41 (1.64)	3.83 (2.02)	3.80 (1.54)
$d_{\text{Pre-Post-test}}$	.77	.69	.66	1.24	.18	.35	.76

A: open modelling tasks, B: mathematical tasks, C: multiple-choice modelling tasks; anal. chem.: analytical chemistry students, engin. reg.: engineering students (regular programme), engin. ext.: engineering students (extended programme); d: Cohen's (1992) effect size d.

as the reference category, because this allows direct contextual comparisons to all South African studies and only the MI group is available in the Chem. study (this eliminates interaction effects with the factor "group" in Tables 2a and 2b). Of particular interest is the predictor "time" and its interactions with the factor "context", because this is used to model the competency gains of MI Ger. and the deviations of the MI groups in the South African studies. In addition, the interaction terms "time x group" and "time x group x context" are also important, as they describe the difference in competency development compared to the German and South African TD groups. Furthermore, the developmental trajectories of all contexts and groups are visualised using line graphs.

Data preparation and all analyses were done using the statistical software *R* (R Core Team, 2024) and additional packages such as "lme4" (Bates et al., 2015), "lmerTest" (Kuznetsova et al., 2017), "MuMIn" (Bartoń, 2023), "psych" (Revelle, 2023) or "ggplot2" (Wickham, 2016). The data used and the analysis code are available online in anonymised form at https://osf.io/t7hak/.

3 Results and Findings

3.1 Quantitative Analyses: (Differential) Effects of the Interventions (RQ 1)

3.1.1 Global Descriptive Progress

For a first overview of the competency levels and developments in all four studies, Table 1 shows the descriptive parameters by test section, group, and pre-test as well as post-test separately and the increase in competency in the entire test for each group using Cohen's (1992) effect size d. This already indicates that the modelling unit had positive effects in all groups ($d \geq .18$), but particularly in the method-integrative groups, since medium to large increases in competency were achieved here ($d \geq .66$). This apparently also applies to lower entry levels, as suggested by the similarly high level of competency development in the MI Chem. group ($d \geq .69$), which had a slightly lower initial level in the pre-test than all other groups (cf. Table 1). The gains in Eng. ext. are remarkably pronounced, with a competency increase of $d = 1.24$ in the MI group and $d = .76$ in the TD group. Possible reasons for the performance strength of this group are considered in the discussion (see Sect. 4.1), but overall these global results confirm our first assumption related to our first research question as stated in Sect. 1, that all studies show progress in students' mathematical modelling competency for both teaching styles, and progress tends to be higher with the method-integrative style.

3.1.2 Descriptive Progress per Test Section

Not only with regard to the test as a whole, but also in its individual sections, the MI groups of all studies show remarkable competency gains, as can also be seen in Table 1. The largest absolute increase in section A, in which the students in the educational context of German secondary schools perform slightly better from the outset, is found between the pre- and post-test for MI Eng. ext., the smallest for TD Eng. reg. For section B, in which the means of German school students in the pre- and post-test are on average slightly lower than those of South African students (exception: Eng. ext.), the biggest absolute progress is again for MI Eng. ext.,

Table 2a Linear mixed regression models for test section A (open modelling tasks) and test section B (mathematical tasks)

	Model A					Model B				
N | *Obs.* | *ICC*	457		889	16.5%		457		888	20.4%	
Fixed effects	β	*SE*	*df*	*t*	*p*	β	*SE*	*df*	*t*	*p*
Intercept	−0.03	0.16	808.71	−0.18	.86	−0.20	0.16	791.26	−1.24	.21
Chem.	−0.32	0.21	808.71	−1.51	.13	0.02	0.22	791.26	0.08	.93
Eng. reg.	−0.23	0.18	809.72	−1.27	.20	−0.10	0.18	793.12	−0.56	.58
Eng. ext.	−0.23	0.21	810.82	−1.10	.27	−0.14	0.21	793.77	−0.67	.51
Group	**−0.52**	**0.23**	**815.00**	**−2.31**	**.02**	**0.50**	**0.23**	**798.79**	**2.20**	**.03**
Time	**0.68**	**0.20**	**477.34**	**3.35**	**<.01**	0.27	0.20	472.54	1.33	.19
Group x Eng. reg.	0.49	0.26	816.32	1.87	.06	−0.43	0.27	800.72	−1.63	.10
Group x Eng. ext.	0.31	0.28	815.01	1.10	.27	**−0.61**	**0.29**	**798.80**	**−2.13**	**.03**
Time x Chem.	0.02	0.27	458.09	0.09	.93	0.10	0.26	453.97	0.36	.72
Time x Eng. reg.	−0.07	0.23	467.48	−0.31	.75	0.38	0.23	463.39	1.66	.10
Time x Eng. ext.	0.30	0.26	460.95	1.14	.25	**0.70**	**0.26**	**456.73**	**2.71**	**.01**
Group x Time	−0.08	0.29	489.93	−0.28	.78	**−0.79**	**0.29**	**484.96**	**−2.76**	**.01**
Group x Time x Eng. reg.	−0.06	0.33	476.66	−0.18	.86	0.56	0.33	472.27	1.71	.09
Group x Time x Eng. ext.	−0.20	0.35	473.31	−0.56	.58	0.58	0.35	468.86	1.66	.10
marg. R^2 | cond. R^2	.13			.39		.10			.40	

Note: N = sample size; Obs. = number of observations; ICC = intraclass correlation; β = (standardized) regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = probability of committing a Type I error; R^2: coefficient of determination.

while in TD Ger. the mean even decreases (see also Table 2a, model B). In section C, the MI groups of all samples tend to show a competency growth, whereas the achievement of the TD groups even decreases slightly in two out of three studies (Eng. reg. and ext.). Despite psychometric limitations in the scale of test section C (cf. Sect. 2.4), this remarkable finding thus reflects specific effects of the two teaching styles, which are reliable and worthy of discussion in view of the identical item and test booklet distribution and the parallel design in all studies (cf. Sect. 2.5), thus confirming again our assumption regarding the first research question.

3.1.3 Comparable and Significantly Different Effects Between the Four Studies

To disentangle the reciprocal effects between the different educational contexts of the individual studies and the learning environments in more detail, linear mixed models were estimated as described in Sect. 2.6. In the first model for test section A (open modelling tasks), Table 2a shows a significant increase in competency for MI Ger., coded here as the reference group ($\beta_{\text{Time}} = .68$, $p < .01$). The competency development of no learning environment (neither MI nor TD) in the other studies deviated significantly from this (cf. Figure 2a), even if their performance in the pre-test was significantly lower in some cases (e. g., TD Ger.: $\beta_{\text{Group}} = -.52$, $p = .02$). In the educational context of MI Eng. ext., the increase in competency tended to be even slightly bigger than that of MI Ger. ($\beta_{\text{Time x Eng. ext.}} = .30$, $p = .25$).

Regarding test section B (mathematical tasks; Table 2a), the reference group MI Ger. showed a positive trend ($\beta_{\text{Time}} = .27$, $p = .19$), while only MI Eng. ext. showed a significant additional increase in competency ($\beta_{\text{Time x Eng. ext.}} = .70$, $p = .01$). TD Ger. has significantly higher performance in the pre-test than MI Ger. ($\beta_{\text{Group}} = .50$, $p = .03$), but this performance decreases significantly in the post-test ($\beta_{\text{Group x Time}} = -.79$, $p = .01$, cf. Figure 2b). Possible reasons for this surprising phenomenon are reflected in Sect. 4.2, they seem to be essentially an effect of the teaching style.

With respect to the multiple-choice modelling tasks in test section C, there is, as shown in Table 2b, a significant competency increase in MI Ger. ($\beta_{\text{Time}} = .82$, $p < .01$), which is not achieved in this range by any MI group of the South African studies, but only that of Eng. reg. is significantly smaller (cf. Table 2b). The competence growth of TD Ger. tends to be lower, but still present ($\beta_{\text{Group x Time}} = -.56$, $p = .09$), while the results in the post-test of Eng. ext. and reg. are even lower than in the pre-test (cf. Figure 3c, Tables 1 and 2b).

The model for the overall test (see Table 2b) shows a significant competency growth for MI Ger. ($\beta_{\text{Time}} = .80$, $p < .01$), which is even slightly higher for Eng. ext. ($\beta_{\text{Time x Eng. ext.}} = .40$, $p = .10$). In contrast, the competency increase in the TD Ger. group is significantly lower ($\beta_{\text{Group x Time}} = -.66$, $p = .02$), whereby the difference in the competency gains between the teaching styles in

Table 2b Linear mixed regression models for test section C (multiple-choice modelling tasks) and the complete test

	Model C					Model D				
N | *Obs.* | *ICC*	457		892	11.0%		457		888	22.6%	
Fixed effects	β	SE	df	t	p	β	SE	df	t	p
Intercept	−0.22	0.17	865.10	−1.29	.20	−0.21	0.16	769.76	−1.31	.19
Chem.	0.05	0.23	865.10	0.24	.81	−0.12	0.22	769.76	−0.56	.58
Eng. reg.	0.16	0.19	865.30	0.81	.42	−0.10	0.18	772.09	−0.57	.57
Eng. ext.	−0.03	0.22	865.57	−0.12	.91	−0.25	0.21	772.73	−1.19	.23
Group	−0.06	0.24	866.49	−0.25	.81	−0.02	0.23	778.69	−0.09	.93
Time	**0.82**	**0.24**	**484.44**	**3.48**	**<.01**	**0.80**	**0.19**	**472.78**	**4.18**	**<.01**
Group x Eng reg.	0.23	0.28	866.94	0.82	.41	0.14	0.26	781.18	0.55	.58
Group x Eng ext.	0.44	0.30	866.49	1.47	.14	0.08	0.28	778.70	0.27	.78
Time x Chem.	−0.50	0.31	462.77	−1.61	.11	−0.14	0.25	455.15	−0.58	.57
Time x Eng. reg.	**−0.74**	**0.27**	**473.29**	**−2.76**	**.01**	−0.16	0.22	464.10	−0.73	.47
Time x Eng. ext.	−0.53	0.31	465.96	−1.72	.09	0.40	0.25	457.78	1.63	.10
Group x Time	−0.56	0.34	492.68	−1.68	.09	**−0.66**	**0.27**	**484.88**	**−2.39**	**.02**
Group x Time x Eng. reg.	0.30	0.38	478.33	0.78	.43	0.35	0.31	472.81	1.13	.26
Group x Time x Eng. ext.	0.08	0.41	475.73	0.19	.85	0.11	0.33	469.55	0.32	.75
marg. R^2 | cond. R^2	.03			.15		.12			.45	

Note: N = sample size; Obs. = number of observations; ICC = intraclass correlation; β = (standardized) regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = probability of committing a Type I error; R^2: coefficient of determination.

other studies (cf. Figure 3d) is somewhat more moderate ($\beta_{\text{Group x Time x Eng. reg.}} = .35$, $p = .26$; $\beta_{\text{Group x Time x Eng. ext.}} = .11$, $p = .75$).

In summary, and in line with our research hypotheses, a substantial overall competency increase can be observed in all studies as a result of the modelling unit, which is generally slightly higher in the MI groups ($d \geq .66$) than in the TD groups ($d \geq .18$). It is – as far as can be proven based on the available data – independent of the initial level, with the educational context playing a role above all in the specific test sections and competency facets. This is reflected in various forms of processing and different solution rates in the individual modelling tasks, which will be discussed in more detail in Sect. 3.2.

3.2 Qualitative Analyses: Focal Points Perceived from Students' Work (RQ 2)

A detailed qualitative analysis of the students' work on the test items revealed the striking differences between the educational contexts, presumably referring to curricular or cultural effects (secondary level, Germany vs. tertiary level, South Africa), as well as interesting differences between the learning environments, indicating special effects of the teaching styles (MI vs. TD). The most notable of these effects are summarised below for items A1 and A2 (from test section A: open modelling tasks), B3a and B3b (from section B: mathematical tasks), and C1, C3 and C5 (from section C: multiple-choice modelling tasks). These supply a global answer to our second research question by confirming that both the learning environment and the educational context have a relevant influence. Appendix I contains the corresponding results in anonymised form at item level (means and standard deviations of all items; available at: https://osf.io/t7hak/).

3.2.1 Differences Between Learning Environments

Item A2 seems to be particularly sensitive to the intervention as all MI groups performed better in the post-test than all TD groups and tended to show a bigger increase in performance (pre: $t(449) = 0.90$, $p = .37$; post: $t(436) = 2.59$, $p = .01$; see also Appendix I). This item contains a list of data values linked to a real-life situation and students were asked to find a suitable mathematical model to represent the data. An investigation in students' work revealed that the crucial idea of 'making a sketch' was present in both TD and MI groups, but far more in the MI groups (for example, 30.6% German TD cases vs. 74.3% German MI cases). This observation is again consistent with the differences in mean achievement (see Table 1). By 'making a sketch' students from the MI groups could demonstrate, in all four studies, that they have learned an essential heuristic element of solving modelling tasks, contained in the Solution Plan (see Sect. 2.3). This was found with low, medium and high performing students in both educational contexts.

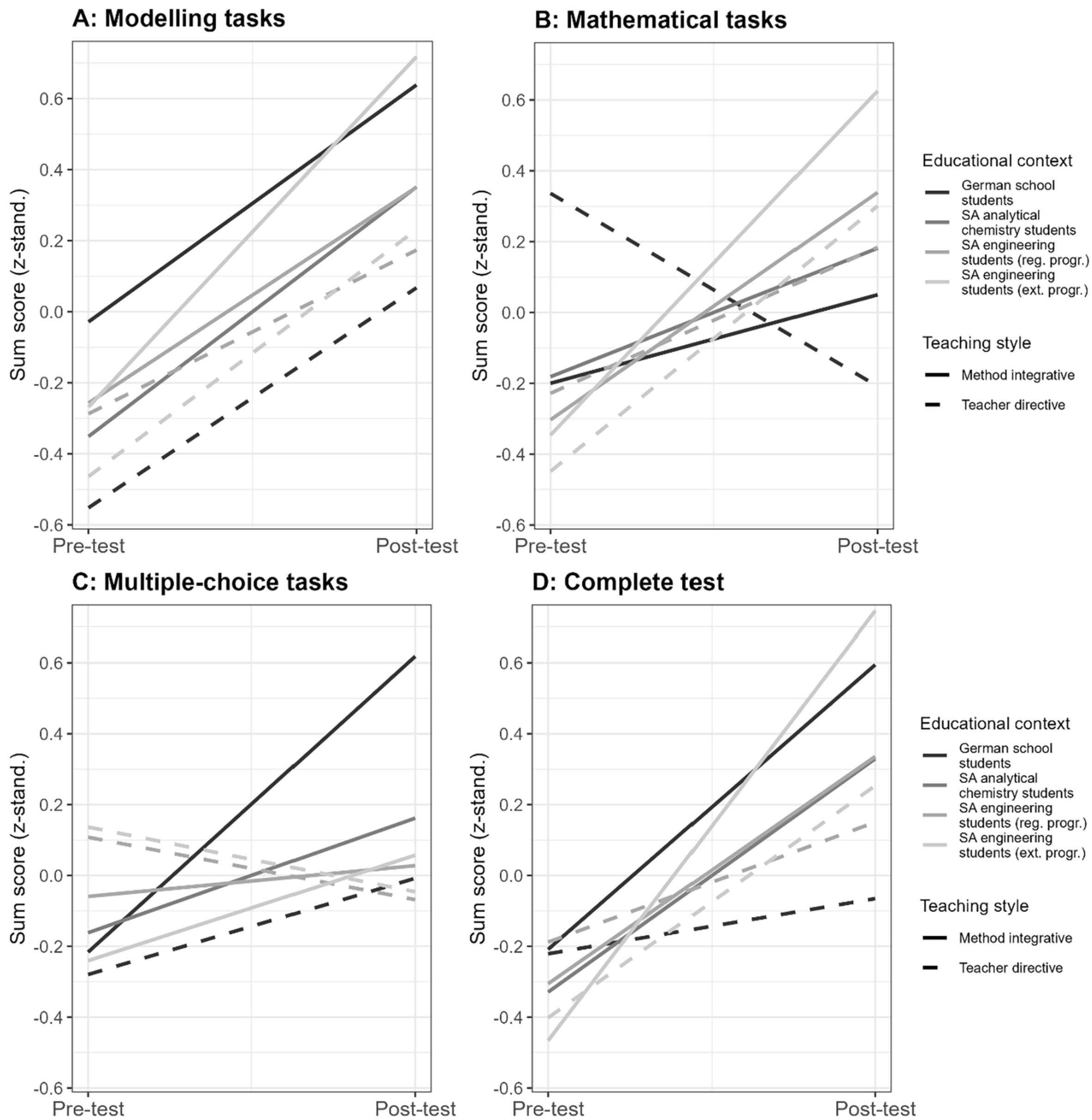


Fig. 3 (**a**) Line graph for open modelling tasks (section A), (**b**) line graph for mathematical tasks (section B), (**c**) line graph for multiple-choice modelling tasks (section C), and (**d**) line graph for the complete test

3.2.2 Differences Between Educational Contexts

The German group at secondary level performed significantly better on item A1 than the South African groups at tertiary level, for both TD and MI samples and in both the pre-and post-tests (pre: $t(449) = 12.23$, $p < .01$; post: $t(436) = 7.29$, $p < .01$). This coincides with the higher average achievement for the secondary students in test section A (see Table 1). Item A1 is an open modelling task which requires from a student a clear understanding of the situation before reasonable assumptions are made. The most obvious difference in students' work was the use of graphical representations to display the information in the question, and these were used much more frequently by the secondary students. A possible explanation is that students from the German educational context are more familiar with modelling items, and more used to elements of a more independence-oriented teaching style. Moreover, the use of graphical representations might be a more natural first step for all students at the secondary level. The German students could thus more easily visualise the given information.

Task B3 was more difficult for the secondary than for the tertiary students, for both TD and MI groups. This task contains two parts: item B3a requires to draw the graph of a rational function (pre: $t(449) = -4.62$, $p < .01$; post: $t(436) = -5.13$, $p < .01$), and item B3b asks for an estimation of its asymptotic behaviour (pre: $t(448) = -1.81$, $p = .07$; post: $t(436) = -2.18$, $p = .03$). If students struggled with the first item and could not visualise the graph, the second item was most likely also wrong. The content of rational functions is obviously more familiar to students at the tertiary level, due to experiences at the upper secondary level.

Differences in the means between the studies showed that the German secondary students performed significantly better than the SA tertiary students on item C1 (pre: $t(449) = 3.37$, $p < .01$; post: $t(439) = 2.12$, $p = .03$), and tended to perform better on C3 (pre: $t(449) = 0.05$, $p = .96$; post: $t(439) = 2.25$, $p = .03$), with a comparatively strong increase in the performance for both MI and TD samples, while the SA students performed better on item C5 (pre: $t(449) = -3.34$, $p < .01$; post: $t(439) = -1.37$, $p = .17$). A deeper look into these multiple-choice items reveals:

- Item C1 contains a narrative text which describes a real-life situation (emergency evacuation of an aircraft) and possible answer options (analogous to the item in Figure 1b). A plausible explanation for the better performance of German secondary students may be that narrative texts in modelling questions, presented in the home language of the student, are easier to understand, or South African students might not have been sufficiently familiar with the situation described in the item.
- Item C3 contains a narrative text which describes a real-life situation (speed over distance) and corresponding graphical representations. The slightly better performance of German secondary students may relate again to the better understanding of narrative texts, as well as to the observation from our study that the secondary students used more often graphical representations when working with modelling items and therefore could more easily interpret the answer options.
- Item C5 contains a narrative text to describe a real-life situation (supermarket queues) and mathematical equations/inequalities to describe the answer options. The better performance of SA tertiary students in this item may be explained by the mathematical formulations in the answer options, regardless of the narrative text. Students at the secondary level are certainly less familiar with equations/inequalities including multiple variables than students at the tertiary level.

The qualitative analysis of individual items thus clearly shows how students' approaches and solution rates differ depending on the educational context. German secondary school students tend to perform better in open modelling tasks (e.g., A1) and in multiple-choice tasks with (somewhat more extensive) narrative instruction (e.g., C1), which require a good knowledge of both the teaching language and the situation, whereas South African students tend to score better in tasks that require a confident handling of different mathematical representations, and symbolic, formal, and technical elements of mathematics (e.g., B3, C5). Regarding our second research question, it can therefore be stated that progress in mathematical modelling competency is made in both learning environments but that the individual effects achieved are not – as assumed – independent of the educational context.

4 Discussion, Limitations, and Conclusion

4.1 Context-(Un)Related Effectiveness

This paper compares the effects of four intervention studies with a uniform longitudinal research design, the same teaching unit, and identical evaluation instruments, collected in the CoSTAMM project. The design included a five-lesson modelling teaching unit, with two different teaching styles ("method-integrative" vs. "teacher-directive"), and was originally designed for students at the start of the tertiary level. The central goal of the teaching unit is to foster students' mathematical modelling competency. Three studies were implemented at the tertiary level in South Africa (in 2019 and 2022), and one study at the secondary level in Germany (in 2022), with a total of 453 participants.

To address the first research question (RQ 1), all four studies were integrated into an exploratory meta-analytic linear mixed modelling approach to analyse all results at once, particularly to what extent the effectiveness of the learning environments can be replicated across all CoSTAMM studies, and what influence the educational context (secondary level, Germany vs. tertiary level, South Africa) has. Overall, the modelling unit proved to be effective independently of the educational context, and it was more effective for the groups that followed the method-integrative style. These results confirm that teaching mathematical modelling seems to be particularly effective if instructional and constructional elements are appropriately intertwined (see Sect. 1).

We know from former studies in teacher education about the unfamiliarity of South African tertiary students towards modelling tasks and that students require substantial support in the modelling process (e.g., Biccard, 2024; Durandt, 2018, 2021; Durandt & Lautenbach, 2020). Against this background, and thus considering our second research question (RQ 2), the performance strength of the method-integrative group of South African engineering students in the extended curriculum programme is remarkable ($d = 1.24$). It shows how effective the intervention can be despite comparatively lower prior knowledge, but under an optimal learning environment. Since this was not the only group taught by the second author of this paper (Durandt et al., 2022a), administration effects during implementation appear insufficient as a justification. However, one plausible reason is the nature of the extended curriculum programme itself which contains additional developmental modules and support structures for students provided by the university (e.g., support with academic language, reading and writing, and smaller classes). The programme characteristics, in combination with our modelling unit, provided an ideal learning environment for students. These encouraging results call for a substantial increase in the importance of mathematical modelling in the South African school and university curriculum. So, the results may have consequences for teaching mathematical modelling at secondary and tertiary levels, including teacher education.

4.2 Differential Findings of the Quantitative and Qualitative Analysis, and Their Limitations

An unexpected and at first glance even contra-intuitive result is the performance decrease between pre- and post-test of both German TD classes in test section B (mathematical tasks). This raises questions about the selection of the sample, although it does not visibly differ from that of the other two German classes (Durandt et al., 2025), and about the design-based implementation of the study (Sect. 2.5), which, however, did not reveal any irregularities in subsequent interviews with the teachers of these classes. Both TD teachers, independently of each other, explained this decrease rather as an undesirable effect of the teacher-directive teaching in these classes The teaching unit contained only modelling tasks, and since the students followed the teachers' instructions and concentrated on what was recognisably important for the teacher, they set their minds on modelling tasks only and were amazed or even confused by the mathematical tasks in the post-test. Also, in their previous class tests, these students had experienced that only the content that was covered in the hours before the test were queried. If this interpretation of the TD decrease in section B is valid, then it is fully justified, in the light of RQ 2, to regard also the results in section B as an indication of the superiority of the MI teaching style.

When interpreting the results, it should also be noted that the reliabilities of the test sections as well as of the entire test, which serve as dependent variables in all models, are rather low. Such values are not unusual when measuring knowledge and understanding, as one sub-competency can increase without another sub-competency (which may be focused on in another item) also increasing. This affects the internal consistency of a test if the respective competency aspect is to be assessed in its thematic and procedural heterogeneity, aligned with the topics of the modelling unit and as valid as possible, but with a manageable number of items. These low reliability values can lead to an underestimation of the relationships and effects in the models (e.g., Schmitt, 1996; Stadler et al., 2021; Taber, 2018). Given this limitation, a qualitative evaluation of the participants' test responses for open modelling, mathematical and multiple-choice modelling items supplies not only additional insights, particularly how the solution of mathematical modelling tasks differs depending on learning environment and educational context, according to RQ 2. Rather, this item-by-item analysis, as presented in Sect. 3.2 and Appendix I, can verify and validate the results at scale level but of course only has a very selective, specific informative value and limited generalisability.

An interesting finding from this qualitative evaluation is, for example, that students from the MI groups more often used graphical representations than students from the TD groups. The idea of 'making a sketch' is an important exploratory element of solving modelling tasks, and we know that there are relations between the accuracy of the drawing and students' modelling competency (see Rellensmann et al., 2017).

Further, our investigation suggests a link between learning mathematical modelling and learning in the home language, specifically in questions with narrative texts. We know from several studies (see, e.g., Plath & Leiss, 2018) how important language proficiency is for understanding the real-world context of a modelling task and that improving the students' fluency in the education language is critical to improving student achievement (see, e.g., Setati et al., 2009). Nevertheless, South African tertiary students are taught entirely in English (the samples met the entrance qualification of the institution related to language proficiency) and research has shown that multilingualism does not necessarily imply a weaker performance for narrative text understanding (Essien, 2024). An obvious alternative reason might be related to the socio-cultural context – most tertiary students in the sample might, for instance, not have been exposed to an aircraft before. Overall, our study may also be regarded as preliminary research into the issue of how socio-cultural context (e.g., Bishop, 1988; Dreher et al., 2024), language and multilingualism (Essien, 2024) may influence the learning of mathematics. Further studies addressing this issue are needed.

Finally, there are, of course, many more variables that may influence the results of such intervention studies, in particular the professional competencies of the involved teachers. However, we have no teacher data available for our studies, and our background information on the respective samples is limited, which is certainly a limitation of our study. Therefore, we have to restrict ourselves to plausible interpretations based on the methodologically controlled, rigorous pre-post-test design and on our knowledge about the interventions themselves as well as about the educational background of the samples.

4.3 Summary and Further Research

In the CoSTAMM studies, two teaching styles were implemented in varying samples to foster students' modelling competency, and the method-integrative style which combines both constructional and instructional elements in a specific manner proved to be more effective than the teacher-directed style in all educational contexts. Some differential effects could also be identified, due to some specifics of the samples. In conclusion, this study provides empirical evidence how a learning environment for mathematical modelling can effectively be transferred from one educational context to another while yielding simultaneously positive learning gains. The comparative approach presented in this paper shows that it is worthwhile implementing the same intervention in different contexts and exploring, through both quantitative and qualitative methods, communalities and differences in the outcomes.

Plans and ideas for further CoSTAMM studies comprise the extension of the present modelling unit by implementing a broader spectrum of modelling tasks and the inclusion of teacher competencies as a variable (see Buchholtz et al., 2023) with an instrument which can be used both at the secondary and the tertiary level. We also intend to replicate our intervention in further samples in both countries, to investigate more deeply the potential sensitivity of teaching styles to socio-cultural diversity, and to review our results repeatedly under different circumstances and conditions in order to obtain robust evidence for improving the teaching and learning of mathematical modelling.

Acknowledgements This work is based on research partially supported by the National Research Foundation (NRF) of South Africa, Unique Grant No. 121969, and a TWAS-DFG scholarship of the second author in 2022 at the University of Kassel in Germany.

Funding Information Open Access funding enabled and organized by Projekt DEAL.

References

Ärlebäck, J. B., Albarracín, L., Orey, D., Rosa, M., & Sevinc, S. (2024). Exploring the potential of using Fermi problems to elicit and develop cultural aspects in modelling processes. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 525–535). Berlin: Springer. https://doi.org/10.1007/978-3-031-53322-8_43.

Bartoń, K. (2023). MuMIn: Multi-Model inference. R package version 1.47.5. https://CRAN.R-project.org/package=MuMIn.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Biccard, P. (2024). Productive struggle in mathematical modelling. *The Mathematics Enthusiast*, *21*(1), 99–112. https://doi.org/10.54870/1551-3440.1620.

Bishop, A. J. (1988). Mathematics education in its cultural context. *Educational Studies in Mathematics*, *19*(2), 179–191. https://doi.org/10.1007/BF00751231.

Blum, W. (2011). Can modelling be taught and learnt? Some answers from empirical research. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling (ICTMA 14)* (pp. 15–30). Berlin: Springer. https://doi.org/10.1007/978-94-007-0910-2_3.

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do? In S. J. Cho (Ed.), *The proceedings of the 12th international congress on mathematical education – intellectual and attitudinal challenges* (pp. 73–96). Berlin: Springer. https://doi.org/10.1007/978-3-319-12688-3_9.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*, 32–42. https://doi.org/10.3102/0013189X018001032.

Brunner, E., & Star, J. (2024). The quality of mathematics teaching from a mathematics educational perspective: What do we actually know and which questions are still open? *ZDM – Mathematics Education*, *56*, 775–787. https://doi.org/10.1007/s11858-024-01600-z.

Buchholtz, N., Kaiser, G., & Schwarz, B. (2023). The evolution of research on mathematics' teachers competencies, knowledge and skills. In A. Manizade, N. Buchholtz, & K. Beswick (Eds.), *The evolution of research in teaching mathematics* (pp. 55–89). Berlin: Springer. https://doi.org/10.1007/978-3-031-31193-2_3.

Cai, J., Morris, A. K., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, *49*(1), 2–8. https://doi.org/10.5951/jresematheduc.49.1.0002.

Carreira, S. & Blum, W. (Eds.) (2021). Mathematical modelling in the teaching and learning of mathematics: Part 1/ Part 2. *Quadrante*, *30*(1/2).

Cevikbas, M., Kaiser, G., & Schukajlow, S. (2022). A systematic literature review of the current discussion on mathematical modelling competencies: State-of-the-art developments in conceptualizing, measuring, and fostering. *Educational Studies in Mathematics*, *109*, 205–236. https://doi.org/10.1007/s10649-021-10104-6.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

De Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, E. C., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence – the case for combining inquiry-based and direct instruction. *Educational Research Review*, *39*, 1–14. https://doi.org/10.1016/j.edurev.2023.100536.

Dreher, A., Wang, T.-Y., Feltes, P., Hsieh, F.-J., & Lindmeier, A. (2024). High-quality use of representations in the mathematics classroom – a matter of the cultural perspective? *ZDM–Mathematics Education*, *56*, 965–980. https://doi.org/10.1007/s11858-024-01597-5.

Durandt, R. (2018). *A strategy for the integration of mathematical modelling into the formal education of mathematics student teachers*. Doctoral dissertation, University of Johannesburg.

Durandt, R. (2021). Design principles to consider when student teachers are expected to learn mathematical modelling. *Pythagoras*, *42*(1), a618. https://doi.org/10.4102/pythagoras.v42i1.618.

Durandt, R., & Lautenbach, G. (2020). Strategic support to students' competency development in the mathematical modelling process: A qualitative study. *Perspectives in Education*, *38*(1), 211–223. https://doi.org/10.18820/2519593X/pie.v38i1.15.

Durandt, R., Blum, W., & Lindl, A. (2021). Exploring first-year engineering students' prior knowledge in mathematics. In M. Qhobela & L. G. Mohafa (Eds.), *The 29th annual conference of the southern African association for research in mathematics, science and technology education* (pp. 28–31). National University of Lesotho.

Durandt, R., Blum, W., & Lindl, A. (2022a). Fostering mathematical modelling competency of first year South African engineering students: What influence does the teaching design have? *Educational Studies in Mathematics*, *109*, 361–381. https://doi.org/10.1007/s10649-021-10068-7.

Durandt, D., Blum, W., & Lindl, A. (2022b). A mathematical modelling unit for first-year engineering students. *Modelling in Science Education and Learning*, *15*(1), 77–92. https://doi.org/10.4995/msel.2022.16646.

Durandt, R., Blum, W., & Lindl, A. (2023). The effects of different teaching approaches on engineering students' modelling competency. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th conference of the international group of the psychology of mathematics education* (Vol. *II*, pp. 251–258). PME.

Durandt, R., Blum, W., & Lindl, A. (2024). The effects of an independence-oriented teaching design on the development of tertiary students' modelling competency. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 247–257). Berlin: Springer. https://doi.org/10.1007/978-3-031-53322-8_19.

Durandt, R., Blum, W., & Lindl, A. (2025). A comparative case study on teaching modelling at the secondary level with a unit developed for the tertiary level. In T. Ikeda, A. Saeki, V. Geiger, & G. Kaiser (Eds.), *International horizons in mathematics modelling education*, Berlin: Springer.

Essien, A. A. (2024). *Multilingualism in mathematics education in Africa*. Bloomsbury Academic.

Haines, C., Crouch, R., & Davies, J. (2001). Understanding students' modelling skills. In J. Matos, W. Blum, K. Houston, & S. Carreira (Eds.), *Modelling and mathematics education, ICTMA 9: Applications in science and technology* (pp. 366–380). Chichester: Horwood. https://doi.org/10.1533/9780857099655.5.366.

Hattie, J. (2023). Visible learning, the sequel. A synthesis of over 2, 100 meta-analyses relating to achievement. *Routledge*. https://doi.org/10.4324/9781003380542.

Hidayat, R., Mohd Ayub, A. F., Abd Hamid, H. S., & Nasir, N. (2025). Difficulties in instructing mathematical modeling: Case study in Malaysia. *International Electronic Journal of Mathematics Education*, *20*(2), em0811. https://doi.org/10.29333/iejme/15815.

Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *Testing, Psychometrics, Methodology in Applied Psychology*, *26*, 101–119.

Kaiser, G. & Schukajlow, S. (Eds.) (2022). Innovations in measuring and fostering modelling competencies. *Educational Studies in Mathematics*, *109*(2).

König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Felske, C., Musekamp, F., & Kaiser, G. (2021). The link between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom. *Educational Studies in Mathematics*, *107*, 189–212. https://doi.org/10.1007/s10649-020-10021-0.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers – results from the COACTIV project*. Berlin: Springer. https://doi.org/10.1007/978-1-4614-5149-5.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Leiss, D., Schukajlow, S., Blum, W., Messner, R., & Pekrun, R. (2010). The role of the situation model in mathematical modelling: Task analyses, student competencies and teacher interventions. *Journal für Mathematik-Didaktik*, *31*, 119–141. https://doi.org/10.1007/s13138-010-0006-y.

Lindl, A., Krauss, S., Schilcher, A., & Hilbert, S. (2020). Statistical methods in transdisciplinary educational research. *Frontiers in Education*, *5*, 97. https://doi.org/10.3389/feduc.2020.00097.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale: Erlbaum.

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling*. London: Routledge. https://doi.org/10.4324/9781315189314.

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, *102*, 9–28. https://doi.org/10.1007/s10649-019-09903-9.

Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM – Mathematics Education*, *50*(1/2), 159–171.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality. The German framework of three basic dimensions. *ZDM – Mathematics Education*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4.

R Core Team (2024). *R: A language and environment for statistical computing. R foundation for statistical computing*. https://www.R-project.org/.

Reddy, V., Winnaar, L., Arends, F., Juan, A., Harvey, J., Hannan, S., & Isdale, K. (2022). *The South African TIMSS 2019 grade 9 results: Building achievement and bridging achievement gaps*. HSRC Press. http://hdl.handle.net/20.500.11910/19286.

Rellensmann, J., Schukajlow, S., & Leopold, C. (2017). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educational Studies in Mathematics*, *95*, 53–78. https://doi.org/10.1007/s10649-016-9736-1.

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research*. Evanston: Northwestern University. R package version 2.3.9. https://CRAN.R-project.org/package=psych.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350.

Schukajlow, S., & Blum, W. (2023). Methods for teaching modelling problems. In G. Greefrath, S. Carreira, & G. A. Stillman (Eds.), *Advancing and consolidating mathematical modelling – research from ICME-14* (pp. 327–339). Berlin: Springer. https://doi.org/10.1007/978-3-031-27115-1_20.

Schukajlow, S., Kaiser, G., & Stillman, G. (2018). Empirical research on teaching and learning of mathematical modelling: A survey on the current state-of-the-art. *ZDM – Mathematics Education*, *50*, 5–18. https://doi.org/10.1007/s11858-018-0933-5.

Schukajlow, S., Kaiser, G., & Stillman, G. (Eds.) (2023). Modeling from a cognitive perspective. *Mathematical Thinking and Learning*, *25*(3).

Schukajlow, S., Kolter, J., & Blum, W. (2015). Scaffolding mathematical modelling with a solution plan. *ZDM – Mathematics Education*, *47*, 1241–1254. https://doi.org/10.1007/s11858-015-0707-2.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics*, *79*, 215–237. https://doi.org/10.1007/s10649-011-9341-2.

Setati, M., Chitera, N., & Essien, A. (2009). Research on multilingualism in mathematics education in South Africa: 2000–2007. *African Journal of Research in Mathematics, Science and Technology Education*, *13*(sup1), 65–80. https://doi.org/10.1080/10288457.2009.10740662.

Siller, H.-S., Geiger, V., & Kaiser, G. (Eds.) (2024). *Researching mathematical modelling education in disruptive times*. Berlin: Springer. https://doi.org/10.1007/978-3-031-53322-8.

Spooner, K. (2024). Investigating independent student work and instructor guidance for tertiary mathematical modelling activities. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 365–375). Berlin: Springer. https://doi.org/10.1007/978-3-031-53322-8_29.

Stadler, A., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, *60*, 100832. https://doi.org/10.1016/j.newideapsych.2020.100832.

Stender, P., Krosanke, N., & Kaiser, G. (2017). Scaffolding complex modelling processes: An in-depth study. In G. A. Stillman, W. Blum, & G. Kaiser (Eds.), *Mathematical modelling and applications: Crossing and researching boundaries in mathematics education* (pp. 467–477). Berlin: Springer. https://doi.org/10.1007/978-3-319-62968-1_9.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Vorhölter, K. (2024). Structure, role, and promotion of metacognitive modelling strategies: Theoretical derivations and empirical findings. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 63–80). Berlin: Springer. https://doi.org/10.1007/978-3-031-53322-8_5.

Watson, A. (Ed.) (1998). *Situated cognition and the learning of mathematics*. Oxford: University of Oxford.

Wedelin, D., Adawi, T., Jahan, T., & Andersson, S. (2015). Investigating and developing engineering students' mathematical modelling and problem-solving skills. *European Journal of Engineering Education*, *40*(5), 557–572. https://doi.org/10.1080/03043797.2014.987648.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Berlin: Springer. https://doi.org/10.1007/978-3-319-24277-4_9.

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137.