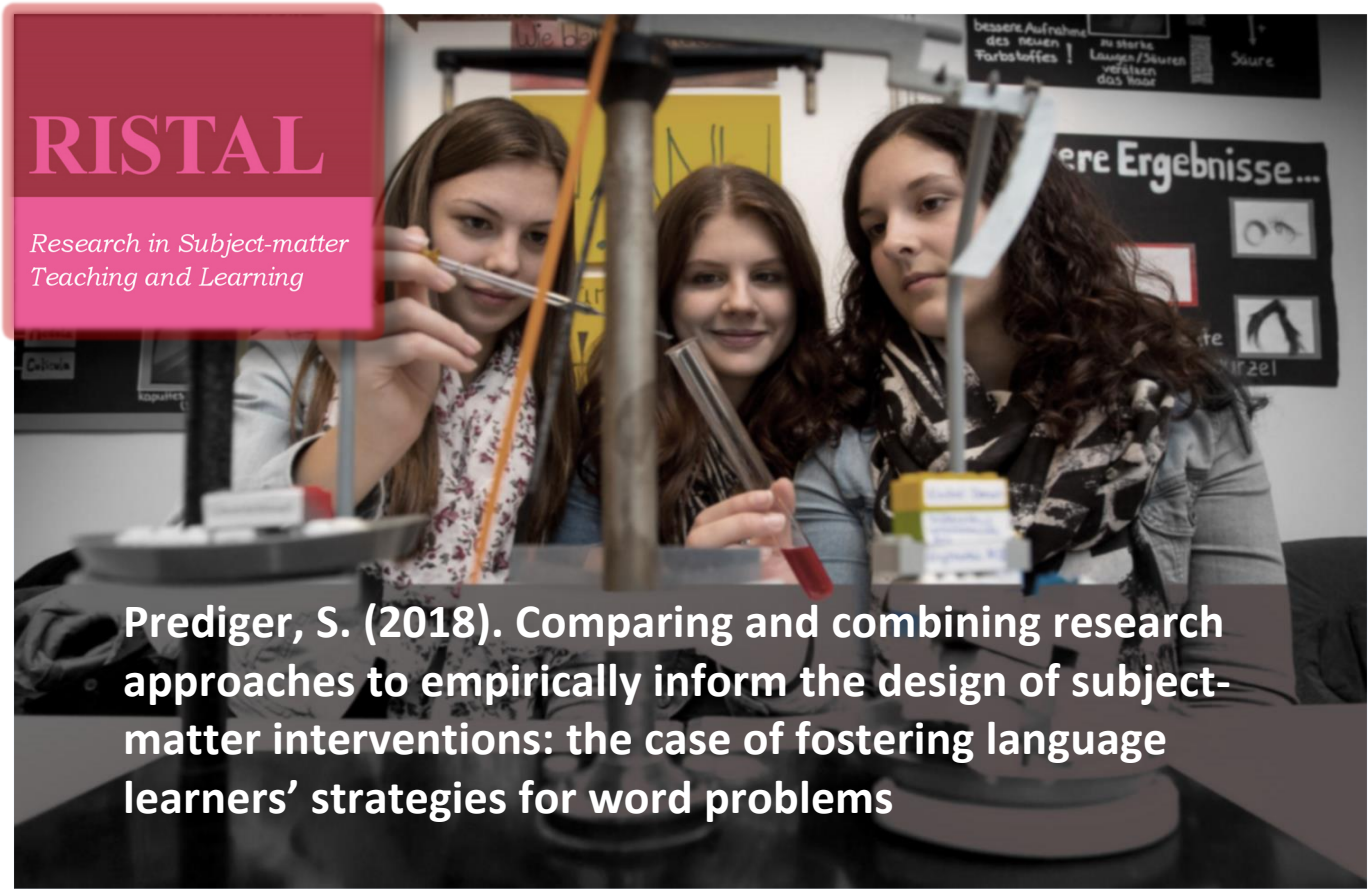
RISTAL

Research in Subject-matter Teaching and Learning



Prediger, S. (2018). Comparing and combining research approaches to empirically inform the design of subject-matter interventions: the case of fostering language learners' strategies for word problems

RISTAL 1 / 2018

Research in Subject-matter Teaching and Learning

Launch issue

Citation:
Prediger, S. (2018). Comparing and combining research approaches to empirically inform the design of subject-matter interventions: the case of fostering language learners' strategies for word problems. *RISTAL, 1*, 4–18.

DOI: https://doi.org/10.23770/rt1808

ISSN pending

Comparing and combining research approaches to empirically inform the design of subject-matter interventions: the case of fostering language learners' strategies for word problems

Susanne Prediger

Abstract

Most research approaches in subject-matter education aim at informing the design of subject-matter interventions and the work of teachers in subject-matter classrooms. However, different research foci result in it being done in a more or less direct form, and conclusions that are too far reaching are often drawn that are not covered by the research study. In particular, few research questions also address the WHAT-question (What exactly must be learned?) rather than only the HOW-question (How can it be learned most effectively?). This article examines the exemplary field of fostering language learners' reading strategies in subject-matter classrooms in order to compare what different research approaches can contribute and how they must be combined in order to provide a substantial foundation for the design of language-responsive classrooms from multiple perspectives. Using design research, language demands can be specified in subject-matter learning situations; it turns out to be crucial to take this research step before designing and evaluation interventions.

Keywords

Language proficiency, subject-matter achievement, design principles for fostering language learners, research approaches, subject specificity

For each classroom innovation, there is a long journey of research, development, and implementation from recognizing its principal relevance to implementing it effectively in classrooms, and many researchers discuss which research approaches can best provide the empirical foundation for designing and implementing classroom innovations (e.g., Burkhardt & Schoenfeld, 2003, and Cobb et al., 2003). Some academic disciplines still prioritize single research approaches, such as experimental or quasi-experimental controlled trials, to showing efficacy in experimental psychology, which was once even declared to be the only research approach deserving funding (No Child Left Behind Act, US Congress, 2001). However, this restriction must be widened in two ways:

- *From generic to subject-specific and topic-specific research:* The focus on one single research approach is critical, especially if it is accompanied by a *generic claim*: Providing empirical evidence for the efficacy of a certain design element or teaching component for a specific learning content does not immediately mean that the same design element works for other learning topics in the same way. This is why subject-specific (e.g., physics, mathematics, history) or even topic-specific (e.g., physical concept of force, mathematical procedure of multiplying multi-digit numbers) research is required. In their widely cited meta-analysis, Seidel and Shavelson (2007) emphasize that "providing opportunities for students to engage in domain-specific

DOI: https://doi.org/10.23770/rt1808

learning activities was shown to be the component with the highest effect sizes" (Seidel & Shavelson, 2007, p. 483). This article argues that beyond the domain-specificity, even the topic-specificity of the research is crucial.

- *Acknowledging and systematically combining multiple research approaches:* In contrast to experimental psychology, subject-matter education disciplines can be characterized by the big diversity of accepted research approaches applied (GFD, 2016; Bikner-Ahsbahs, Knipping, & Presmeg, 2016). However, these multiple research approaches often stay unconnected next to each other.
- These two extensions for research approaches are discussed in this article using the concrete case of one specific innovation: fostering language learners' abilities to solve mathematical word problems. The need to foster students' language learning is now widely acknowledged, as many large-scale studies have shown the language gap in students' achievements. However, much research is still required in order to clarify what exactly is relevant in coping with word problems and how it can be fostered in subject-specific or subject-independent ways. This article discusses different research approaches, many of which have been combined by the MuM (Mathematiklernen unter Bedingungen der Mehrsprachigkeit [Mathematics Learning in Multilingual Contexts], founded in 2008) research group in Dortmund. The aim of the article is to provide examples of how different research approaches can contribute to providing empirical foundations for the design of language-responsive classrooms without presenting a complete research report for each of the studies.

Section 1 starts by summarizing the state of the research on the language gap and students' difficulties with word problems. Section 2 resumes the research journey of the MuM research group by combining different research approaches. Section 3 looks back on this concrete research journey by providing a brief meta-level comparison and a plea for combining research approaches.

1 State of the research on language learners' difficulties with word problems

1.1 Large-scale assessment studies revealing a language gap

In general, the research approach of using *large-scale assessments* aims at comparing the achievement of students with different backgrounds. In this approach, tests are administered to a large sample of students in order to assess selected competences (in our case, mathematics achievement) and some background factors (in our case, socio-economic status, language proficiency, and others).

In the concrete research field we are examining, large-scale assessment studies have repeatedly shown that privileged students outperform the underprivileged students. In most cases, the privileges are conceptualized by factors of socio-economic status, immigrant background, or family language (Haag, Heppt, Stanat, Kuhl, & Pant, 2013; OECD, 2007). These factors are used more often than language proficiency because they can easily be measured by students' self-reports or existing school data, such as free school meals.

DOI: https://doi.org/10.23770/rt1808

However, when language proficiency in the language of instruction is also controlled, it turns out to be the factor with the strongest connection to mathematics achievement: stronger than socio-economic status, immigrant background, or family language (Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2018; Ufer, Reiss, & Mehringer, 2013). Among the different facets on the lexical, syntactical, and discursive level in language production and language reception, *general reading proficiency* seems to be a good predictor of limited mathematics achievement (Abedi & Lord, 2001; Paetsch, Felbrich, & Stanat, 2015); general vocabulary knowledge is also a good predictor, but syntactical skills is not (Paetsch et al., 2015).

The language gap occurs for multilingual as well as monolingual learners. Hence, for this article, the term *language learner* refers not only to second language learners but also to monolingual students with low academic language proficiency. The focus is on students who grew up in Germany, not on newly immigrated students.

These kinds of findings from large-scale assessments are crucial for detecting a field of action. Indeed, policy makers have reacted and decided that fostering language learning must be the task for all subject-matter classrooms, on the district, state, and European level (e.g., Thürmann, Vollmer, and Pieper, 2010; Moschkovich, 2013). However, large-scale assessments can only inform general school policy and cannot provide an empirical base for the design of language-responsive classrooms.

1.2 Example for interventions studies with insufficient short cuts: training general reading proficiency does not increase mathematics achievement

In general, the research approach of *intervention studies using quasi-randomized controlled trials* aims at comparing the effects of two or more interventions with respect to learning gains that are measured using pre-tests, post-tests, and follow-up tests for two comparable (quasi-randomly assigned) subsamples of students. In the context of language learners, a randomized controlled trial can compare different interventions enhancing the language proficiency.

A typical shortcut when large-scale assessments are used as an immediate base for action is exemplified by two intervention studies with quasi-randomized controlled trials that confounded predictors with the content to be learned: Starting from the finding that general reading proficiency is the strongest predictor for mathematics achievement, these two intervention studies developed trainings for general reading proficiency in Grade 7 (Hagena, Leiß, & Schwippert, 2017) and Grade 3 (Hellmich & Förster, 2015); the latter was also combined with general vocabulary work. Both studies were able to show some learning gains in general reading proficiency, but neither reached significant learning gains in mathematical achievement that was measured by solving word problems (Hagena et al., 2017) or by a general mathematics test covering procedural skills and word problems.

The failure of the two intervention studies gives the first indication that generic interventions with a focus on subject-independent competencies are not focused enough to reach mathematical learning gains, thus illuminating the need for further investigation into what exactly hinders language learners' successful solution of the test items. Specifically, this experience calls for *subject-specific* investigations.

DOI: https://doi.org/10.23770/rt1808

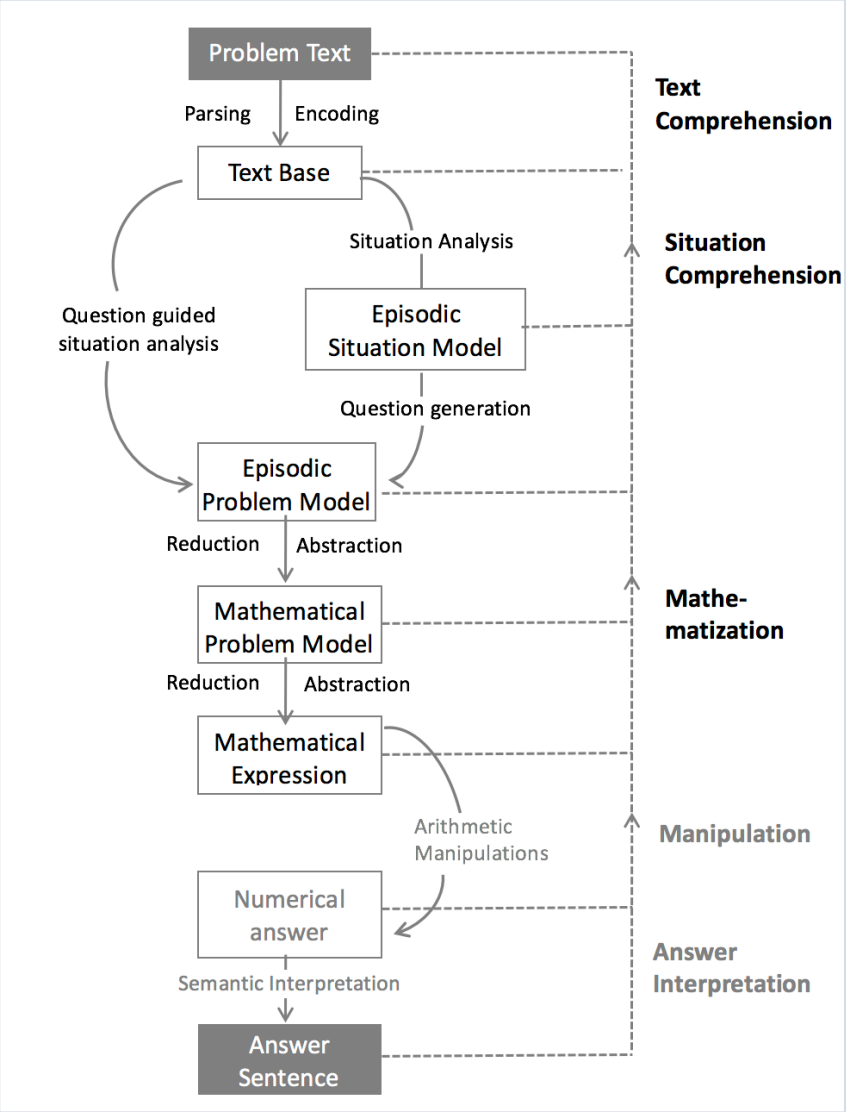


Fig. 1 Model for treating word problems (with slight adaptations) from Reusser (1990)

1.3 Cognitive lab interviews for developing models on students' cognitive processes

In general, the research approach of *cognitive lab interviews* aims at investigating students' processes in laboratory settings in which one or two students work, often with think-aloud protocols. The videotaped processes are then analyzed qualitatively or quantitatively in order to develop models of students' cognitive processes.

Since the 1980s, cognitive psychologists and mathematics educators have studied students' processes with word problems by conducting cognitive lab interviews. One of the most established process models has been developed by Reusser (1990), who describes the process of solving mathematical word problems "as a *strategic process from text to situation to equation*" (Reusser, 1990, p. 481).

His model (see Fig. 1) provides a useful analytic (rather than chronological) distinction of steps from the episodic situation model (an often temporarily structured internal reconstruction of the situation in the text) to the episodic problem model (referring also

DOI: https://doi.org/10.23770/rt1808

to the question of the word problem) and then to the mathematical problem model (in which the situation is structured according to relevant mathematical relations).

Many researchers later adapted Reusser's model in various ways, and these models have proved of value for locating students' difficulties more precisely in the process (Borromeo Ferri, 2006). For language learners, these difficulties were believed to be located mainly in the first steps of text comprehension and situation comprehension (Duarte et al., 2011).

Most documented is the obstacle of superficial comprehension when significant steps in the process model are omitted and instead a direct step is made from the text to a mathematical model. Cognitive lab interviews have identified typical inappropriate compensation strategies:

- Select some numbers and immediately start calculation (e.g., Verschaffel, Greer, & De Corte, 2000)
- Focus only on key words (Nesher & Teubal, 1975)
- Select numbers and chose operation currently taught (e.g., Verschaffel et al., 2000).

However, knowing the process steps and the inappropriate strategies has not yet provided any help in figuring out what students need to learn to overcome them.

1.4 Item analysis for isolating language difficulties for language learners

Two typical research approaches based on item analysis exist that aim at isolating difficulties for language learners in assessment studies:

- *DIF analysis* (Differential Item Functioning). Language biases can be revealed when the frequencies of solutions for language learners are compared with expected difficulties on the Rasch scale (Abedi, 2006; Haag et al., 2013, Prediger et al., 2018).
- *Experimental assessment designs*. Items are systematically modified with respect to specific language demands in order to study the effects on students' frequency of solution (Abedi & Lord, 2001; Haag, Heppt, Roppelt, & Stanat, 2015).

It is interesting to see that most of these studies could not isolate specific language difficulties such as unfamiliar vocabulary or complex sentence structures. Instead, the results were similar for simplified and normal test items, and the DIF-values were difficult to trace back to isolated language features. Thus, the crucial language demands seem to be difficult to capture using isolated language features.

1.5 Summarized state of research

When the research journey of the MuM research group started, it was able to build upon the following state of previous research:

- Language learners have disadvantages in tests compared to students with higher language proficiency; some researchers assume that word problems pose specific difficulties for them.
- Enhancing general reading proficiency does not have an impact on mathematics achievement, so more focused analysis of difficulties is required.

DOI: https://doi.org/10.23770/rt1808

- The greatest difficulties with word problems are located in text comprehension and situation comprehension, for which students' inappropriate compensation strategies are identified.
- Several attempts at item analysis have not succeeded in isolating language features that make test items more difficult for language learners.

However, this research has not provided an empirical foundation for questions such as:

- *What* exactly do students need to learn in order to overcome their difficulties with word problems?
- *How* can they learn it?

As will be shown, this WHAT-question can also be treated as an empirical question (van den Heuvel Panhuizen, 2005).

2 From identifying difficulties to supporting learners in topic-specific ways: a research journey focusing on design research

Based on the state of the research reported in Section 1, the MuM research group in Dortmund applied different research approaches to pursuing these two questions. These approaches are presented in the following subsections.

2.1 Corpus linguistic analysis: potential obstacles are not necessarily real obstacles

In general, the *corpus linguistic analysis* research approach aims at specifying typical (lexical or syntactical) language features by comparing different corpora of written or spoken text. It is an often-recommended approach for specifying language demands (e.g., Bailey, 2007).

With respect to difficulties with mathematical word problems, a corpus linguistic analysis was conducted in the project MuM-Percentages, which compared four textbook corpora with a corpus of newspapers and a corpus of exam items (Niederhaus, Pöhler, & Prediger, 2016). It enabled identification of various potential lexical and syntactical obstacles, e.g., complex prepositional or participle syntactical constructions and a limited lexical overlap between textbooks and exam items (ibid.).

The main outcome of the corpus analysis is a list of language features that frequently appear in the different corpora. However, as this corpus analysis only considered the texts without taking into account the students' solutions, it cannot determine whether these potential obstacles really pose difficulties for students' understanding. This calls for item analysis or cognitive lab interviews.

DOI: https://doi.org/10.23770/rt1808

2.2 Large-scale assessments, item analyses, and cognitive lab interviews to specify obstacles

Large-scale assessments and item analysis in the project MuM-Exam, which examined a high-stakes central exam for Grade 10 in Nordrhein-Westfalen, replicated existing findings (see Sections 1.1 and 1.2): In a high-stakes test for 1,495 tenth graders, language proficiency was the factor with the strongest connection to mathematics achievement. The DIF analysis could not isolate specific language features that underprivilege students with low language proficiency (Prediger et al., 2018, replicating Ufer et al., 2013).

The item analysis in the project MuM-Exam showed that some items with reading obstacles were difficult for nearly all students, but no simple pattern of isolated language features could be determined. In terms of the specific difficulties of language learners, the items with significant DIF values for language learners contained no reading obstacles but only higher conceptual demands or processual demands (Prediger et al., 2018). This means that for students with low language proficiency, acquiring conceptual understanding and higher process skills is even more crucial than reading comprehension (Prediger, in press). However, reading comprehension can be an obstacle for all students. These quantitative findings were validated by cognitive lab interviews in which 47 tenth graders' solution processes for the same items were analyzed qualitatively: Again, reading obstacles were constantly accompanied by conceptual obstacles.

These results called for replication in a second assessment study with another age group: In the project MuM-Percentages, a test was conducted with 308 seventh graders. The items consisted of percentage problems of three different types (Find the amount, find the base, and find the base after reduction) and three different formats (text format, which reflected typical word problems; visual format, which presented the problem structure visually in a percent bar; and pure format, which directly addressed the problem structure using the technical terms). Students' scores were analyzed statistically using a cognitive diagnosis model (Pöhler, George, Prediger, & Weinert, 2017). For the familiar problem types, the items in text format had the lowest frequency of solution, so this again confirmed that word problems can pose difficulties. However, the common assumption that language learners' difficulties can be traced back to reading difficulties (Duarte et al., 2011) was deconstructed: the difference between language learners' frequency of solving items in pure format and text format is the same as for students with high language proficiency.

Summing up, these studies in two different MuM-projects show that:

- Reading proficiency for word problems should be enhanced for all students with respect to their comprehension strategies.
- For language learners, an additional focus must be set on conceptual understanding of the mathematical concepts.

Although the WHAT-question could be narrowed a bit, these studies still do not give sufficient insight into which comprehension strategies are to be fostered.

DOI: https://doi.org/10.23770/rt1808

2.3 Design experiments for specifying the learning content: revealing target strategies for mathematical word problems

In general, design experiment studies aim at developing and investigating teaching-learning arrangements in order to understand what is to be learned and how it can be learned (Cobb et al., 2003; Gravemeijer & Cobb, 2006). They can be conducted in classroom settings or laboratory pair settings; the latter is appropriate as long as the specification of learning contexts is aimed at, while the former is more suitable to grasping ecological validity of classroom conditions. Design experiments differ from cognitive lab interviews by their interventionist rather than diagnostic character (Cobb et al. 2003).

In the concrete research context, design experiments have been applied to identify successful comprehension strategies as the crucial learning content. Whereas non-successful strategies for word problems have already been well identified (see Section 1.3), astonishingly little was known about the comprehension strategies adopted in successful processes (with the exception of Reusser, 1994, mentioned below). Thus, in spite of 30 years of research on word problems, it has not been clear what students had to learn.

In the MuM-project, we again started again with cognitive lab interviews for determining which comprehension strategies can support language learners in solving word problems. However, due to students' general inhibitions about immersing themselves in the process of solving word problems, this research approach proved not to be efficient for the research question and only surface strategies could be found.

Thus, we followed Gravemeijer and Cobb (2006) in transferring Kurt Lewin's slogan for design experiments: "If you want to truly understand something, try to change it." (p. 46). In the projects MuM-Reading (Dröse & Prediger, in press) and MuM-Algebra (Prediger & Krägeloh, 2015b), design experiments were conducted in pair settings in order to specify successful reading strategies. Only fostering students' processes allowed their inhibition to be overcome, and strategies for solving word problems became apparent. By trying to modify them, we learned to distinguish the relevant and powerful strategies from the ones that were not. The following six target strategies were identified for algebraic word problems (Prediger & Krägeloh, 2015a; some already identified by Reusser, 1994):

(S1) Find relevant information

(S2) Focus on information with its meaning, e.g. quantities with the units

(S3) Focus on relations connecting pieces of information, e.g., relations between quantities

(S4) Decompose in several steps and work forward

(S5) Work forward and backward in order to decompose in several steps towards a goal

(S6) Get started anyway with first ideas in order to overcome affective obstacles

Whereas the first strategy is a subject-independent reading strategy that applies for many genres, the other strategies are typical for the subject-specific genre of

DOI: https://doi.org/10.23770/rt1808

mathematical word problems. Strategies S4 are S5 are required when the word problems require multiple steps to be sequenced properly. S6 is required to overcome the inhibition of not daring to start.

2.4 Design experiments developing strategic scaffolding for comprehension strategies

How can strategies be taught and learned? Strategies cannot simply be told by the teacher to the students, e.g., by giving them a list of strategies. Instead, the design principle of strategic scaffolding was transferred to this subject-specific context (Prediger & Krägeloh, 2015a). While the WHAT-questions must be answered topic specifically, the HOW-questions can be treated by transfer from other fields.

The general notion of "scaffolding" is widely used in language education and cognitive science. A scaffold is a "temporary framework to support learners when assistance is needed and is removed when no longer needed," i.e., the so-called fading out (Lajoie, 2005, p. 542). Hannafin et al. (1999) define strategic scaffolding as "guides in analyzing and approaching learning tasks or problems" by focusing "on approaches for identifying and selecting needed information" (p. 131, 133). The aim of strategic scaffolds is to "trigger a series of related strategies" (ibid, p. 134) and was suggested by Reusser (1994) for comprehension strategies for word problems.

In the projects MuM-Algebra (Prediger & Krägeloh, 2015a), MuM-Percentages (Pöhler & Prediger, 2015), and MuM-Reading (Dröse & Prediger, in press), different scaffolds were developed and optimized for Grades 8, 7, and 5. Design experiments were conducted in pair settings and classroom settings in order to investigate typical learning processes when students get familiar with the scaffold, work with it, and fade it out. The design had to be optimized in order to be flexible enough for different individual preferences, this was crucial for assuring adaptivity.

2.5 Design experiments for developing conceptual understanding

Whereas the strategy trainings were crucial for most students, the language learners had additional difficulties in developing the conceptual understanding of the involved mathematical concepts. In order to support language learners in developing conceptual understanding, topic-specific design research was required to

- identify which language demands hindered language learners' acquisition of the specific mathematical concept and
- design interventions that combine the conceptual learning trajectory with a language learning trajectory offering those words and phrases required for developing conceptual understanding.

Again, for the HOW-questions, the design could build upon subject-independent design principles such as the design principles of

- discursive activation (Erath, 2017);
- relating registers and representations that ask students to permanently relate the graphical, symbolical, and verbal presentations, including everyday register, school academic register, and technical register (Moschkovich, 2013); and

DOI: https://doi.org/10.23770/rt1808

- macro-scaffolding, according to which language learning trajectories must be sequenced from the everyday via the school academic to the technical registers and intertwined with the conceptual learning trajectories (Pöhler & Prediger, 2015).

However, in order to answer the WHAT-questions (Which exact language demands hinder language learners' development of conceptual understanding?), topic-specific design experiments were again required, because trying to change the students' conceptual understanding was the precondition to identifying their language obstacles (following Lewin's slogan from Section 2.3). Even more than for comprehension strategies in word problems, this research had to be topic specific and focused on language production rather than language reception.

The qualitative in-depth analysis of students' learning processes towards developing conceptual understanding revealed language demands on different levels:

- **On the discourse level**, the language demands for developing conceptual understanding can be specified by the discourse practice of *explaining meanings*: Whereas language learners succeed in participating in the discourse practice of *reporting procedures*, their participation in explaining meanings is often dangerously limited (Erath, 2017; Huang, Normandia, & Greer, 2005).
- **On the sentence level**, the discourse practice of explaining meanings does not require specific syntactical constructs, but phrases expressing coherence between arguments such as "we can see this connection in the figure by…."
- **On the word level,** in particular, students with low language proficiency require specific support in extending their lexicon in order to participate in the discourse practices of explaining meanings: rather than the vocabulary usually focused on in the technical lexicon ("percentage," "amount," "base," "multiply"), it is the so-called meaning-related vocabulary that partly stems from the school academic register that is most crucial for explaining meanings ("the part of a whole," "the old price," "the money to be saved"). For each mathematical topic in view, this meaning-related vocabulary must be specified in order to provide focused language support for the most relevant discourse practice. So far, these have been specified for the mathematical concepts of fractions (Prediger & Wessel, 2013), percentages (Pöhler & Prediger, 2015), variables as generalizers (Prediger & Krägeloh, 2015b), and functions (Prediger & Zindel, 2017).

The specification of relevant language demands in the learning processes was achieved by qualitatively investigating learning processes initiated in design experiments on content- and language-integrated teaching-learning arrangements. A second outcome of these design research studies has been the teaching-learning arrangements themselves, optimized to support language learners' conceptual development. The iterative approach also allowed the researchers to successively narrow the qualitative investigation on different topic-specific foci, e.g., relevance of relating registers (Prediger & Wessel, 2013) and the micro-processes of negotiating meanings (Prediger & Krägeloh, 2015b).

DOI: https://doi.org/10.23770/rt1808

2.5 Intervention studies providing empirical evidence for efficacy and effectiveness

In general, intervention studies with quasi-randomized controlled trials are conducted for teaching-learning arrangements with the aim of showing the efficacy of certain design principles in clinical laboratory settings (e.g., conducted in small groups and with researchers as teachers) and then showing their effectiveness in field studies (usually in classroom settings with the regular teachers under more ecologically valid conditions; Burkhardt & Schoenfeld, 2003). In both cases, pre-tests, post-tests, and follow-up tests are conducted with the intervention group (working with the teaching-learning arrangement) and the control group (which is comparable in the pre-test achievements and all relevant control variables but being taught with other interventions).

In the project MuM-Fractions, it was possible to provide empirical evidence for the efficacy of fostering language learners' conceptual understanding in a laboratory study (Prediger & Wessel, 2013): The seventh graders working in a content- and language integrated teaching-learning arrangement fostering conceptual understandings of fractions had significantly higher learning gains than the students in the control group.

In contrast, the empirical evidence was more difficult to achieve for fostering students' comprehension strategies: In the project MuM-Reading, a teaching-learning arrangement on developing comprehension strategies with strategic scaffolding was tested in a field study with four classes. The intervention group had slightly higher learning gains than the control group, but the difference was not significant. We see two major reasons why only tendencies could be found that still did not have significance:

- The focus on comprehension strategies alone might be too isolated; it must be combined with either a more consequent focus on conceptual understanding or on the syntactical complexities that appear in the word problems.
- The teaching-learning arrangement should have first been evaluated in a laboratory study under more controlled conditions; it could then be further improved with respect to robustness in field studies so that more teachers can work with it in the intended ways (Burkhardt & Schoenfeld, 2003).

For the first reason, an intervention study in the project MuM-Percentages can provide some plausibility: In this project, a teaching-learning arrangement was designed in which the development of conceptual understanding was combined with a scaffolding of comprehension strategies, both based on the subject matter of the percent bar (Pöhler & Prediger, 2015). This combined teaching-learning arrangement even achieved empirical evidence in an effectiveness field study with 108 students: The ANCOVA shows that after 15 sessions of intervention, the intervention group significantly outperformed the control group (with comparable pre-knowledge) and medium effect sizes. This shows that, by using a suitable scaffold, teachers can foster language learners' achievement in percentage word problems when they foster both the development of conceptual understanding and comprehension strategies (Pöhler, Prediger, & Neugebauer, 2017).

Summing up the investigation of effects, qualitative studies have shown potential learning gains that have not yet all been robust in intervention studies. Whereas fostering conceptual understanding showed significant learning gains in laboratory and

DOI: https://doi.org/10.23770/rt1808

field studies, the field intervention study fostering only comprehension strategies without also fostering conceptual understanding could not yet provide significant learning gains. Although the studies cannot immediately be compared, this can be interpreted as a hint that the fostering of comprehension strategies might require further investigation before being adaptive and focused enough for all learners.

3 Meta-perspective: comparing and combining research approaches

The research journey of the MuM research group presented in Section 2 is of course one among many research journeys. However, it carries a central message: In order to bring innovations into school, much more than a new syllabus demanding the fostering of language learners in every subject is required. The often-underestimated research challenge is to specify what exactly has to be learned before asking how it can be learned. This WHAT-question requires subject-specific and even topic-specific research that combines different research approaches, because no single research approach can provide the necessary subject-specific and topic-specific empirical foundations for teaching-learning arrangements. Even if there are many more research approaches than those applied in the MuM research journey, looking back on these allows us to illuminate how they complement each other:

- **Large-scale assessments** with written tests and large samples can show the necessity of an innovation by determining the correlative connection between subject-matter achievement and language proficiency. However, they cannot unfold explanative power or underpin the teaching approaches in classrooms unless they are complemented by item analysis.
- **Corpus linguistic studies and linguistic item analysis** reveal potential obstacles of the texts themselves, but they must be complemented by empirical analysis in order to study which potential obstacles really pose difficulties for students of different language backgrounds. The repeated empirical findings that no language feature can be isolated as a singular difficulty in the DIF analysis shows that language is always embedded in complex structures.
- **Cognitive lab interviews** that investigate students' processes provide cognitive models for students' solution processes and identify the obstacles which really pose difficulties. The findings that many language learners fail due to their limited conceptual understanding of mathematical concepts made it necessary to investigate the learning processes in learning process studies.
- **Design research studies** can not only place learning process studies in regular classrooms (where some learning opportunities do not even appear), but can also design the learning situation first that can then be investigated. The iterative interplay of (re)design and design experiments allow researchers to specify the main language demands during the learning processes and then to successively optimize the teaching-learning arrangement that supports language learners in acquiring the relevant language aspects. These comprise subject-specific but topic-independent target comprehension strategies as well as topic-specific phrases and notions in the

DOI: https://doi.org/10.23770/rt1808

meaning-related lexicon required for the discourse practice of explaining meanings. The qualitative investigation of initiated learning processes first provides qualitative indications for students' learning gains that must then be complemented by quantitative evidence.

- **Quasi-randomized intervention studies** in laboratory and field settings can show the efficacy or effectiveness of interventions when the intervention group has significantly higher learning gains than the control group. These kinds of studies should not be started too early; two examples of non-successful short cuts have shown why the teaching-learning arrangements must first be empirically founded by qualitative research in other approaches. Even in our own studies, one field study failed because the robustness of the teaching-learning arrangement was not sufficiently assured before going to complex field conditions. Additionally, isolating comprehension strategies from sensitizing for other language features and from conceptual understanding was not successful. The teaching-learning arrangement on percentages, which combined both, could even succeed under field conditions. Again, we see that research journeys should not be shortened and that language features must not be treated in isolated ways but in complex approaches.

Summing up, none of the applied research approaches could have treated the relevant research questions alone. Most explanative power and empirical foundation for classroom practices is gained by successively combining and systematically intertwining studies rather than standing them next to each other (Burkhardt & Schoenfeld, 2003). Of course, many other research approaches could also contribute to the challenge and will need to be taken into account in future research.

However, the limited research journey already specifically shows how the explanative power increases with the topic specificity of the research. Rather than posing only generic, subject-independent questions, a lot was to be learned by treating them for a specific subject and specific topics within the subject. This might contribute to explaining the findings by Seidel and Shavelson (2007) of domain-specific teaching approaches being more effective for learning gains than domain-independent ones.

References

Abedi, J. (2006). Language issues in item-development. In S. M. Downing & T. M. Haldyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah: Erlbaum.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234.

Bailey, A. L. (2007). *The language demands of school: Putting academic English to the test*. New Haven: Yale.

Bikner-Ahsbahs, A., Knipping, C., & Presmeg, N. (Eds.). (2016). *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods*. Dordrecht: Springer.

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *ZDM Mathematics Education, 38*(2), 86–95.

Burkhardt, H., & Schoenfeld, A. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher, 32*(9), 3–14.

DOI: https://doi.org/10.23770/rt1808

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9–13.

Duarte, J., Gogolin, I., & Kaiser, G. (2011). Sprachlich bedingte Schwierigkeiten von mehrsprachigen Schülerinnen und Schülern bei Textaufgaben. In S. Prediger & E. Özdil (Eds.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (pp. 35–53). Münster: Waxmann.

Dröse, J. & Prediger, S. (in press). Scaffolding für fachbezogene textsortenspezifische Lesestrategien. To appear in B. Ahrenholz, S. Jeuk, B. Lütke, J. Paetsch, & H. Roll (Eds.), *Fachunterricht, Sprachbildung und Sprachkompetenzen.* Berlin: De Gruyter.

Erath, K. (2017). Talking about conceptual knowledge. In B. Kaur, W.K. Ho, T.L. Toh, & B.H. Choy (Eds*.), Proceedings of PME 41* (Vol. 2, pp. 321–328). Singapore: PME.

GFD – Gesellschaft für Fachdidaktik (2016). *Diskussionspapier der GFD - Formate Fachdidaktischer Forschung.* http://www.fachdidaktik.org/Home.html.

Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. v. d. Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). London: Routledge.

Haag, N., Heppt, B., Roppelt, A., & Stanat, P. (2015). Linguistic simplification of mathematics items. *European Journal of Psychology of Education, 30*(2), 145–167.

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics. *Learning and Instruction, 28*(0), 24-34

Hagena, M., Leiß, D., & Schwippert, K. (2017). Using reading strategy training to foster students' mathematical modelling competencies. *Eurasia Journal of Mathematics Science and Technology Education, 13*(7b), 4057–4085.

Hannafin, M., Land, S., & Oliver, K. (1999). Open Learning Environments. In C.M. Reigeluth (Ed.), *Instructional-Design Theories and Models* (Vol. 2, pp. 115–140). Mahwah: Lawrence Erlbaum.

Hellmich, F., & Förster, S. (2015). Transfereffekte eines wortschatzbasierten Lesetrainings auf die Bearbeitungsqualität mathematischer Textaufgaben. In D. Blömer, M. Lichtblau, A.-K. Jüttner, K. Koch, M. Krüger, & R. Werning (Eds.), *Perspektiven auf inklusive Bildung* (pp. 255–260). Wiesbaden: Springer.

Huang, J., Normandia, B., & Greer, S. (2005). Communicating Mathematically: Comparison of Knowledge Structures in Teacher and Student Discourse in a Secondary Math Classroom. *Communication Education, 54*(1), 34–51. doi:10.1080/14613190500077002.

Lajoie, S. (2005). Extending the scaffolding metaphor. *Instructional Science*, *33*(5-6), 541–557.

Moschkovich, J. (2013). Principles and Guidelines for Equitable Mathematics Teaching Practices and Materials for English Language Learners. *Journal of Urban Mathematics Education, 6*(1), 45–57.

Nesher, P. & Teubal. E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics, 6*(1), 41–51.

Niederhaus, C., Pöhler, B., & Prediger, S. (2016). Relevante Sprachmittel für mathematische Textaufgaben. In E. Tschirner, O. Bärenfänger, & J. Möhring (Eds.). *Deutsch als fremde Bildungssprache* (pp. 135–162). Stauffenberg: Tübingen.

OECD. (2007). PISA 2006. Science competencies for tomorrow's world (Vol. 2). Paris: OECD.

Paetsch, J., Felbrich, A., & Stanat, P. (2015). Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie, 29*(1), 19–29.

Pöhler, B., & Prediger, S. (2015). Intertwining lexical and conceptual learning trajectories. *Eurasia Journal of Mathematics, Science, & Technology Education, 11*(6), 1697–1722.

Pöhler, B., George, A. C., Prediger, S., & Weinert, H. (2017). Are word problems really more difficult for students with low language proficiency?. *IEJME – Mathematics Education, 12*(7), 667–687.

DOI: https://doi.org/10.23770/rt1808

Pöhler, B., Prediger, S., & Neugebauer, P. (2017). Content- and language integrated learning: A field experiment for percentages. In B. Kaur, W. K. Ho, T. L. Toh, & B. H. Choy (Eds.), *Proceedings of PME 41* (Vol. 4, pp. 73–80). Singapore: PME.

Prediger, S. (in press). Welche Forschung kann Sprachbildung im Fachunterricht empirisch fundieren? Ein Überblick zu mathematikspezifischen Studien und ihre forschungsstrategische Einordnung. To appear in B. Ahrenholz, S. Jeuk, B. Lütke, J. Paetsch, & H. Roll (Eds.), *Fachunterricht, Sprachbildung und Sprachkompetenzen.* Berlin: De Gruyter.

Prediger, S., & Krägeloh, N. (2015a). Low achieving eighth graders learn to crack word problems: a design research project. *ZDM Mathematics Education, 47*(6), 947–962. doi:10.1007/s11858-015-0702-7.

Prediger, S., & Krägeloh, N. (2015b). The epistemic role of languages while constructing meaning for the variable as generalizers. In A. Halai & P. Clarkson (Eds.). *Teaching and Learning Mathematics in Multilingual Classrooms* (pp. 89–108). Rotterdam: Sense.

Prediger, S., & Wessel, L. (2013). Fostering German language learners' constructions of meanings for fractions. *Mathematics Education Research Journal, 25*(3), 435–456.

Prediger, S., & Zindel, C. (2017). School Academic Language Demands for Understanding Functional Relationships: A Design Research Project on the Role of Language in Reading and Learning. *EURASIA Journal of Mathematics, Science and Technology Education, 13*(7b), 4157–4188. doi: 10.12973/eurasia.2017.00804a.

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E. & Benholz, C. (2018). Language proficiency and mathematics achievement– Empirical study of language-induced obstacles in a high stakes test, the central exam ZP10. Translation of the German article Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10, *Journal für Mathematik-Didaktik*, 36(1) 2015, 77-104. English doi: 10.1007/s13138-018-0126-3.

Reusser K. (1990). From Text to Situation to Equation. In H. Mandl, E. De Corte, N. Bennet, & H.F. Friedrich (Eds.), *Learning and Instruction* (Vol. II, pp. 477–498). New York: Pergamon.

Reusser, K. (1994). Tutoring Mathematical Text Problems: From Cognitive Task Analysis to Didactic Tools. In S. Vosniadou, E. de Corte, & H. Mandl (Eds.), *Technology-Based Learning Environments* (Vol. 137, pp. 174–182). Berlin: Springer.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research, 77*(4), 454–499.

Thürmann, E., Vollmer, H. J., & Pieper, I. (2010). *Language(s) of Schooling: Focusing on vulnerable learners* (Vol. 2). Straßbourg: Council of Europe.

Ufer, S., Reiss, K., & Mehringer, V. (2013). Sprachstand, soziale Herkunft und Bilingualität: Effekte auf Facetten mathematischer Kompetenz. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Eds.), *Sprache im Fach* (pp. 167–184). Münster: Waxmann.

US Congress (2001). *No child left behind act*. Washington D.C.

Van den Heuvel-Panhuizen, M. (2005). Can scientific research answer the 'what' question of mathematics education? *Cambridge Journal of Education, 35(1)*, 35–53.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.

Prof. Dr. Susanne Prediger

Full Professor for Mathematics Education at the Institute for Development and Research in Mathematics Education in Dortmund, Germany.

DOI: https://doi.org/10.23770/rt1808