

---

**Towards the Implementation of a Digital Learning Program in  
Mathematics: Determinants of Practice Behavior and Effects on  
Performance, Self-Concept, and Anxiety**

---

**DISSERTATION**

in fulfillment of the requirements for the doctoral degree at the Faculty of  
Arts and Humanities at Kiel University

submitted by  
Anna Katharina Maria Hilz, Dipl.-Psych.

Kiel,  
February 2024

First examiner

Prof. Dr. Olaf Köller

Second examiner

Prof. Dr. Uta Klusmann

Date of the oral examination

28.06.2024

Approved for printing by the second Vice Dean for  
Studies and Examinations, Prof. Dr. Elmar Eggert

13.01.2025

# DANKSAGUNG

Zunächst möchte ich mich herzlich bei Olaf Köller, Karin Guill und Janina Roloff-Bruchmann bedanken, die mir ermöglichten am IPN im Projekt IMPULSE zu promovieren. Weiterhin danke ich Olaf Köller für die Begutachtung dieser Arbeit. Ebenso möchte ich mich bei Uta Klusmann bedanken, einerseits für die Begutachtung meiner Dissertation, andererseits aber auch für ihr wertvolles Feedback in den letzten Jahren. Außerdem danke ich Gráinne Newcombe für die wertvolle Sprachkorrektur. Mein ganz besonderer Dank gilt Karen Aldrup für die Betreuung meiner Promotion in den letzten 2 Jahren. Durch deine unglaublich wertschätzende Art und dein wahnsinnig lehrreiches Feedback hast du mich maßgeblich in meinem kreativen und kritischen Denken gefördert. Ich bin sehr dankbar für alle Dinge, die ich von dir lernen durfte.

Insbesondere möchte ich mich bei meinen IMPULSE-Kolleginnen Lisa und Katharina für den lehrreichen Austausch und die wertvolle Unterstützung in den letzten Jahren bedanken. Ich bin sehr dankbar für die schönen Erfahrungen, die wir gemeinsam sammeln durften und die vielen Momente, in denen wir gelacht haben. Weiterhin danke ich allen Menschen, die den WiPa als Ort der Inspiration und des wertvollen Austauschs geprägt haben. Mein besonderer Dank gilt dabei Luca, Ute, Lucas und Livia, nicht nur für unsere aufregenden und erlebnisreichen Konferenzreisen, sondern auch für eure uneingeschränkte Unterstützung. Ihr habt vor allem die letzten 2 Jahre sehr besonders gemacht.

Weiterhin danke ich meiner Freundin Merle. Du bist immer für mich da, wenn ich emotionale Unterstützung brauche und dein tiefes Vertrauen in mich und meine Fähigkeiten ehrt mich sehr. Von ganzem Herzen möchte ich mich außerdem bei meinen Eltern bedanken. Eure bedingungslose Unterstützung gibt mir Kraft und den nötigen Rückhalt bei der Verwirklichung all meiner kleinen und großen Pläne.

Mein größter Dank gebührt allerdings meinem Mann Tim. Timmy, ich danke dir dafür, dass du an mich glaubst und mir immer eine Stütze bist, wenn es manchmal schwierig wird. Deine unerschütterliche Verbundenheit gibt mir die Kraft über mich hinauszuwachsen.



# ABSTRACT

Acquiring high mathematical competence throughout their academic careers is central to students' successful participation in modern society (OECD, 2023). However, countries such as Germany are facing a worryingly strong decline in students' mathematical competence, as well as a widening achievement gap (OECD, 2023; Stanat et al., 2022). To address these challenges, it has been suggested that the integration of computer-based math learning programs into education could be a promising solution because such programs are cost-effective and easily implementable. However, studies that investigated the effects of such learning programs on students' performance revealed heterogeneous result patterns (e.g., Byun & Joung, 2018; Tokac et al., 2019), and overall outcomes have been more disappointing than promising (Pellegrini et al., 2021). Nonetheless, it is probably not the programs themselves that lead to the disappointing results but the way in which the effects are investigated; previous research usually compared the program provision in an experimental group to a waitlist control group (e.g., see Bai et al., 2012; Chang et al., 2015; Hung et al., 2014; Shin et al., 2012). Thereby, a factor that could possibly explain the heterogeneous effects was largely not considered, namely, students' actual practice behavior, even though there is evidence that regular and sustained practice is essential for learning success (Gerbier & Toppino, 2015; Toppino & Gerbier, 2014) and even though the process measures often stored within learning programs offer a huge potential to economically and objectively measure students' practice behavior (Azevedo, 2015).

According to the supply-use model (Helmke & Weinert, 1997; Seidel, 2014), two central assumptions can be made. First, as the mere provision of a program does not automatically mean that all students practice to the same degree, one could suggest that individual student characteristics are associated with their practice behavior. Surprisingly, no research so far has investigated this suggestion within the context of math learning programs. Second, varying practice behavior may explain why some students achieve learning success with math learning programs while others do not. In this sense, students could not only benefit regarding their performance but also in terms of achievement-related constructs such as academic motivation and emotions. Indeed, some studies provided first evidence that students' process-measured practice behavior (i.e., measured as the number of tasks completed or the amount of time spent with the program; see Haleva et al., 2021; Hassler Hallstedt et al., 2018; B. R. J. Jansen et al., 2013; Spitzer, 2022) is related to potential performance improvements. Still, very little is known about how the provision of a math learning program and, in particular, how students' practice

behavior relates to students' academic motivation and emotions (see Dondio et al., 2023; Fadda et al., 2022; B. R. J. Jansen et al., 2013).

To narrow these gaps in the research, the aim of this dissertation with its focus on students' practice behavior was twofold. A supply-use model was adapted to the context of math learning programs in order to investigate (1) links between students' individual characteristics (i.e., prior math knowledge, personality, affective-motivational preconditions, and family background) and their practice behavior, and (2) students' practice behavior and its relation to performance, academic motivation, and emotions (i.e., math self-concept and math anxiety). To address these research questions, three empirical studies and two additional replication analyses were conducted with the adaptive arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014; German: Prowise Learn). The studies were based on two implementation periods, varying between 22 and 45 weeks, in which the program was provided to fifth-grade students. Regarding students' practice behavior, all studies took advantage of the objective process measures stored within Math Garden during students' practice. Thus, practice behavior was primarily operationalized by the number of tasks completed by the students.

Addressing the first research question, Study 1 investigated whether individual student characteristics, namely, prior math knowledge, conscientiousness, math self-concept, math anxiety, socioeconomic status, and a migration background predicted students' practice behavior (i.e., the number of tasks completed) with Math Garden over 22 weeks. Based on longitudinal data from 203 students in the first implementation period, multilevel analyses were conducted, as the completed tasks (measured weekly) were clustered within students. To ensure the robustness of the results, Study 1 was replicated with a different sample comprising 890 students who practiced with Math Garden for 45 weeks, that is, in the second implementation period.

Addressing the second research question, Study 2 focused on the effects of Math Garden regarding students' learning outcomes in terms of their addition and subtraction performance, their math self-concept, and their math anxiety over the period of 22 weeks. The study included 376 students from the first implementation period, who belonged either to the experimental group who practiced with Math Garden or to a waitlist control group. Regression analyses were used to examine the effects of the program provision and students' practice behavior (i.e., the number of tasks completed/number of weeks practiced) on the learning outcomes investigated. Study 2 was also replicated with students from the second implementation period. This sample

included 1,093 students, whereby 890 students in the experimental group practiced with Math Garden.

Study 3 connected the assumptions investigated in Study 1 and Study 2 and, thus, contributed to both research questions of this dissertation. Based on the longitudinal data of the 890 students in the second implementation period, Study 3 investigated whether math anxious students avoided practicing with Math Garden (i.e., completed fewer tasks) and, in turn, whether this inhibited them in improving their math performance. Mediation analyses were conducted.

Regarding the relationship between students' individual characteristics and their practice behavior (Research Question 1), two patterns of findings emerged. Both math anxious students and students with a migration background practiced less with Math Garden; however, it remained unclear which factor had more predictive power (Study 1, the replication of Study 1, and Study 3). For all other student characteristics (prior math knowledge, conscientiousness, math self-concept, and socioeconomic status), no statistically significant relationship was found.

With respect to the second research question, that is, the effects on students' learning outcomes, the results did not find any effects of the mere program provision on students' performance. However, analyses regarding the experimental group revealed a significant link between students' practice behavior and their performance improvement (i.e., the number of completed tasks; Study 2, the replication of Study 2, and Study 3). A reverse pattern emerged with regard to the effects on math self-concept. That is, the analyses showed that an increase in students' math self-concept was related to the program provision but, at the same time, appeared to be independent of students' actual practice (Study 2). However, the promising finding of the effect of the program provision on math self-concept was not statistically robust across the two implementation periods (replication of Study 2). Regarding math anxiety, no statistically significant links were found.

This dissertation provides nuanced support for the assumption that practice behavior might play an important role in students' practice in a math learning program. It was able to show that specific individual student characteristics were related to students' practice behavior and, in turn, that varying practice behavior had the potential to explain differences regarding students' performance improvement. However, the link between practice behavior and academic motivation and emotions remains unclear at this point. Practitioners need to ensure that students systematically and consistently practice with a math learning program if students are to benefit from that program regarding their performance. Such support from practitioners

might be especially relevant for math anxious students and students with a migration background.



## CONTENT

<b>Abstract .....</b>	<b>i</b>
<b>Chapter 1 - Introduction.....</b>	<b>11</b>
<b>Chapter 2 - Theoretical and Empirical Background .....</b>	<b>17</b>
2.1 Student Characteristics, Their Practice Behavior, and Their Learning Outcomes ...	18
2.2 Math Learning Programs as a Supply Structure .....	20
2.2.1 Categorizing Math Learning Programs According to Their Aims and Scopes ..	20
2.2.2 Categorizing Math Learning Programs According to Program Complexity .....	22
2.3 Rethinking Students' Engagement and Process Measures .....	27
2.4 The Role of Students' Characteristics in Their Practice Behavior .....	29
2.4.1 Prior Knowledge .....	29
2.4.2 Personality.....	30
2.4.3 Academic Motivation.....	31
2.4.4 Academic Emotions .....	31
2.4.5 Family Background.....	32
2.5 Effects of Practice Behavior on Students' Learning Outcomes .....	33
2.5.1 Performance .....	34
2.5.2 Academic Motivation.....	38
2.5.3 Academic Emotions .....	40
2.6 Research Questions and Studies Conducted .....	41
2.6.1 Study 1 and Additional Replication Analyses: Students' Individual Characteristics as Predictors of Practice Behavior .....	44
2.6.2 Study 2 and Additional Replication Analyses: Effects of Providing and Practice With Math Garden on Students' Learning Outcomes.....	45
2.6.3 Study 3: The Mediating Role of Practice Behavior Regarding the Relationship Between Math Anxiety and Math Performance.....	45
2.7 References.....	46
<b>Chapter 3 - Study 1: The Relationship Between Individual Characteristics and Practice Behavior Within an Adaptive Arithmetic Learning Program.....</b>	<b>65</b>
3.1 Introduction.....	67
3.1.1 Individual Characteristics as Predictors of Practice Behaviour .....	68
3.1.2 The Present Study .....	71
3.2 Methods .....	72
3.2.1 Design and Procedure .....	72
3.2.2 Sample.....	73

3.2.3 Arithmetic Learning Program .....	73
3.2.4 Measures .....	74
3.2.5 Data Analysis .....	76
3.3 Results.....	77
3.3.1 Descriptives and Preliminary Analyses .....	77
3.3.2 Relationship Between Individual Characteristics and Practice Behaviour .....	79
3.4 Discussion .....	81
3.4.1 Individual Characteristics as Predictors of Practising With an Arithmetic Learning Program .....	81
3.4.2 Strengths and Limitations .....	83
3.4.3 Theoretical and Practical Implications.....	84
3.5 Conclusions.....	86
3.6 References.....	87
3.7 Supplementary Tables.....	97
<b>Chapter 4 - Additional Replication Analysis Study 1 .....</b>	<b>99</b>
4.1 Introduction.....	100
4.2 Research Question and Hypotheses .....	100
4.3 Methods .....	101
4.3.1 Design and Procedure .....	101
4.3.2 Sample.....	101
4.3.3 Measures .....	101
4.3.4 Data Analysis .....	103
4.4 Results.....	104
4.4.1 Descriptives and Preliminary Analyses .....	104
4.4.2 Relationship Between Individual Characteristics and Practice Behavior .....	106
4.5 Discussion.....	108
4.6 References.....	110
<b>Chapter 5 - Study 2: How to continue? New Approaches to Investigating the Effects of Adaptive Math Learning Programs on Students' Performance, Self-Concept, and Anxiety .....</b>	<b>113</b>
5.1 Introduction.....	115
5.1.1 Promising Mechanisms of Math Learning Programs .....	116
5.1.2 New Approaches to Investigating the Effectiveness of Math Learning Programs .....	118
5.1.3 The Present Study .....	121
5.2 Materials and Methods.....	123

5.2.1 Design and Procedure .....	123
5.2.2 Sample.....	124
5.2.3 Measures .....	125
5.2.4 Data Analysis .....	127
5.3 Results.....	129
5.3.1 Descriptives and Preliminary Analyses .....	129
5.3.2 Effects of Providing Math Garden .....	133
5.3.3 Effects of Practiced Tasks.....	133
5.3.4 Effects of Practiced Weeks .....	133
5.4 Discussion.....	135
5.4.1 Effects of Math Garden on Math Performance .....	136
5.4.2 Effects of Math Garden on Math Self-Concept .....	138
5.4.3 Effects of Math Garden on Math Anxiety .....	139
5.4.4 Limitations .....	140
5.4.5 Practical Implications.....	142
5.5 Conclusions.....	143
5.6 References.....	144
5.7 Supplementary Tables.....	155
<b>Chapter 6 - Additional Replication Analysis Study 2 .....</b>	<b>161</b>
6.1 Introduction.....	162
6.2 Research Question and Hypotheses .....	162
6.3 Methods .....	163
6.3.1 Design and Procedure .....	163
6.3.2 Sample.....	163
6.3.3 Measures .....	164
6.3.4 Data Analysis .....	165
6.4 Results.....	166
6.4.1 Descriptives and Preliminary Analyses .....	166
6.4.2 Effects of Providing Math Garden and of Practice Behavior .....	170
6.5 Discussion.....	173
6.6 References.....	174
<b>Chapter 7 - Study 3: Tracing Students' Practice Behavior in an Adaptive Math Learning Program: Does it Mediate the Math Anxiety–Performance Link? .....</b>	<b>177</b>
7.1 Introduction.....	179
7.1.1 Defining Math Anxiety .....	180
7.1.2 How is Math Anxiety Linked to Math Performance?.....	180

7.1.3 Conceptualizing Avoidance Behavior .....	181
7.1.4 The Potential of Process Measures .....	182
7.1.5 The Present Study .....	182
7.2 Methods .....	184
7.2.1 Design and Procedure .....	184
7.2.2 Sample.....	185
7.2.3 Measures .....	185
7.2.4 Data Analysis .....	186
7.3 Results.....	187
7.3.1 Descriptives and Preliminary Analyses .....	187
7.3.2 The Relationship Between Math Anxiety, Avoidance Behavior, and Performance .....	190
7.4 Discussion.....	192
7.4.1 Math Anxiety and Avoidance Behavior.....	193
7.4.2 The Potential of Process Measures .....	194
7.4.3 Limitations .....	195
7.4.4 Implications.....	196
7.4.5 Conclusions.....	197
7.5 References.....	197
7.6 Supplementary Tables.....	205
<b>Chapter 8 - General Discussion.....</b>	<b>211</b>
8.1 RQ1: The Associations Between Individual Student Characteristics and Students' Practice Behavior With a Math Learning Program.....	214
8.1.1 Math Anxiety and Students with a Migration Background .....	214
8.1.2 Prior Knowledge, Conscientiousness, Math Self-Concept, and SES .....	217
8.2. RQ2: Effects of Program Provision and Students' Practice with a Math Learning Program on Students' Performance, Academic Motivation, and Emotions .....	219
8.2.1 Performance .....	220
8.2.2. Math Self-Concept .....	223
8.2.3 Math Anxiety .....	225
8.3 Theoretical and Practical Implications.....	226
8.3.1 Theoretical Implications .....	226
8.3.2 Practical Implications.....	228
8.4 Limitations .....	231
8.5 Future Research .....	232
8.6 Conclusions.....	235

8.7 References.....	237
<b>Appendix .....</b>	<b>249</b>
Deutschsprachige Zusammenfassung .....	250



# 1

## INTRODUCTION

### 1 Introduction

In recent times, the widespread digitalization in many areas of life, the presence of data that influence personal decisions, and the need to tackle major societal challenges such as climate change, national debt, population growth, and the spread of pandemic diseases have fundamentally increased the relevance of being a mathematically competent individual (OECD, 2023). This is because all of these (critical) issues have a quantitative component to them. Thus, mathematics is undeniably one of the most crucial competences that students acquire throughout their academic career as it prepares them to act as thoughtful, engaged, and reflected citizens in light of the above-mentioned 21st-century global challenges (Van Aalderen-Smeets et al., 2019). Across the lifespan, students with high mathematical competence are more likely to successfully complete their education and pursue a successful career (Bailey et al., 2014; Barbieri et al., 2021; Duncan et al., 2007; OECD, 2019), whereas students with insufficiently developed mathematical competence are more prone to school dropout, unemployment, and, therefore, an increased likelihood of developing mental health disorders (Bynner & Parsons, 2006). Against this background, the poor results of the latest PISA study are very worrying, especially for Germany (OECD, 2023). Overall, the average mathematical competence of young people in many OECD countries has decreased compared to the previous assessment in 2018. However, the decline in Germany was above average, so the results are even below those of the first PISA shock in 2000. Thus, Germany is currently facing a second PISA shock after 23 years. It must be said, however, that the observed phenomenon is not incidental; the deficits have persisted over a longer period as students' mathematical competence has consistently decreased since 2012. In addition to the mathematical competence drop, a growing achievement gap also challenges the German education system. This gap is particularly evident during the transition from elementary to secondary school (Stanat et al., 2022). These two unfavorable developments, that is, a general competence drop and a widening achievement gap, are alarming and put teachers in schools under double pressure to provide high-quality instruction tailored to a wide range of performance levels. Thus, there is a significant need for remedial action, for instance, supporting teachers by giving them opportunities to provide more individualized instruction and personalized learning (Brasiel et al., 2016).

In times of progressing digitalization, one possible way to confront this challenge might be to integrate computer-based math learning programs into math instruction. Unlike complex interventions, which are often costly and usually reach only few students, most math learning programs are cost-effective and have the advantage that they can be easily implemented for a



broad range of recipients. In addition, many programs allow a task assignment according to the learners' skill level and thus offering opportunities for individualized learning.

Previous research has already shown that various types of math learning programs could have the potential to positively affect students' math achievement (e.g., Brasiel et al., 2016; Byun & Joung, 2018; Higgins et al., 2019; Hillmayr et al., 2020; Ran et al., 2021; Sailer & Homner, 2020; Tokac et al., 2019). However, meta-analyses have also pointed out the high heterogeneity regarding the effectiveness of such interventions, as many empirical studies revealed null effects. Thus, even though the use of technology in education was expected to revolutionize students' learning, the overall results obtained up until now seem to be more disappointing than promising (Pellegrini et al., 2021). One problem with traditional experimental approaches to investigating the effects of math learning programs on students' performance is that researchers mostly only focus on comparing the program provision in an experimental group with a waitlist control group (e.g., Bai et al., 2012; Chang et al., 2015; Shin et al., 2012). When they do not find any effects, they consequently conclude that math learning programs might not be suitable tools to support students in fostering their math skills (Pellegrini et al., 2021). However, this approach lacks the depth needed to fully understand the underlying mechanisms that potentially make such programs effective and it does not offer testable explanations for nonexistent effects.

One possible starting point here would be to give greater consideration to the complexity of individuals' interaction with the program. Hereby, students' varying practice behavior might play a central role, as there is already some empirical evidence that positive program effects are related to students' usage intensity (Cheung & Slavin, 2013; Ozel et al., 2008). Hence, just offering a math learning program to students might not be sufficient to support them; it is important that students practice with such learning programs regularly over a longer period in order for them to achieve learning success. Even though this general assumption on successful learning has also been made in the field of cognitive psychology (Gerbier & Toppino, 2015; Toppino & Gerbier, 2014), empirical studies investigating the effects of math learning programs on students' performance have, so far, mostly neglected their actual practice behavior (e.g., see Bai et al., 2012; Brezovszky et al., 2019; Chang et al., 2015; Huang et al., 2014; Hung et al., 2014; Shin et al., 2012).

Two central assumptions emerge from this, which indicate new valuable directions that can be taken when investigating the potential of math learning programs. Firstly, the question arises as to which factors are associated with students' practice behavior, as the mere provision of a program does not automatically mean that all students practice to an equal degree.

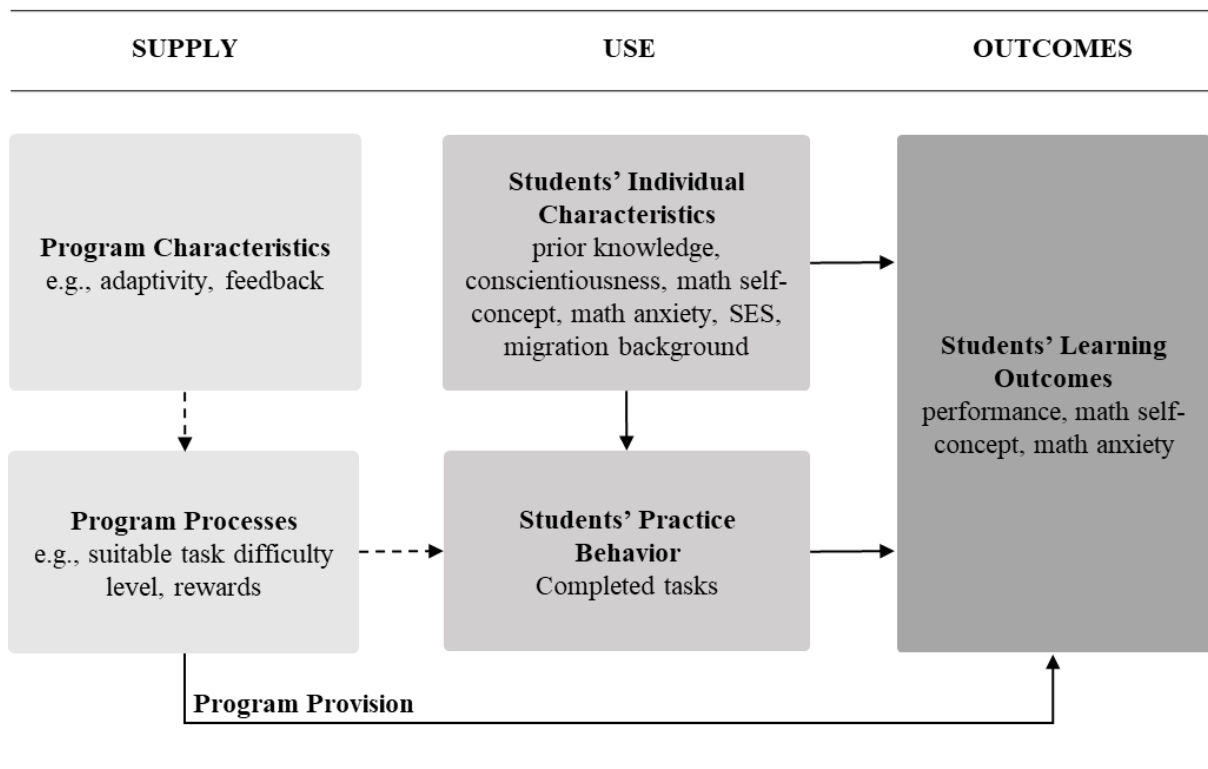
Accordingly, an investigation of whether individual student characteristics, such as cognitive, affective-motivational, or family-related preconditions, are associated with their practice behavior could provide important information. However, no research yet has investigated this association within the context of a math learning program.

Secondly, varying practice behavior might explain why some students benefit from math learning program interventions, whereas others do not. Due to their design, most studies that examine math learning programs have researchers or teachers monitor students' practice to ensure that the students practice a designated number of tasks or for a designated time period (see Brezovszky et al., 2019; Huang et al., 2014; Rodrigo, 2011; Shin et al., 2012; Van den Heuvel-Panhuizen et al., 2013). However, students' level of practice could still vary and, furthermore, this approach can only provide limited information about how students would interact with the program when they are unsupervised, not only in school but also, for instance, at home. Very few studies have provided students with a math learning program in an ecologically valid setting and considered their practice behavior in terms of process measures (i.e., the number of tasks completed or the time spent with the program). They all provided first empirical evidence that students' performance increased the more they practiced with a program (Haleva et al., 2021; Hassler Hallstedt et al., 2018; B. R. J. Jansen et al., 2013; Spitzer, 2022). At this point, it is important to think about how students can further benefit from learning programs, apart from in their performance. In accordance with a long tradition in psychological research, constructs closely related to performance such as academic motivation and emotions might be particularly relevant in the present context (e.g., Eccles et al., 1983; Eccles & Wigfield, 2020; Pekrun, 1988; Pekrun, 2006). However, only very little is known about how math learning programs, and especially students' practice behavior with them, are related to students' academic motivation and emotion (Dondio et al., 2023; Fadda et al., 2022; B. R. J. Jansen et al., 2013).

In line with these research desiderata, a supply-use model adapted to the context of math learning programs provided the heuristic framework for this dissertation (Helmke & Weinert, 1997; Seidel, 2014; see Figure 1.1). The model strongly supports (1) the assumption that there are links between students' individual characteristics (e.g., prior math knowledge, personality, affective-motivational preconditions, and family background) and their practice behavior with a learning program, and (2) the assumption that students' practice behavior is related to their performance, academic motivation, and emotions (e.g., math self-concept and math anxiety). Thus, interindividual student differences might explain diverging practice behavior, which could result in favorable or unfavorable learning outcomes.

**Figure 1.1**

*Heuristic Framework Displaying the Central Variables of the Dissertation: Students' Individual Characteristics, Their Practice Behavior With a Math Learning Program, and Potential Learning Outcomes*



*Note.* The framework is based on a supply-use model adapted to the context of math learning programs (Helmke & Weinert, 1997; Seidel, 2014; also see Section 2.1). Dotted lines represent theoretical assumptions; continuous lines refer to relationships investigated in this dissertation.

To investigate the two theoretical assumptions presented, three empirical studies and two additional replication analyses were conducted. Regarding their methodological design, two relevant approaches can be emphasized. First, all investigations followed a longitudinal pretest-posttest design, in which students practiced with the arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014; German: Prowise Learn). Hereby, the practice periods varied between 22 and 45 weeks. This made it possible to examine how specific student characteristics predicted students' practice over a substantial amount of time and, furthermore, how practice behavior contributed to students' learning outcomes. Second, in all investigations, process measures that were recorded during students' practice were used as indicators of students' practice behavior. This approach is underrepresented in current research (Azevedo,

2015), but it provides valuable and objective ways to investigate students' engagement with math learning programs.

Chapter 2 describes the important theoretical and empirical assumptions on which this dissertation is based. First, Section 2.1 presents the adapted supply-use model and its components, on which the heuristic framework of this dissertation is based. In Section 2.2, different learning program definitions are presented according to their functionality in order to categorize the Math Garden program. Next, Section 2.3 demonstrates the need to operationalize students' engagement (i.e. practice behavior) with objective process measures and suggests how math learning programs can possibly satisfy this need. Sections 2.4 and 2.5 give a more detailed overview of the determinants and effects of students' practice behavior with math learning programs, which are the focus of this dissertation. Closing Chapter 2, Section 2.6 derives the main research questions. Subsequently, three empirical studies as well as two additional replication analyses are presented in Chapters 3 to Chapter 7. Finally, all findings are juxtaposed and discussed in Chapter 8.

# 2

## THEORETICAL AND EMPIRICAL BACKGROUND

### 2 Theoretical and Empirical Background

When considering how math learning programs can support students' learning, the media debate established by Clark (1994) and Kozma (1994) already suggested that it is not merely the digital tool that has an effect on students' performance but, rather, that a large number of contextual factors, for instance, the instructional design features of the digital math learning program, are associated with the effect. Based on these considerations, a research field emerged that specifically examines the effectiveness of differently functioning tools (see Section 2.2.2 for a further description). Indeed, meta-analytic results have shown that adaptivity and feedback are two central promising characteristics of math learning programs and are important factors when considering their effect on students' performance (Hillmayr et al., 2020). The main conclusion drawn from this rather program-centered view is that tasks within math learning programs need to adapt to learners' individual skill level to effectively foster their performance and that learners benefit from receiving at least corrective feedback (Hillmayr et al., 2020).

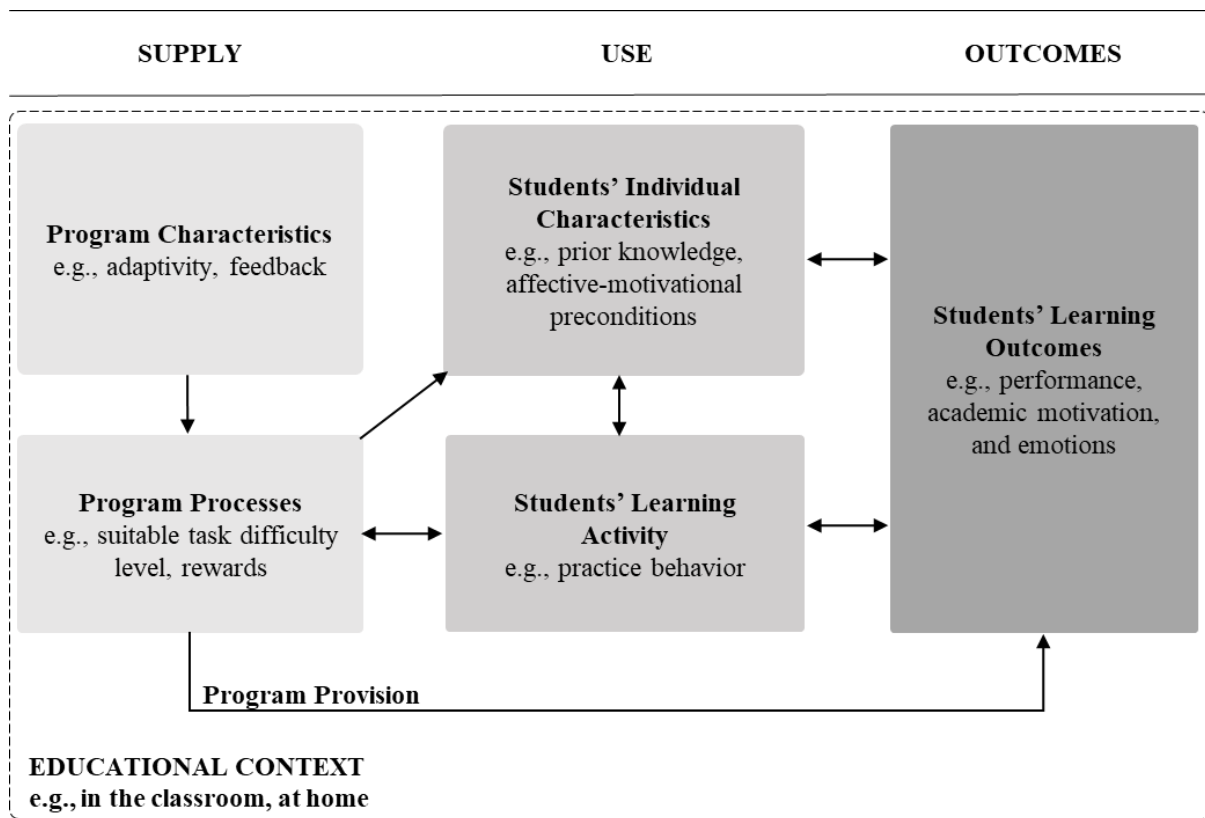
In contrast to this view, the aspect of how the complex individual interacts with the math learning program could be considered as a contextual factor and, thus, a more person-centered approach could be taken. The central assumption of this approach would be that students have different person-related preconditions, such as their prior math knowledge or their motivation regarding math. These varying preconditions might lead to different practice behavior with the program, which, in turn, would result in different effects on learning outcomes (Seidel, 2014). Whereas the program-centered perspective has gained increasing attention in recent years, this more person-centered view is rather underrepresented in current research on math learning programs. To gain new insights regarding this approach, a supply-use model, adapted to the context of math learning programs (see Figure 2.1) offers a sound theoretical basis, as it displays the suggested interplay between student-related determinants of practice behavior and the consequences of that practice behavior (Seidel, 2014).

#### 2.1 Student Characteristics, Their Practice Behavior, and Their Learning Outcomes

The supply-use model (Helmke & Weinert, 1997; Seidel, 2014) was originally developed to describe the complex conditions of students' learning in the classroom. It is divided into three connected levels: *supply*, students' *use*, and their *learning outcomes*, and thus represents the idea that the interplay of academic supply characteristics, students' individual characteristics, and their individual learning activities plays a crucial role in the effectiveness for students' cognitive and affective-motivational learning outcomes (Helmke & Weinert, 1997; Seidel, 2014; see also Brühwiler & Blatchford, 2011; Reinhold et al., 2024).

**Figure 2.1**

*Adapted Supply-Use Model of the Relationship Between the Provision of a Math Learning Program, Students' Characteristics, Their Learning Activity, and Their Learning Outcomes*



A math learning program can be perceived as one form of a supply. This supply structure includes specific characteristics, such as the beneficial characteristics of adaptivity and feedback, which are emphasized in the program-centered approach (Hillmayr et al., 2020). These characteristics then cause processes; adaptivity leads to tasks that are adapted to the learner's skill level, while feedback and its associated rewards make students feel valued for their practice. Thus, the provision of a program could indeed be assumed to form a direct association between the processes of the learning program and students' learning outcomes (see Figure 2.1). But, even though all students are provided with the same math learning program, the model suggests that they might still differ regarding their use of it (i.e., in their practice behavior). A relevant factor that affects practice behavior encompasses students' individual characteristics. This includes, for instance, their prior knowledge, personality, their affective-motivational preconditions, and their family background. Finally, due to their diverging practice behavior, the model further assumes that this also leads to differences in students' learning outcomes, primarily in their performance but also in their academic motivation and emotions.

In the following three sections, the single levels of the supply-use model in the context of math learning programs are described in more detail. First, the following section displays how different math learning programs can be categorized on the basis of two approaches. Hereby, a special focus is placed on describing the Math Garden program (Klinkenberg et al., 2011; Straatemeier, 2014), which was offered to students for the purpose of investigating the research questions of this dissertation.

### **2.2 Math Learning Programs as a Supply Structure**

A wide variety of digital math learning programs have been developed in different countries to support students in acquiring and developing mathematical competence. However, many of them differ in their aims and scope, as well as their conceptual functioning, which is why findings from one specific program cannot be generalized to all math learning programs. It is therefore important to categorize them in order to differentiate between the effects of different programs. Hereby, at least two approaches, which are not mutually exclusive, can be presented to create a common basis.

#### **2.2.1 Categorizing Math Learning Programs According to Their Aims and Scopes**

Within the first approach, math learning programs can be categorized according to their varying aims and scope. Some programs are more curriculum-oriented, whereas other programs are more skill-oriented. Curriculum-oriented programs can be used as a complement to traditional math instruction to support students in acquiring new competences. For instance, the program Bettermarks<sup>1</sup> from Germany was designed for secondary school students with the aim of providing a wide range of learning material from the regular German curriculum for Grades 4 to 11. It is structured in a similar way to a textbook; the program is divided into different topics that cover the components of the curriculum, whereby each topic contains several sections that build on each other. Within each section, teachers can choose between problem sets and assign them to their students. Similar mechanisms apply, for example, to the program CLEARMath<sup>2</sup> (Carnegie Learning) from the United States. In contrast, skill-oriented learning programs focus on supporting students in strengthening specific skills that they have already learned. For instance, the Finnish program Semideus<sup>3</sup> was developed to support students in fraction processing and numerical development, whereas the Dutch program Math Garden<sup>4</sup> aims to foster students' quick mental arithmetic skills, as well as their numerical reasoning. Thus, Math

---

<sup>1</sup> <https://de.bettermarks.com/>

<sup>2</sup> <https://www.carnegielearning.com/solutions/math/>

<sup>3</sup> <https://seriousgamessociety.org/2016/09/22/semideus-a-game-for-mastering-rational-numbers/>

<sup>4</sup> <https://www.prowise.com/de/product/prowise-learn/>



Garden is well suited for students in elementary school and at the beginning of secondary school. Different domains are implemented within the program, for instance, addition, subtraction, multiplication, and division, but also domains where students need to complete a given number series, or games in which numbers must be combined to obtain a desired target number. Students have the opportunity to choose between these different domains within their own garden, which then results in a session, and every session consists of 10 tasks that have to be solved. In line with its name, Math Garden includes special visual elements, which means that after completing one session, flowers and plants start to grow within the learner's garden, visually representing students' ability to grow (see Figure 2.2). Such emotional design elements, which aim to motivate students, are typical within skill-oriented programs. These elements mean that such programs can be defined as serious educational games or, at least, as having a game-based character (Kiili et al., 2014; Plass et al., 2015; Plass et al., 2020). This applies to both Semideus and Math Garden.

**Figure 2.2**

*Student Interface of Math Garden*



*Note.* Addieren = addition, Subtrahieren = subtraction, Multiplizieren = multiplication, Dividieren = division, Zählen = counting, Einmaleins = multiplication tables.

### 2.2.1.1 Digital Game-Based Learning

Digital game-based learning is about utilizing the qualities of common digital games for educational purposes (Prensky, 2001) by integrating a mixture of learning and game components (Beserra et al., 2014). Thus, when basic instructional tasks are redesigned, they should become more interesting, meaningful, and, consequently, more effective for learning (Plass et al., 2020). Features that turn a learning program into a more serious educational game

are, for instance, the integration of visual aesthetic elements, a narrative design, for example, including a scaffolding (human-like) pedagogical agent, and incentive structures, such as stars and trophies (Plass et al., 2015). In other words, serious educational games aim to combine entertainment with an educational intent (Bellotti et al., 2013), which is why they aim to improve not only students' performance but also their motivational and emotional states (Plass et al., 2015). A growing digital game-based research field focuses on examining how specific design elements can foster students' performance and motivation (e.g., see Kiili et al., 2014). For instance, Koskinen et al. (2022) showed that when emotional design was applied to scaffolds, it was able to improve fifth-grade students' interest and self-efficacy when practicing fractions.

In line with the name of the society that created it (i.e., Serious Games Society), the program Semideus refers to itself as a serious educational game. It follows a storyline, in which the main character Semideus wants to climb Mount Olympus to find out his origins. By successfully completing tasks, students receive rewards, which they can use to bring Semideus closer to the top of Mount Olympus. Furthermore, Semideus also uses a pedagogical agent. Math Garden, in contrast, does not fully fulfill the definition of a serious educational game, as it does not follow a specific storyline during practice, and it does not include a pedagogical agent. Instead, one goal for students can be to create a garden full of flowers and plants due to sufficient practice in the different domains, which will cause the flowers and plants to display their growing abilities (see Figure 2.2). Nonetheless, the program includes an incentive system based on coins that can be earned. Hereby, speed and accuracy during practice determine the number of coins earned, as the tasks must be solved within a time limit. After practice, the coins can be exchanged for trophies within the program, such as cups or medals, which can then be stored within the learner's own digital trophy shelf.

Due to its aim and scope, the Math Garden program can be categorized as being a skill-oriented program with a game-based character. However, this categorization neglects to include certain other program features, such as the fact that some programs like Math Garden also work adaptively, whereas other programs do not. Thus, a second categorization approach is needed, which focuses on the program's complexity.

### **2.2.2 Categorizing Math Learning Programs According to Program Complexity**

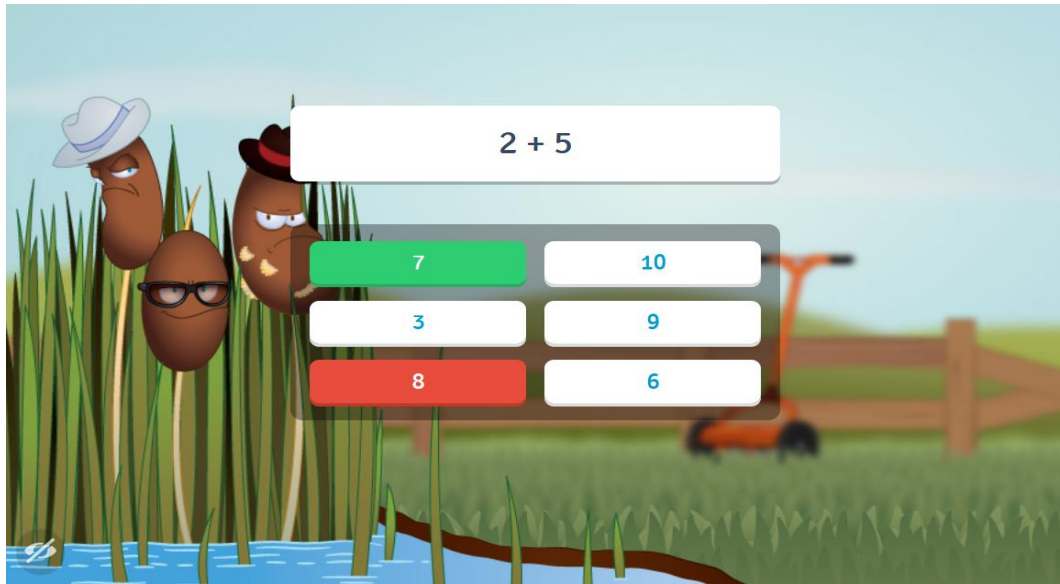
According to the classification framework proposed by Nattland and Kerres (2009), digital learning tools can be further categorized into five commonly used types with different levels of complexity. From less complex to most complex, the authors differentiate between drill and practice programs, tutoring systems, intelligent tutoring systems, hypermedia systems, and

simulations. Drill and practice programs aim to strengthen previously acquired content knowledge by allowing learners to practice at their own pace and repeat exercises as needed. In contrast, tutoring systems provide new knowledge in small units and additionally provide learners with opportunities to practice and repeat the newly acquired knowledge. Both types of programs offer immediate feedback on the correctness of the learner's solutions (Hattie & Timperley, 2007). Regarding an increasing level of complexity, intelligent tutoring systems include "adaptive features", which can be implemented on a macro or a micro level. For the macro level, this means that persistent characteristics of the learner (e.g., preferences for pictures or colors) are assessed in advance and used for (design) decisions within the program. Adaptivity on a micro level, however, refers to the learner's prior knowledge, allowing personalized adjustments to task difficulty and the pace of presenting the content. These systems also provide differentiated feedback or hints on how to optimize the learning process (Nattland & Kerres, 2009). Hypermedia systems, functioning as nonlinear hypertext systems, do not follow a structured teaching approach. Instead, they serve as encyclopedias within educational settings, offering information connected through hyperlinks. As the most complex type of digital learning tool, simulations aim to represent real-world situations that can be altered by manipulating various parameters. Simulations are valuable for applying or expanding knowledge through explorative learning, for instance, in virtual laboratories.

With regard to the digital tools described, it is plausible to assume that programs that vary in their complexity also vary in their effectiveness on student performance. Moreno and Mayer (2007) suggested that interactivity might be the central variable that explains this mechanism. That is, a program is most effective if it interactively responds, for instance, to students' ability level by providing appropriate tasks and also gives immediate feedback on students' solutions, which should help them to reflect on their answers and potentially reconsider inefficient solution strategies. Based on this idea, a research field emerged that investigates instructional design features within digital tools and their effects on students' performance. Within this field, the most prominent feature is certainly feedback. Two meta-analyses conducted by Van der Kleij et al. (2015) and Mertens et al. (2022) investigated the relationship between different methods of providing feedback in computer-based learning environments and students' performance. They differentiated between feedback that indicated whether a response was correct or incorrect (i.e. knowledge of results), feedback that additionally provided the correct answer (i.e., knowledge of correct response), and feedback that provided an explanation as to why the answer was correct or incorrect (i.e., elaborated feedback). Both studies reported positive effects across all types of feedback; however, elaborated feedback showed the largest

effect sizes, whereas feedback that provided the correct answer yielded larger effect sizes than feedback that only indicated the correctness of the answer. Due to these findings, researchers have suggested that intelligent tutoring systems—which provide elaborated feedback—might have a larger impact on students’ math performance (W. Ma et al., 2014) than drill and practice programs that mostly only provide the correct answer or indicate the correctness of an answer (Bayraktar, 2001). These assumptions were confirmed by a meta-analysis conducted by Hillmayr et al. (2020). In that meta-analysis, the authors also pointed out that an adaptive function is positively linked to students’ performance (Hillmayr et al., 2020).

In general, most curriculum-oriented math learning programs are (intelligent) tutoring systems, whereas skill-oriented programs are mostly drill and practice programs. Regarding the programs mentioned above, Bettermarks and CLEARMath both offer feedback to the learner, they work adaptively at least on a micro level, they aim to provide students with new knowledge from their curriculum, and they also offer opportunities to strengthen and repeat content already learned; they can thus be categorized as intelligent tutoring systems. With regard to Semideus and Math Garden, however, the boundaries between these categories become blurred. For instance, the Math Garden program aims to strengthen quick mental arithmetic and gives direct corrective feedback after each task (see Figure 2.3), which is why it could be categorized as a drill and practice program. However, Math Garden also works adaptively, which means that it includes features of an intelligent tutoring system. The same applies to Semideus. Thus, both programs can be described as intelligent drill and practice programs. Nonetheless, the programs vary strongly in terms of how complex adaptivity is implemented. In Semideus, adaptivity means that the order of superordinate levels can be adjusted by the practitioners if the students have problems solving the levels in the originally specified order. In contrast, Math Garden’s adaptivity is much more complex as it functions on the item level. This means that an underlying mechanism provides students with tasks that are adapted to their skill level.

**Figure 2.3***Example of Corrective Feedback After Solving a Task*

#### *2.2.2.1 Math Garden's Adaptive Algorithm*

In order to implement adaptive practice within Math Garden, the technique of computerized adaptive testing (CAT) was used (Van der Linden & Glas, 2000; Wainer et al., 2000), which is based on the item response theory (Lord & Novick, 1968). This theory involves statistical models that link responses to an item to the latent ability of a person (Lord & Novick, 1968). Hence, the idea is that CAT can dynamically determine a person's ability level and consequently adjust the item difficulty based on that person's previous responses. Although CAT was originally designed for measurement only, the goal of the Math Garden program was to combine CAT measurement approaches within a practical tool in order to foster students' arithmetic skills (Straatemeier, 2014). One challenge within CAT is that the difficulty of an individual item must be known in advance; this requires pretests, which require resources and this, in turn, means that CAT is not often used in the everyday school context. In contrast, Math Garden's algorithm uses the Elo rating system (Elo, 2008), which eliminates the need for elaborate pretests. In this algorithm, learners and items act as opponents, and solving an item is considered a match. Both learners and items have their own expected ratings, which reflect their abilities. When a student correctly solves an item, he or she wins the match, gaining points on the rating scale while the item loses points. This increases the student's expected skill level and reduces the item's difficulty level. The CAT system used by Math Garden can therefore autonomously determine item difficulty and predict the learner's skill level after the learner has solved approximately 10 to 15 items per domain, resulting in a dynamic performance score.

The algorithm aims for each student to have a 75% probability of solving tasks at their individual skill level per domain as a default setting. However, this probability can be manually adjusted (easy = 90%, difficult = 60%).

Math Garden is web-based; this comes with two additional advantages as all user responses are recorded and stored. Firstly, this improves Math Garden's adaptive algorithm, as the prediction of an item's difficulty is more precise the more users have answered that item (i.e., item-learner matches). Secondly, a huge number of learner-related process measures are generated. Math Garden therefore offers math teachers a valuable diagnostic tool: On the one hand, the program generates performance scores based on the difficulty of the tasks that the students completed correctly. On the other hand, Math Garden stores the number of tasks completed, which is an indicator of how much students have engaged with the program. Both types of measures are displayed for teachers on a digital dashboard within their own account. The students' performance scores can be compared based on various characteristics (e.g., age or grade level) with a reference group (all other students with the same characteristic). This enables the classification of both above-average and below-average performance and, if necessary, the initiation of interventions (Straatemeier, 2014). Furthermore, teachers can monitor how many tasks their students complete within each domain and see which specific tasks students prefer to practice. Thus, if, for example, a student only practices the domain of addition, the teacher can also temporarily block this domain and assign domains to the student's garden that he or she thinks the student needs to practice more (e.g., subtraction).

The process measures generated (i.e., performance scores, completed tasks) within Math Garden are not only beneficial for teachers; they also offer a valuable data source for researchers. In line with the underlying model of this dissertation, the stored number of tasks completed within a math learning program can be understood as students' learning activity. This measurement approach is unique within the field of educational psychology, as most research conducted in the last decades assessed students' learning activity with self-reports (e.g., see Hasty et al., 2021; Louw et al., 2008; Muenks et al., 2017; Skinner et al., 2009). However, self-reports are potentially biased (Donaldson & Grant-Vallone, 2002), which is why a debate on the necessary revolution in how to assess students' learning activity was held by scientists from the engagement research tradition (see Azevedo, 2015). Contributing to that debate, the following section displays students' learning activity from an engagement framework perspective and highlights how process measures stored while students practice with math learning programs can offer valuable new perspectives on how to assess their behavioral engagement.

### 2.3 Rethinking Students' Engagement and Process Measures

Students' engagement has been described as “the holy grail of learning” (Sinatra et al., 2015, p. 1) as it is the thing that makes learning effective. It has been characterized as “students' active participation in academic and curriculum- or school-related activities, and commitment to educational goals and learning” (Christenson et al., 2012, p. 816). Thus, it represents the visible expression of students' motivation (Skinner et al., 2009) and serves as a predictor of various beneficial student outcomes, especially performance and students' well-being (Wong et al., 2023). However, due to its very broad definition, there has only been little consensus among researchers regarding a uniform engagement framework (Azevedo, 2015; Wong & Liem, 2022). Nonetheless, in several studies published at the beginning of the 2000s (e.g., Appleton et al., 2006; Fredricks et al., 2004; Furlong et al., 2003), researchers agreed that students' engagement is a meta-construct consisting of at least three dimensions, namely, affective, behavioral, and cognitive engagement (Wong & Liem, 2022).

However, not all three dimensions might have the same importance. Regarding students' performance, meta-analyses revealed the strongest relationship for the behavioral engagement component, ranging between .35 and .39 (Lei et al., 2018; Wong et al., 2023). As behavioral engagement has been defined as “the extent to which students intentionally exert effort during learning activities” (Wong & Liem, 2022, p. 123), it could therefore be argued that displaying high levels of behavioral engagement is most important for students' learning success. Although the finding that the behavioral component of students' engagement is most closely related to students' performance is very valuable, the fact that the measurement of students' behavioral engagement is usually based on self-reports within the engagement research tradition might be problematic (Azevedo, 2015). Using self-reports means, for example, asking students how much time they have spent, on average, per week working on the subject of math (e.g., Hasty et al., 2021; Louw et al., 2008). The problem hereby is that self-reports are grounded in students' perception of how they behaved within certain processes and these perceptions might not align with what actually occurs during learning (Azevedo, 2015; Zimmerman, 2008). For instance, students could overestimate their learning time due to social desirability because they might know that math is important (Birtwistle et al., 2022; Donaldson & Grant-Vallone, 2002). Conversely, students might also underestimate their learning time. By doing so, they would avoid attributing possible failure to their own abilities, instead attributing it to too little engagement, which would be self-serving (Birtwistle et al., 2022; Donaldson & Grant-Vallone, 2002). Thus, such traditional self-report approaches might hinder the understanding of students' actual engagement processes. To complete the picture of the importance of behavioral

engagement by minimizing the bias, Podsakoff et al. (2012) recommended using additional data sources.

To meet this need for the unbiased measurement of the behavioral engagement component within the context of math learning programs, the integration of objective process measures could be a possible starting point. For instance, researchers could use students' practice behavior in terms of the number of tasks completed or the amount of time spent on a math learning program. As learning programs can easily assess such background measures, one would certainly expect learning program researchers to already have made extensive use of this function; this could also be expected because there should be little doubt that behavioral engagement is a highly relevant factor that needs to be considered when examining the effect mechanisms of math learning programs.

However, learning program researchers have mostly not yet included objective measures of students' practice behavior (cf. Haleva et al., 2021; Hassler Hallstedt et al., 2018; Spitzer, 2022). This is surprising given the untapped scientific potential of digital learning programs. Considering objectively measured practice behavior could significantly contribute to (1) the ongoing debate on how to improve the assessment of students' behavioral engagement, and (2) improving the quality of studies within the field of math learning programs.

Hereby, two relevant questions arise. The first one concerns the factors that are related to students' practice behavior, as the mere provision of a program does not guarantee equally high practice levels among all students. Thus, specific groups of students might practice less and, consequently, might not get the chance to benefit from a math learning program implementation, which is why they might need additional support. A few studies from the context of homework research offered promising insights regarding this assumption, for instance, by showing that students with a higher math self-concept reported an increased willingness to invest effort when doing math homework (Trautwein & Lüdtke, 2007, see Section 2.4 for further examples). But, again, effort (i.e., students' learning activity) was measured with self-reports in the study of Trautwein and Lüdtke (2007). Notably, no previous research has explored the link between students' characteristics and students' objectively measured practice behavior within the context of a math learning program.

With regard to the idea that students with specific characteristics might differ regarding their levels of practice, one could further assume that varying practice behavior in turn contributes to varying learning outcomes. Up to now, only very few studies have investigated this question in the context of performance (Haleva et al., 2021; Hassler Hallstedt et al., 2018; Spitzer, 2022) and even fewer investigations have investigated it in the context of students'



affective-motivational outcomes (B. R. J. Jansen et al., 2013). Both approaches, that is, examining the potential determinants and the possible consequences of students' practice behavior, as well as their potential contribution to the field of math learning program investigations are explored in more detail in the next two sections.

### **2.4 The Role of Students' Characteristics in Their Practice Behavior**

Coming from a general theoretical perspective that aims to derive the characteristics and student-related preconditions that might influence students' practice behavior with a math learning program, the above-mentioned adapted supply-use model clearly suggests that various cognitive, personality-related, motivational, affective, and family background-related variables are associated with students' individual learning activity (Helmke & Weinert, 1997; Seidel, 2014). Nonetheless, this model only provides a broad view and does not offer theoretical explanations of the expected direction of effects. For example, the model suggests that students' motivational preconditions are related to their practice behavior, but one can hardly argue that highly motivated students would display higher levels of practice. Thus, in order to determine why specific individual student characteristics might explain potentially diverging practice behavior with a learning program, further theoretical assumptions are needed. As discussed above, practice behavior represents one facet of behavioral engagement. Behavioral engagement is grounded within various prominent psychological theories that outline associations with different student characteristics (e.g., Eccles, 2007; Eccles & Wigfield, 2020; Eysenck et al., 2007; Sweller et al., 1998; Tett et al., 2021).

On the basis of these theories, the following sections discuss how prior knowledge, conscientiousness as a personality trait, math self-concept and math anxiety as aspects of academic motivation and emotion, and family background variables such as socioeconomic background (SES) or a migration background are related to students' behavioral engagement, that is, students' practice behavior with a math learning program.

#### **2.4.1 Prior Knowledge**

Prior knowledge has long been considered the most important factor that influences student learning and achievement due to the important role it plays when students acquire new knowledge (Tobias, 1994; Reinhold et al., 2024). The cognitive load theory states that, during the process of acquiring new knowledge, relevant information is processed in the working memory, but the working memory has limited capacity (Sweller et al., 1998). A central task during knowledge acquirement (i.e., the learning process) is therefore to keep the load on the

working memory as low as possible. Thus, certain instructional design-related, but also person-related, characteristics promote or hinder successful learning (Kirschner et al., 2011; Paas et al., 1994). For instance, students with higher prior knowledge have more working memory capacity available to process a learning task. Prior knowledge therefore has the potential to decrease students' cognitive load; this can be assumed to lead to higher task engagement during the learning process (Paas et al., 2003). Empirical evidence supports this assumption by showing that students' prior knowledge was positively related to their behavioral engagement when proceeding with a learning task (Dong et al., 2020; Mihalca et al., 2011; Myhill & Brackley, 2004). Regarding students' prior math knowledge, similar mechanisms might emerge when students practice with math learning programs.

### **2.4.2 Personality**

With regard to different personality traits and engagement, the trait activation theory considers personality traits to have the latent potential to cause individuals to engage in specific situations and behave in certain ways in those situations (Tett et al., 2021), that is, an individual expresses their traits when exposed to situational cues related to those traits. The trait activation theory was originally developed to investigate employees' job performance, but it can also be transferred to the educational context (see Meyer et al., 2023).

Within education, conscientiousness has been found to be the most important personality trait, as it comes with various characteristics that favor students' individual learning processes (Mammadov, 2022; Poropat, 2009; Vedel, 2014). Conscientious students can be described as being organized, ambitious, self-disciplined, perseverant, and achievement-oriented (Costa & McCrae, 1992; Dumfart & Neubauer, 2016); these characteristics are positively related to engagement and achievement-related behavior (Heaven et al., 2007). Thus, in terms of the trait activation theory, conscientious students regard a specific learning situation as a cue to act in the manner described above (i.e., in an organized, ambitious, etc., manner), which consequently leads to high task engagement (Tett et al., 2021). Studies conducted to investigate the relationship between conscientiousness and students' engagement were able to support the theoretically assumed relationship, for instance, by showing that conscientious students spent more time learning tasks (Poropat, 2014), had higher self-reported work engagement (Sulea et al., 2015), and invested more academic effort (Trautwein et al., 2009). Such positive effects of conscientious students could therefore also be expected in the context of practicing with math learning programs.

### 2.4.3 Academic Motivation

Motivation is a key construct in different areas of psychology and can broadly be characterized as the energy behind individuals' actions (T. Jansen et al., 2022). In the context of educational psychology, researchers mostly focus on academic motivation, which is understood as “the process whereby goal-directed academic activity is initiated and sustained” (Koenka, 2020, p. 3). Over the past four decades, diverse academic motivation constructs have been investigated, for instance, self-efficacy, self-concept, intrinsic and extrinsic motivation, interest, task value, and achievement-related goals (see T. Jansen et al., 2022). These constructs have their foundations in several well-known psychological theories such as the situated expectancy-value theory (Eccles & Wigfield, 2020). Regarding achievement activities, this theory clearly states that beliefs about one's expectancy of success as well as task values are closely linked to achievement-related choices, engagement, and persistence (Eccles & Wigfield, 2020). The theory also incorporates students' motivational characteristics, such as math self-concept, which represents the self-evaluation of one's ability in the subject of math (Marsh & Martin, 2011). Notably, there is a consensus in the research community that expectancy of success is empirically indistinguishable from domain-specific self-concepts (Eccles & Wigfield, 2002; Wigfield & Eccles, 1992), which is why, according to the situated expectancy-value theory, math self-concept can be seen as the empirical operationalization of success expectancies within the subject of math (e.g., Trautwein et al., 2012; Nagengast et al., 2011). In other words, students' math self-concept is assumed to be directly related to achievement-related behavior.

This assumed link is supported by several empirical studies. Guo et al. (2015), Nagy et al. (2007), and Meece et al. (1990) showed that students' math self-concept is related to their course enrollment intentions. In addition, Köller et al. (2000) as well as Trautwein and Lüdtke (2007) also showed that math self-concept is associated with students' learning activity in the subject of math and with their willingness to invest effort when doing homework. Thus, students' math self-concept might play a central role in students' practice with a math learning program.

### 2.4.4 Academic Emotions

There is widespread agreement that emotions are complex experiences that involve various interrelated psychological processes that include affective, cognitive, physiological, and motivational components (Kleinginna & Kleinginna, 1981; Scherer, 2000). Within the context of educational psychology, academic emotions are defined as emotions that are tied to achievement activities, such as studying or practicing (Pekrun & Stephens, 2010). Hereby, a

prominent academic emotion is math anxiety, which is characterized as an unpleasant feeling associated with numerical manipulations and math problem solving (Richardson & Suinn, 1972). Regarding its relationship with achievement activities, the attentional control theory clearly suggests that math anxious students avoid math-related tasks (i.e., they show low levels of engagement) due to unpleasant negative physiological and psychological reactions (Eysenck et al., 2007). Empirical research has supported this theoretical assumption by showing that math anxious students avoid choosing math-intensive courses (Hembree, 1990; LeFevre et al., 1992; Ramirez et al., 2018). Furthermore, Quintero et al. (2022) provided evidence that math anxious students display low behavioral engagement when it comes to math-related tasks in math instruction. In light of this, one could assume that similar mechanisms might apply when math anxious students practice with math learning programs.

### **2.4.5 Family Background**

Researchers in sociology, economics, and psychology have extensively documented the link between students' family background, such as their SES or migration background, and students' academic achievement (e.g., Hattie, 2023; Laosa, 1999; Teachman et al., 1997). They have provided consistent empirical evidence that students growing up in families with more financial, time, and intellectual resources are more likely to be successful in their educational career (Eccles, 2007).

A theoretical framework provided by Eccles (2007) explains these mechanisms; Eccles adapted the situated expectancy-value approach to the context of family-related characteristics and processes. Regarding students' learning activity, Eccles (2007) suggested that through the interplay of parents' beliefs and parents' activity-specific behavior, family characteristics such as students' SES or a migration background are related to students' achievement-related choices and engagement. This means that, for instance, a higher SES makes parents more likely to seek, identify, summarize, and evaluate information about their child's education, which might lead to specific beliefs that include assumptions about what their children need in order to succeed in school. Such beliefs result in beneficial parental behavior (e.g., talking, modeling positive interactions, providing support), which, in turn, enriches the child's environment, subsequently resulting in positive learning-related outcomes (Davis-Kean et al., 2019).

For both family background variables (i.e., SES and a migration background), there is already empirical evidence linking them to students' individual learning activity. Previous studies showed that students with higher SES reported higher levels of behavioral engagement compared to their peers with a lower SES (Fullarton, 2002; Tomaszewski et al., 2020). An

extensive investigation by Chiu et al. (2012), including 41 different countries, showed that a migration background was negatively linked to school engagement, as students with a migration background often did not have the necessary resources at home and may have received less support from their parents. Regarding students' practice behavior with a math learning program, students' SES as well as a migration background could therefore be expected to play an important role.

To summarize, the theories mentioned within this section provide comprehensive and detailed frameworks on the mechanisms that link various individual student characteristics (i.e., prior math knowledge, conscientiousness, math self-concept, math anxiety, SES, and migration background) to students' behavioral engagement, and these theoretical frameworks have been supported by empirical evidence. Still, no study has yet investigated student characteristics in the context of students engaging with supportive digital tools, for instance, a math learning program, even though this context is becoming more and more relevant. Certain student characteristics (e.g., conscientiousness) could be expected to favor students' practice behavior, whereas other characteristics, such as math anxiety, might hinder their practice. Thus, based on the supply-use model (Helmke & Weinert, 1997; Seidel, 2014; see Figure 2.1), as well as the theoretical considerations and empirical evidence discussed above, this dissertation focuses on students' prior math knowledge, conscientiousness, math self-concept, math anxiety, SES, and migration background as potential determinants of students' practice behavior within a math learning program. The results of such this approach could inform researchers and, more importantly, practitioners about which students might be at risk and, therefore, need more initial support and encouragement in terms of practicing. This leads to the question of how practice behavior contributes to students' learning outcomes, which is presented in the next section.

### **2.5 Effects of Practice Behavior on Students' Learning Outcomes**

Regarding students' learning outcomes, studies investigating the effects of math learning programs have, so far, mostly focused on performance (Hattie, 2023). That certainly makes sense; however, students might also benefit regarding their emotional and motivational outcomes, which are often closely related to their math performance (Marsh & Martin, 2011; Ashcraft & Krause, 2007). Thus, even if a program does not increase performance in the short term, it can still be effective in the long term by improving students' motivation or by reducing negative academic emotions. There is thus a need for a broader and more in-depth investigation of the effects of learning programs; this was recently stated in the meta-analysis of Dondio et al. (2023), which is why Vanbecelaere et al. (2023) introduced a comprehensive framework for

assessing the effects of learning programs. At the core of this framework, the authors distinguish between cognitive and noncognitive outcomes, thus encouraging researchers to consider a wider range of students' learning outcomes within their investigations.

Many previous studies only provided students with a math learning program and then evaluated whether such an intervention made a difference in the performance of this group compared to a control group (e.g., see Bai et al., 2012; Brezovszky et al., 2019; Chang et al., 2015; Huang et al., 2014; Hung et al., 2014; Shin et al., 2012). However, this approach does not consider students' actual engagement with the program, although the supply-use model strongly suggests that practice behavior might be one of the most decisive factors in whether or not an implementation is successful (Helmke & Weinert, 1997; Seidel, 2014). Thus, there is a pressing need to take practice behavior into account when evaluating learning program implementations.

In sum, two central research desiderata can be derived. First, in addition to investigating the effects of learning programs on students' performance, researchers should also consider the effects on students' academic motivation and emotions. Second, when investigating a broader range of outcomes, it is important not only to test the provision of the program but also to include students' actual practice behavior.

To narrow the gap within research on math learning programs, this dissertation therefore focuses on a broad approach to investigate students' learning outcomes. Previous research paradigms of the effect of math learning programs on students' learning outcomes, that is, on performance, academic motivation, and academic emotions, are presented in the following three sections and, additionally, the important role that students' practice behavior might play is outlined.

### **2.5.1 Performance**

Students' performance improvement is certainly the key learning outcome when it comes to the potential effects of math learning programs. The cognitive theory of multimedia learning, proposed by Mayer (2014), explains why math learning programs such as Math Garden can foster students' academic performance. According to the author, active engagement with learning content is crucial for comprehending new content. During such engagement, interactivity plays a key role in successful learning, which means that the program should repeatedly respond to the learners' actions (Moreno & Mayer, 2007). In line with this, programs such as Math Garden interact by adapting task difficulty according to students' performance

levels, and they also offer immediate feedback (see also Section 2.2.2). This facilitates direct reflection and may lead to reconsideration of inefficient calculation strategies.

To investigate the effects of learning programs on students' performance, at least two different research paradigms have emerged. The first is the common gold standard approach, which researchers traditionally use to evaluate their interventions. In this approach, an intervention, such as providing a math learning program in an experimental group, is tested in comparison to a control group (see Hariton & Locascio, 2018; Webber & Prouse, 2018). The second is a process measure approach from educational data mining or the learning analytics field. It focuses more on investigating students' engagement during practice (see Baker & Inventado, 2014). In the following sections, the advantages and limitations of these two approaches are discussed and the potential of combining these two approaches is pointed out.

### *2.5.1.1 The Gold Standard Approach to Evaluating Interventions*

When evaluating math learning programs, a distinction can be made between efficacy and effectiveness. Efficacy refers to the achievement of the intended goals in an environment in which experimental control is maintained at a very high level in order to improve internal validity. This ensures that the observed performance increase results directly from the treatment and that the influence of other variables is minimized. In contrast, effectiveness also refers to the attainment of the intended goals but under real-world conditions, thus focusing on high external validity (Brook & Lohr, 1991; Flay et al., 2005; Hunsley et al., 2014). Thus, although studies on the efficacy of math learning programs are theoretically feasible, for instance, if students come into a laboratory to practice, the relevance of this approach is seriously questioned as real learning usually takes place in less controlled environments, for instance, in the classroom (All et al., 2015).

To evaluate the effectiveness of a math learning program, the use of the common and popular gold standard approach is now well established (e.g., see Bai et al., 2012; Brezovszky et al., 2019; Chang et al., 2015; Huang et al., 2014; Hung et al., 2014; Shin et al., 2012). In this approach, an intervention (i.e., using a math learning program) that is carried out in an experimental group is evaluated by comparing it with a waitlist control group that does not receive the treatment, usually in a pretest-posttest design (Hariton & Locascio, 2018; Webber & Prouse, 2018). Due to the randomized assignment of participants to each group, various unknown person-related third variables are thereby expected to be held constant (Eid et al., 2017). Regarding the statistical analysis, researchers then examine the program provision, that is, whether a student received the program (experimental group) or not (waitlist control group). This approach fulfills important requirements regarding inferences about causality, such as a

potential covariation of performance improvement within the experimental group but not within the control group, a temporal order of the potential performance effect ensured through a pretest and posttest, and the preclusion of alternative explanations through randomization and the inclusion of known third variables (Eid et al., 2017).

Various meta-analyses have already tried to measure the general size of a performance effect based on the results of empirical studies that evaluated math learning program interventions using the gold standard approach (e.g., Byun & Joung, 2018; W. Ma et al., 2014; Tokac et al., 2019). However, according to Pellegrini et al. (2021), the results are disappointing and seem to indicate that math learning programs do not really fulfill their aim of supporting students' learning. Hereby, the authors argued that the effect sizes were mostly small and often turned out to be insignificant ( $0.05 \leq d \leq 0.42$ ; see Byun & Joung, 2018; Cheung & Slavin, 2013; Li & Ma, 2010; W. Ma et al., 2014; Ran et al., 2021; Tokac et al., 2019). Nonetheless, as math learning programs represent cost-saving interventions, their smaller effect sizes can still have practical relevance (Kraft, 2020). Furthermore, students from the control group generally also attend math instruction; thus, they also have access to resources similar to those available to the treatment group, which is why smaller effect sizes could also be expected (Kraft, 2020). In conclusion, the small effect sizes might be unproblematic and might even be more in line with expectations.

What might really be the problem, however, is that the authors of the above-mentioned meta-analyses pointed out a high heterogeneity regarding the effects of the empirical studies included in their meta-analyses (e.g., Byun & Joung, 2018; Tokac et al., 2019). Kraft (2020) suggested that the strength of an intervention effect varied according to who was included in the analysis, that is, when merely the provision of an intervention was tested, the effects could be expected to be smaller than when the actual receipt of an intervention was tested, as it could not be ensured that all students who were targeted by the intervention actually participated in it (Kraft, 2020). Regarding math learning programs, this would mean that, when testing simply the program provision, it cannot be ensured that all students actually practiced with the program. Thus, it is possible that, in studies that reported smaller effect sizes, only the provision of a math learning program was tested, whereas, in studies that reported bigger effect sizes, the actual receipt of a math learning program was tested. This assumption is supported by studies in which teachers or researchers supervised students' practice and thus ensured that they completed the intended number of tasks or for the intended amount of time; these studies consistently report significant increases in students' performance (Brezovszky et al., 2019; Huang et al., 2014; Rodrigo, 2011; Shin et al., 2012; Van den Heuvel-Panhuizen et al., 2013).



However, this approach can only provide limited information about how students interact with the program when they are unsupervised and is therefore only partly ecologically valid.

In summary, this heterogeneous results pattern might be an indicator that not all students make equal use of a program provision. If this were indeed the case, this would be problematic and would mean that it is not the programs themselves that are the source of the heterogeneous effects but, rather, the common methodological gold standard approach used to investigate their potential effects. Although this traditional intervention research paradigm offers several advantages, particularly with regard to the important issue of causality (Eid et al., 2017), it is limited in its ability to draw conclusions about why missing effects occur, as there is little information about the students' actual use of a math learning program (i.e., behavioral engagement; Reinhold et al., 2024), especially when they are unsupervised. Thus, a stronger engagement-related approach should be included in investigations. This might be achieved with methods from the field of educational data mining and learning analytics.

### *2.5.1.2 Process Measures Approach*

In times of digitalization, programs and applications mostly store their users' behavior in terms of process measures, for instance, the number of tasks completed in the program, the time spent on the program, or the correctness or incorrectness of the answers provided. This also applies to math learning programs (e.g., Math Garden, Semideus, and Bettermarks). The utilization of such large-scale process measures is a key element of the educational data mining and learning analytics fields (see Baker & Inventado, 2014). Within these fields, by examining response accuracy and speed, researchers aim to understand how students develop cognitively or how their in-game behavior (e.g., skipping tasks) potentially affects their in-program performance (Buwalda et al., 2016; Brinkhuis et al., 2018; Haleva et al., 2021; Mooij et al., 2021; Savi et al., 2018; Savi et al., 2023; Van der Maas & Nyamsuren, 2017).

Although this approach has the huge advantage of having a big database and also offers nuanced and urgently needed insights into how students engage with math learning programs, it is limited when it comes to drawing objective conclusions about students' potential performance improvements. This is because, although programs such as Math Garden provide performance scores, these performance scores do not undergo a traditional validation process with a representative norm sample, which is why they might not be entirely trustworthy (Moosbrugger & Kelava, 2020). However, if behavior-related process measures were linked with valid performance scores (measured with a pretest and a posttest, as in the traditional gold standard approach), this would narrow the research gap on how students' behavioral engagement with a math learning program contributes to potential performance improvements.

Unfortunately, the number of studies so far that have taken process measures into account and linked them to objective performance scores is still modest. The few studies that used such data mostly considered the time spent practicing with math learning programs (e.g., Hassler Hallstedt et al., 2018) or the number of tasks completed by the students in the program (e.g., B. R. J. Jansen et al., 2013). Promisingly, all studies were able to show that students' practice behavior was significantly related to their performance. This underlines the importance of taking practice behavior into account when drawing conclusions about the effectiveness of math learning programs on students' performance, as missing effects might occur because students systematically practiced too little.

### **2.5.2 Academic Motivation**

There is much empirical evidence that academic motivation is strongly related to students' performance (e.g., Howard et al., 2021; Kriegbaum et al., 2018). Thus, the observed decline of this motivation over the course of secondary school students' educational career is observed with concern (Gaspard et al., 2017; Jacobs et al., 2002; Scherrer & Preckel, 2019). Consequently, efforts to promote academic motivation and create a learning environment that maintains and increases students' academic motivation are considered to be highly important (T. Jansen et al., 2022).

The self-determination theory (Deci & Ryan, 1993, 2000; Deci et al., 1991) can be used to explain why math learning programs might be beneficial for students' motivation. According to this theory, learning environments have to meet students' innate psychological needs for competence, autonomy, and social relatedness. Competence is described as the state in which an individual knows how to effectively achieve certain outcomes, while autonomy involves taking control of initiating and regulating one's own actions (Deci et al., 1991). The third need, social relatedness, pertains to the presence of satisfying and secure relationships (Deci et al., 1991). According to the self-determination theory, students' need for competence might be satisfied by engaging with a learning program, as it provides learning support through direct program feedback and offers the possibility to increase the time spent on learning. Students' need for autonomy can be supported by a learning program if the program provides students with optimal, challenging tasks appropriate for their competence level, because being able to successfully solve a learning task allows the student to strive for autonomy (Krapp, 2005). Finally, learning programs might have the potential to increase students' social relatedness. This point might be counterintuitive at first glance, as one would assume that students primarily work on their own while practicing. However, if, after practicing, students talk about the

experiences they had while practicing, this might be a basis for a common interest, which can satisfy social relatedness (Deci et al., 1991).

Various meta-analyses have already focused on the associations between the use of educational technologies and students' motivation (e.g., Higgins et al., 2019; T. Jansen et al., 2022; Higgins et al., 2019; Fadda et al., 2022). In their systematic review of meta-analyses, T. Jansen et al. (2022) found that the use of technology (i.e., digital tools and online learning opportunities) was positively associated with students' academic motivation. However, their result refers to a broad range of technology use across different subjects. For math, a meta-analysis conducted by Higgins et al. (2019) supported the results pattern that the use of technology in the classroom can foster students' academic motivation but, also in their study, the term "technology" covers a range of different tools and systems (e.g., digital workbooks, digital encyclopedia, video podcasts, computer games). More specifically, within the context of math learning programs, a first and recent meta-analysis was conducted by Fadda et al. (2022). The authors suggested an overall positive effect of math learning programs on students' motivation ( $d = 0.27$ ; Fadda et al., 2022). Thus, all previous meta-analyses on this topic support the assumption that math learning programs might be promising tools to foster students' academic motivation.

Regarding the meta-analysis conducted by Fadda et al. (2022), within the 20 studies included, most studies referred to the motivational constructs of self-efficacy and self-confidence, while only three studies assessed students' motivation in terms of math self-concept, and those studies yielded mixed results. Whereas Main and O'Rourke (2011) found an increase in the math self-concept of 9- to 10-year-old students, Miller and Robertson (2010, 2011) did not find any effects for 10- to 11-year-old students. Still, even though all these studies ensured that students, either supervised by a research assistant or by their teacher, practiced for an intended amount of time each week, none of these studies took students' actual practice behavior into account, which might explain the unclear pattern of findings. So far, only B. R. J. Jansen et al. (2013) took students' practice behavior into account when investigating the effects of Math Garden on students' perceived math competence in a Dutch sample. Perceived math competence as a construct is considered to be similar to math self-concept (Lee, 2009). Interestingly, compared to a control group without a learning program, they found a small effect of the program provision on perceived math competence in one out of three experimental groups. In this group, the task difficulty was adjusted during practice to a 75% solution probability (medium) according to students' individual skills. No effect was found in either the easy group (90% solution probability) or the difficult group (60% solution probability).

However, the effect was not mediated by the number of tasks completed, which might indicate that students' motivational improvement could be independent of how much they actually interact with the program. Still, on average, only 50 students were included in each experimental group, which is why the approach might lack statistical power. Furthermore, the authors also pointed out that the effect was very small, which could have been due to the rather short intervention period (i.e., 6 weeks; B. R. J. Jansen et al., 2013). Thus, intervention studies that are conducted over a substantial period of time (i.e., longer than 6 weeks) and that apply the process measures approach (i.e., consider students' practice behavior) are needed in order to shed light on how math learning programs have the potential to increase students' math self-concept.

### **2.5.3 Academic Emotions**

Within the context of education, academic emotions are tied not only to achievement activities, as stated within Section 2.4, but also to achievement outcomes in terms of success or failure (Pekrun & Stephens, 2010). Most studies that investigate the relationship between academic emotions and achievement outcomes refer to students' anxiety, especially within the subject of math (e.g., see Aldrup et al., 2020; Gunderson et al., 2018; Pekrun et al., 2017; Ramirez et al., 2018). Hereby, researchers aim to understand the unfavorable, negative relationship between students' math anxiety and their math performance, which was consistently found in three prominent meta-analyses (Barroso et al., 2021; Hembree, 1990; X. Ma, 1999). As around one third of students apparently suffer from math anxiety (OECD, 2013; Sorvo et al., 2017), there is an urgent need to break this vicious cycle and understand how different learning environments could have the potential to reduce students' math anxiety.

The control-value theory, proposed by Pekrun (2006), suggests that math anxiety arises when the perceived value of a math-related task is high, but the ability to solve the task is considered uncontrollable due to lacking resources such as skills and knowledge. However, if a math learning program offers the opportunity to adjust task difficulty based on the learner's skill level, this should make tasks more controllable for students with math anxiety, thereby reducing their anxiety (Pekrun, 2006). Positive feedback from the program might further enhance students' perception of task controllability. Additionally, learning programs provide a private space for students to make mistakes. Given that public experiences with math contribute to math anxiety (Bekdemir, 2010), students may feel more secure about making errors during practice with a math learning program than in oral sessions with the entire class. Consequently, by minimizing public (i.e., potentially embarrassing) experiences, such programs might enable

students to gain a heightened sense of control, which could have the potential to reduce their math anxiety when they engage with math learning programs.

In light of these promising theoretical assumptions about how students with math anxiety might benefit from practicing with math learning programs, Dondio et al. (2023) recently published a first meta-analysis of the relationship between digital game-based learning and math anxiety. However, the results are rather sobering, as the authors found only weak and nonrobust evidence that digital games in math have the potential to reduce math anxiety (Dondio et al., 2023). Still, they suggested that the effects would be stronger for longer intervention periods. Indeed, many studies that investigated this relationship referred to a one-day/one-session intervention (e.g., Hung et al., 2014; Vanbecelaere et al., 2020), but this brief experience, even though it might have been positive, can presumably not eliminate the often deep-rooted feelings of math anxiety (Dowker et al., 2016). In other words, according to Pekrun's control-value theory, students must actively perceive that learning environments such as math learning programs offer higher levels of control in order for them to reduce their anxiety (Pekrun, 2006). Therefore, their actual engagement over a substantial period of time could be a decisive factor in this perception coming to fruition.

So far, only B. R. J. Jansen et al. (2013) took students' practice behavior, in terms of their completed tasks, into account when investigating the effects of the Math Garden learning program on students' math anxiety over the course of 6 weeks. However, they found no effect of the program provision or of the number of completed tasks. This nonsignificant result aligns with the mixed and unclear findings within the research field concerning the effects of digital math learning programs on students' math anxiety. Therefore, it is necessary to design intervention studies that are conducted over an even longer period of time (i.e., longer than 6 weeks) and to use the process measures approach (i.e., consider students' practice behavior) to tackle the question of how math learning programs can contribute to decreasing students' math anxiety.

### **2.6 Research Questions and Studies Conducted**

So far, research has mainly concentrated on investigating the effect of providing math learning programs on students' performance or on the differential effects of beneficial features such as adaptivity or feedback, that is, a very program-centered perspective has been used (e.g., Hillmayr et al., 2020). Regarding the provision, the findings are mixed, and there is a growing number of voices claiming that learning programs do not have the expected revolutionary effect and may be more disappointing than promising (see Pellegrini et al., 2021).

The supply-use model (Helmke & Weinert, 1997; Seidel, 2014), adapted to the context of math learning programs, however, clearly states that the individual interaction with the program might play a crucial role when it comes to potential program effects, not only on performance but also on academic motivation and emotions. This can be described as a more person-centered perspective. Hence, there is a pressing need to understand (1) how the individual preconditions of students are related to their practice behavior, and, in turn, (2) whether practice behavior contributes to improving students' performance and academic motivation and to reducing negative academic emotions (i.e., anxiety). The present dissertation aimed to address these theoretical assumptions; the determinants and consequences of students' practice behavior in the arithmetic learning program Math Garden were investigated. In order to answer current calls to integrate objective process measures into educational research (Azevedo, 2015), students' practice behavior was also included, mainly operationalized by the number of tasks completed.

Up until now, the small number of studies that have taken students' practice behavior in terms of process measures into account focused on the effects on performance. They showed that the quantity of practice was consistently related to students' performance increases (Haleva et al., 2021; Hassler Hallstedt et al., 2018; B. R. J. Jansen et al., 2013; Spitzer, 2022). However, as individual student characteristics might promote or even hinder students' practice, practitioners need to know which students potentially need more initial support when a math learning program is being implemented. So far, though, no research has been conducted on such links in the context of a math learning program. Due to their embedding within the theoretical model as well as their evident relevance within educational research (Barroso et al., 2021; Chiu et al., 2012; Dong et al., 2020; Kriegbaum et al., 2018; Poropat, 2009; Tomaszewski et al., 2020), this dissertation therefore focuses on students' prior knowledge, relevant personality traits (conscientiousness), students' academic motivation (math self-concept) and emotions (math anxiety), and their family background (migration background and SES) as individual preconditions that might be related to practice behavior.

Furthermore, students might not only benefit from a learning program regarding their performance; they might also benefit in terms of their academic motivation and emotions due to the beneficial features implemented within many learning programs, such as adaptivity or feedback (Deci et al., 1991; Pekrun, 2006). Still, high-quality studies are limited within this area (Dondio et al., 2023; Fadda et al., 2022), especially when it comes to the effects of students' practice behavior on academic motivation and emotions (B. R. J. Jansen et al., 2013). Thus, this dissertation also focuses on the effects on students' math self-concept as well as their math

anxiety, as both constructs are strongly associated with students' future performance (Marsh & Martin, 2011; Ramirez et al., 2018). Even if a math learning program does not increase performance in the short term, it can be effective in the long term by improving math self-concept or reducing math anxiety.

Taken together, the main goal of this dissertation was to investigate the important role practice behavior might play when providing students with a math learning program. To attain this goal, this dissertation addressed two research questions:

**Research Question 1 (RQ1).** Are individual student characteristics associated with students' practice behavior in a math learning program?

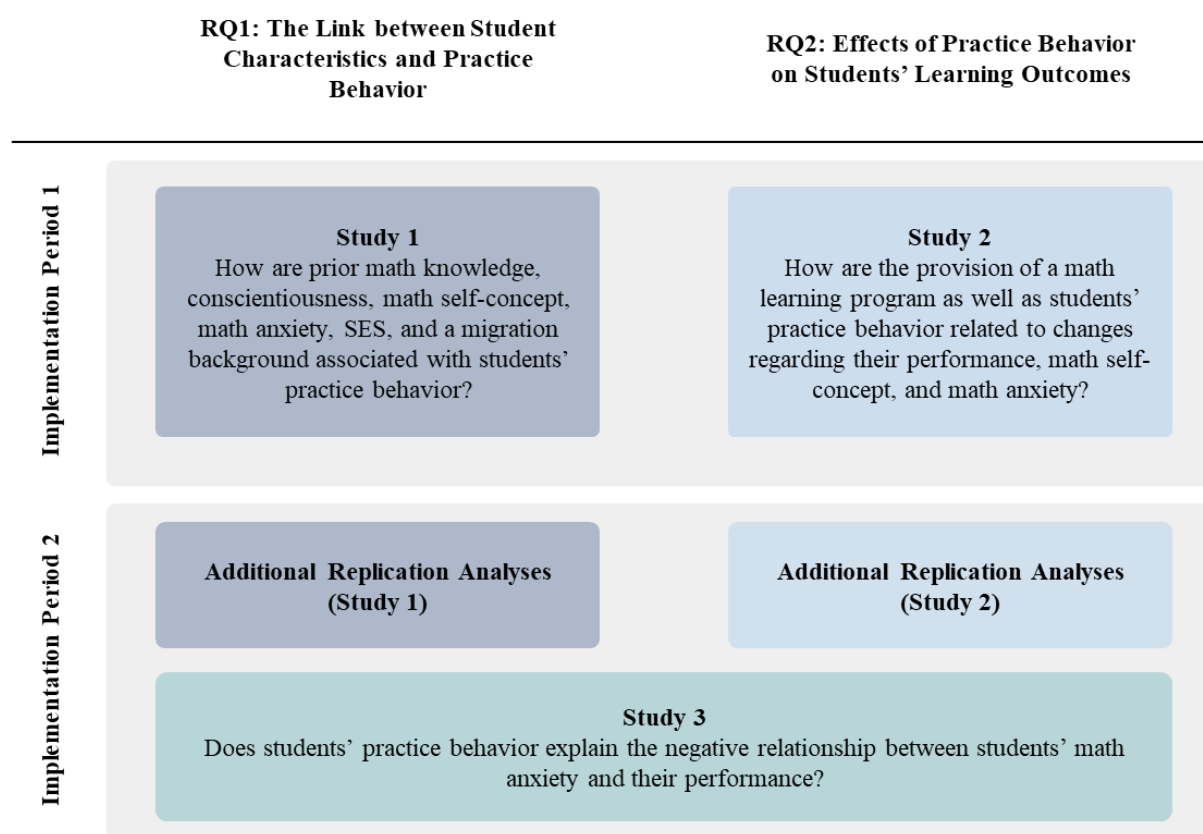
**Research Question 2 (RQ2).** How are the provision of a math learning program and students' practice behavior with it related to their performance, math self-concept, and math anxiety?

In order to address the two research questions, three empirical studies and two additional replication analyses were conducted (see Figure 2.4). For this purpose, the Math Garden program (Klinkenberg et al., 2011; Straatemeier, 2014; German: Prowise Learn) was implemented twice within large German samples over substantially long periods, in both cases as part of a randomized pretest-posttest control group design. The first implementation included a practice period of 22 weeks in the 2019/2020 school year; in the second implementation period, students practiced for 45 weeks with Math Garden in the 2021/2022 school year. During both periods, students' practice behavior was traced. Thus, as students were not closely monitored by researchers when they practiced, but data regarding their behavior were still available, all of the studies conducted demonstrated a high level of ecological validity. Study 1 and Study 2 referred to the first implementation period, whereas Study 3 is based on the data collected within the second implementation period. Additionally, the assumptions of Study 1 and Study 2 were replicated and reported as additional analyses to investigate the robustness of the results. For the replication analyses, the data from the second implementation period were used.

**Figure 2.4**

*Overview of the Three Empirical Studies and Two Additional Replication Analyses of the Dissertation with Regard to the Research Questions and the Implementation Periods*

---



### 2.6.1 Study 1 and Additional Replication Analyses: Students' Individual Characteristics as Predictors of Practice Behavior

To address the first research question of this dissertation, Study 1 investigated how individual student characteristics were related to students' practice behavior in Math Garden over a period of 22 weeks (Implementation Period 1). Based on the results of previous research within different learning contexts (e.g., homework behavior or course choices), it was hypothesized that students with more prior knowledge, higher levels of conscientiousness, a higher math self-concept, and a higher SES would practice more with Math Garden, whereas students with higher levels of math anxiety or students with a migration background would practice less with the program (Chiu et al., 2012; Dong et al., 2020; Dumfart & Neubauer, 2016; LeFevre et al., 1992; Trautwein & Lüdtke, 2007). Students' individual characteristics were assessed in a pretest, and their practice behavior was operationalized as the number of tasks completed,



measured on a weekly basis. Multilevel analyses were calculated to test the derived hypotheses, as the completed tasks, which were measured weekly, were clustered within students.

Using the same theoretical assumptions and methodological approaches, Study 1 was replicated in the second implementation period to ensure the robustness of the findings.

### **2.6.2 Study 2 and Additional Replication Analyses: Effects of Providing and Practice With Math Garden on Students' Learning Outcomes**

To address the second research question of this dissertation, Study 2 focused on the effects of Math Garden on students' learning outcomes in terms of their performance, math self-concept, and math anxiety over the period of 22 weeks (Implementation Period 1). Using a randomized pretest-posttest control group design, not only the effects of the program provision were investigated; how practice behavior within the experimental group was related to students' learning outcomes was also examined. The aim thereby was to extend previous research by following a more in-depth approach (Ran et al., 2021). With regard to students' performance, this means that conclusions were drawn separately about students' addition and subtraction performance, in contrast to conclusions based on a composite score. Furthermore, practice behavior was operationalized not only via students' completed tasks but also in terms of the number of weeks over which students practiced with Math Garden. According to previous research and theoretical assumptions about the beneficial features of adaptivity and feedback (Dondio et al., 2023; Fadda et al., 2022; B. R. J. Jansen et al., 2013; Marsh & Martin, 2011; Mayer, 2014; Pekrun, 2006), it was hypothesized that the provision of Math Garden, as well as the number of tasks completed and the number of weeks spent practicing with Math Garden, would improve students' addition and subtraction performance as well as their math self-concept and would decrease students' math anxiety. Regression analyses were used to statistically test the hypotheses.

Like Study 1, Study 2 was replicated in the second implementation period to ensure the robustness of the results using the same theoretical assumptions and methodological approaches.

### **2.6.3 Study 3: The Mediating Role of Practice Behavior Regarding the Relationship Between Math Anxiety and Math Performance**

Study 3 connected the assumptions and relevant findings of Study 1 and Study 2 and thus addressed both research questions of this dissertation. Based on emotion-theoretical suggestions that math anxious students in particular avoid math-related tasks (Eysenck & Calvo, 1992; Eysenck et al., 2007), this study aimed to answer the research question of whether math anxious

students avoid practicing with Math Garden and, in turn, whether this inhibits them in improving their math performance. As practice behavior was again objectively operationalized with process measures in terms of the number of tasks completed, this study extended previous research that assessed math anxious students' avoidance behavior with potentially biased self-report measures (Hasty et al., 2021; Quintero et al., 2022). Study 3 hypothesized that students' math anxiety would be associated with less practice in Math Garden but also that higher levels of practice behavior would be linked to higher student performance. Thus, lower levels of practice behavior might explain the negative relationship between students' math anxiety and their math performance. The study included three measurement points within the second implementation period: at the beginning, in the middle, and at the end of a school year ( $T_1$ – $T_2$ : 28 weeks of practicing,  $T_1$ – $T_3$ : 45 weeks of practicing), which made it possible to examine whether math anxiety at the beginning of the school year predicted students' practice behavior in the medium and long term, potentially precluding achievement gains due to continued avoidance of practice. Mediation analyses were conducted to test the proposed hypotheses.

### 2.7 References

- Aldrup, K., Klusmann, U., & Lüdtke, O. (2020). Reciprocal associations between students' mathematics anxiety and achievement: Can teacher sensitivity make a difference? *Journal of Educational Psychology, 112*(4), 735–750.  
<https://doi.org/10.1037/edu0000398>
- All, A., Nuñez Castellar, E. P., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers & Education, 88*, 29–37. <https://doi.org/10.1016/j.compedu.2015.04.012>
- Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427–445.  
<https://doi.org/10.1016/j.jsp.2006.04.002>
- Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248. <https://doi.org/10.3758/BF03194059>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle

- school students. *British Journal of Educational Technology*, 43(6), 993–1003.  
<https://doi.org/10.1111/j.1467-8535.2011.01269.x>
- Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, 17(5), 775–785.  
<https://doi.org/10.1111/desc.12155>
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics* (pp. 61–75). Springer New York.  
[https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)
- Barbieri, C. A., Young, L. K., Newton, K. J., & Booth, J. L. (2021). Predicting middle school profiles of algebra performance using fraction knowledge. *Child Development*, 92(5), 1984–2005. <https://doi.org/10.1111/cdev.13568>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Bayraktar, S. (2001). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education*, 34(2), 173–188.  
<https://doi.org/10.1080/15391523.2001.10782344>
- Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. *Educational Studies in Mathematics*, 75(3), 311–328. <https://doi.org/10.1007/s10649-010-9260-7>
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of Serious Games: An Overview. *Advances in Human-Computer Interaction*, 2013, 1–11.  
<https://doi.org/10.1155/2013/136864>
- Beserra, V., Nussbaum, M., Zeni, R., Rodriguez, W., & Wurman, G. (2014). Practising arithmetic using educational video games with an interpersonal computer. *Journal of Educational Technology & Society*, 17(3), 343–358.
- Birtwistle, E., Schoedel, R., Bemann, F., Wirth, A., Sürig, C., Stachl, C., Bühner, M., & Niklas, F. (2022). Mobile sensing in psychological and educational research: examples from two application fields. *International Journal of Testing*, 22(3-4), 264–288.  
<https://doi.org/10.1080/15305058.2022.2036160>
- Brasiel, S., Jeong, S., Ames, C., Lawanto, K., Yuan, M., & Martin, T. (2016). Effects of educational technology on mathematics achievement for K-12 students in Utah. *Journal of Online Learning Research*, 2(3), 205–226.

- Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education*, 128, 63–74.  
<https://doi.org/10.1016/j.compedu.2018.09.011>
- Brinkhuis, M. J. S., Savi, A. O., Hofman, A. D., Coomans, F., Van der Maas, H. L. J., & Maris, G. (2018). Learning as it happens: A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, 5(2), 29–46.  
<https://doi.org/10.18608/jla.2018.52.3>
- Brook, R. H., & Lohr, K. N. (1991). *Efficacy, effectiveness, variations, and quality*. RAND Corporation.
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Buwalda, T., Borst, J., Van der Maas, H., & Taatgen, N. (2016). Explaining mistakes in single digit multiplication: A cognitive model. In D. Reitter, & F. E. Ritter (Eds.), *Proceedings of ICCM 2016 - 14th International Conference on Cognitive Modeling* (pp. 131–136). The Pennsylvania State University.
- Bynner, J., & Parsons, S. (2006). *Does numeracy matter more?* National Research and Development Centre for Adult Literacy and Numeracy.
- Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics*, 118(3-4), 113–126.  
<https://doi.org/10.1111/ssm.12271>
- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International*, 52(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.  
<https://doi.org/10.1016/j.edurev.2013.01.001>
- Chiu, M. M., Pong, S., Mori, I., & Chow, B. W.-Y. (2012). Immigrant students' emotional and cognitive engagement at school: A multilevel analysis of students in 41 countries. *Journal of Youth and Adolescence*, 41(11), 1409–1425.  
<https://doi.org/10.1007/s10964-012-9763-x>

- Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of Research on Student Engagement*. Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. <https://doi.org/10.1007/BF02299088>
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1), 5–13. <https://doi.org/10.1037/1040-3590.4.1.5>
- Davis-Kean P. E., Tang S., Waters N. E. (2019). Parent education attainment and parenting. In Bornstein M. H. (Ed.), *Handbook of parenting: Vol. 2. Biology and ecology of parenting* (3rd ed., pp. 400–420). Routledge.
- Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik [The self-determination theory of motivation and its significance for pedagogy]. *Zeitschrift für Pädagogik*, 39, 223–238. <https://doi.org/10.25656/01:11173>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, 26(3-4), 325–346. <https://doi.org/10.1080/00461520.1991.9653137>
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2), 245–260. <https://doi.org/10.1023/A:1019637632584>
- Dondio, P., Gusev, V., & Rocha, M. (2023). Do games reduce maths anxiety? A meta-analysis. *Computers & Education*, 194, 104650. <https://doi.org/10.1016/j.compedu.2022.104650>
- Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, 11, 591203. <https://doi.org/10.3389/fpsyg.2020.591203>
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years?. *Frontiers in Psychology*, 7, 508. <https://doi.org/10.3389/fpsyg.2016.00508>
- Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences*, 37(1), 8–15. <https://doi.org/10.1027/1614-0001/a000182>

- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Eccles, J. S. (2007). Families, schools, and developing achievement-related motivations and engagement. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 665–691). The Guilford Press.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motivation* (pp. 75–146). Freeman.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <http://dx.doi.org/10.1146/annurev.psych.53.100901.135153>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Elo, A. E. (2008). *The rating of chessplayers, past and present*. Ishi Press International.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden [Statistics and research methods]*. Beltz.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6), 409–434. <https://doi.org/10.1080/02699939208409696>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2022). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning*, 38(1), 304–325. <https://doi.org/10.1111/jcal.12618>
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science: The Official Journal of the Society for Prevention Research*, 6(3), 151–175. <https://doi.org/10.1007/s11121-005-5553-y>

- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fullarton, S. (2002). Student Engagement with school: Individual and school level influences. *Longitudinal Surveys of Australian Youth Research Report No. 27*. Australian Council for Educational Research.
- Furlong, M. J., Whipple, A. D., St. Jean, G., Simental, J., Soliz, A., & Punthuna, S. (2003). Multiple contexts of school engagement: Moving toward a unifying framework for educational research and practice. *The California School Psychologist*, 8(1), 99–113. <https://doi.org/10.1007/BF03340899>
- Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, 48, 67–84. <https://doi.org/10.1016/j.cedpsych.2016.09.003>
- Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, 4(3), 49–59. <https://doi.org/10.1016/j.tine.2015.01.001>
- Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development*, 19(1), 21–46. <https://doi.org/10.1080/15248372.2017.1421538>
- Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, 51(8), 1163–1176. <https://doi.org/10.1037/a0039440>
- Haleva, L., HersHKovitz, A., & Tabach, M. (2021). Students' activity in an online learning environment for mathematics: The role of thinking levels. *Journal of Educational Computing Research*, 59(4), 686–712. <https://doi.org/10.1177/0735633120972057>
- Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: Randomised controlled trials. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13), 1716. <https://doi.org/10.1111/1471-0528.15199>

- Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology, 110*(8), 1127–1148. <https://doi.org/10.1037/edu0000264>
- Hasty, L. M., Malanchini, M [Margherita], Shakeshaft, N., Schofield, K., Malanchini, M [Maddalena], & Wang, Z. (2021). When anxiety becomes my propeller: Mental toughness moderates the relation between academic anxiety and academic avoidance. *The British Journal of Educational Psychology, 91*(1), 368–390. <https://doi.org/10.1111/bjep.12366>
- Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Taylor & Francis.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heaven, P. C., Ciarrochi, J., & Vialle, W. (2007). Conscientiousness and eysenckian psychoticism as predictors of school grades: A one-year longitudinal study. *Personality and Individual Differences, 42*(3), 535–546. <https://doi.org/10.1016/j.paid.2006.07.028>
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen [Conditional factors of academic performance]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie, Bd. 3: Psychologie des Unterrichts und der Schule* (pp. 71–176). Hogrefe.
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33. <https://doi.org/10.2307/749455>
- Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research, 57*(2), 283–319. <https://doi.org/10.1177/0735633117748416>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153*, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 16*(6), 1300–1323. <https://doi.org/10.1177/1745691620966789>



- Huang, Y.-M., Huang, S.-H., & Wu, T.-T. (2014). Embedding diagnostic mechanisms in a digital game for learning mathematics. *Educational Technology Research and Development*, 62(2), 187–207. <https://doi.org/10.1007/s11423-013-9315-4>
- Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1(2-3), 151–166. <https://doi.org/10.1007/s40692-014-0008-8>
- Hunsley, J., Elliott, K., & Therrien, Z. (2014). The efficacy and effectiveness of psychological treatments for mood, anxiety, and related disorders. *Canadian Psychology*, 55(3), 161–176. <https://doi.org/10.1037/a0036933>
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527. <https://doi.org/10.1111/1467-8624.00421>
- Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, 24, 190–197. <https://doi.org/10.1016/j.lindif.2012.12.014>
- Jansen, T., Meyer, J., Wigfield, A., & Möller, J. (2022). Which student and instructional variables are most strongly related to academic motivation in K-12 education? A systematic review of meta-analyses. *Psychological Bulletin*, 148(1-2), 1–26. <https://doi.org/10.1037/bul0000354>
- Kiili, K., Lainema, T., de Freitas, S., & Arnab, S. (2014). Flow framework for analyzing the quality of educational games. *Entertainment Computing*, 5(4), 367–377. <https://doi.org/10.1016/j.entcom.2014.08.002>
- Kirschner, F., Kester, L., & Corbalan, G. (2011). Cognitive load theory and multimedia learning, task characteristics and learning engagement: The current state of the art. *Computers in Human Behavior*, 27(1), 1–4. <https://doi.org/10.1016/j.chb.2010.05.003>
- Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4), 345–379. <https://doi.org/10.1007/BF00992553>
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and

- difficulty estimation. *Computers & Education*, 57(2), 1813–1824.  
<https://doi.org/10.1016/j.compedu.2011.02.003>
- Koenka, A. C. (2020). Academic motivation theories revisited: An interactive dialog between motivation scholars on recent contributions, underexplored issues, and future directions. *Contemporary Educational Psychology*, 61, 101831.  
<https://doi.org/10.1016/j.cedpsych.2019.101831>
- Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse [Course choices of girls and boys in mathematics: The role of subject-specific self-concept and interest]. *Zeitschrift Für Pädagogische Psychologie*, 14(1), 26–37. <https://doi.org/10.1024//1010-0652.14.1.26>
- Koskinen, A., McMullen, J., Ninaus, M., & Kiili, K. (2022). Does the emotional design of scaffolds enhance learning and motivational outcomes in game-based learning? *Journal of Computer Assisted Learning*, 39(1), 77–93.  
<https://doi.org/10.1111/jcal.12728>
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19.  
<https://doi.org/10.1007/BF02299087>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction*, 15(5), 381–395.  
<https://doi.org/10.1016/j.learninstruc.2005.07.007>
- Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, 25, 120–148. <https://doi.org/10.1016/j.edurev.2018.10.001>
- Laosa, L. M. (1999). Intercultural transitions in human development and education. *Journal of Applied Developmental Psychology*, 20(3), 355 – 406. [https://doi.org/10.1016/S0193-3973\(99\)00023-4](https://doi.org/10.1016/S0193-3973(99)00023-4)
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>

- LeFevre, J.-A., Kulak, A. G., & Heymans, S. L. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science*, 24(3), 276–289. <https://doi.org/10.1037/h0078742>
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(3), 517–528. <https://doi.org/10.2224/sbp.7054>
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, 22(3), 215–243. <https://doi.org/10.1007/s10648-010-9125-8>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores; with contributions by Allan Birnbaum*. Addison-Wesley.
- Louw, J., Muller, J., & Tredoux, C. (2008). Time-on-task, technology and mathematics achievement. *Evaluation and Program Planning*, 31(1), 41–50. <https://doi.org/10.1016/j.evalprogplan.2007.11.001>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520. <https://doi.org/10.2307/749772>
- Main, S., & O'Rourke, J. (2011). 'New directions for traditional lessons': Can handheld game consoles enhance mental mathematics skills? *Australian Journal of Teacher Education*, 36(2). <https://doi.org/10.14221/ajte.2011v36n2.4>
- Mammadov, S. (2022). Big five personality traits and academic performance: A meta-analysis. *Journal of Personality*, 90(2), 222–255. <https://doi.org/10.1111/jopy.12663>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics.

- Journal of Educational Psychology*, 82(1), 60–70. <https://doi.org/10.1037/0022-0663.82.1.60>
- Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology*, 114(8), 1743–1772. <https://doi.org/10.1037/edu0000764>
- Meyer, J., Schmidt, F. T., Fleckenstein, J., & Köller, O. (2023). A closer look at the domain-specific associations of openness with language achievement: Evidence on the role of intrinsic value from two large-scale longitudinal studies. *British Journal of Educational Psychology*, 93(1), 113–133. <https://doi.org/10.1111/bjep.12543>
- Mihalca, L., Salden, R. J., Corbalan, G., Paas, F., & Miclea, M. (2011). Effectiveness of cognitive-load based adaptive instruction in genetics education. *Computers in Human Behavior*, 27(1), 82–88. <https://doi.org/10.1016/j.chb.2010.05.027>
- Miller, D. J., & Robertson, D. P. (2010). Using a games console in the primary classroom: Effects of ‘Brain Training’ programme on computation and self-esteem. *British Journal of Educational Technology*, 41(2), 242–255. <https://doi.org/10.1111/j.1467-8535.2008.00918.x>
- Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology*, 42(5), 850–864. <https://doi.org/10.1111/j.1467-8535.2010.01114.x>
- Mooij, S. M. M. de, Raijmakers, M. E. J., Dumontheil, I., Kirkham, N. Z., & Van der Maas, H. L. J. (2021). Error detection through mouse movement in an online adaptive learning environment. *Journal of Computer Assisted Learning*, 37(1), 242–252. <https://doi.org/10.1111/jcal.12483>
- Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9047-2>
- Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion [Test theory and questionnaire construction]*. Springer.
- Muenks, K., Wigfield, A., Yang, J. S., & O’Neal, C. R. (2017). How true is grit? Assessing its relations to high school and college students’ personality characteristics, self-regulation, engagement, and achievement. *Journal of Educational Psychology*, 109(5), 599–620. <https://doi.org/10.1037/edu0000153>
- Myhill, D., & Brackley, M. (2004). Making connections: Teachers’ use of childrens’ prior knowledge in whole class discourse. *British Journal of Educational Studies*, 52(3), 263–275. <https://doi.org/10.1111/j.1467-8527.2004.00267.x>

- Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “×” out of expectancy-value theory? A psychological mystery, a substantive methodological synergy, and a cross-national generalization. *Psychological Science*, 22(8), 1058–1066. <http://dx.doi.org/10.1177/0956797611415540>
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2007). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12(4), 323–345. <https://doi.org/10.1080/13803610600765687>
- Nattland, A., & Kerres, M. (2009). Computerbasierte Methoden im Unterricht [Computer-based methods in the classroom]. In K.-H. Arnold, U. Sandfuchs, & J. Wiechmann (Eds.), *UTB Schulpädagogik, Pädagogik: Vol. 8423. Handbuch Unterricht* (2., aktualisierte Aufl., pp. 317–324). Klinkhardt.
- OECD. (2013). *PISA 2012 results: Ready to learn: Students' engagement, drive and self-beliefs* (Vol. III). OECD Publishing.
- OECD. (2019). *PISA 2018 results: What students know and can do* (Vol. I). OECD Publishing.
- OECD. (2023). *PISA 2022 results: The state of learning and equity in education* (Vol. I). OECD Publishing.
- Ozel, S., Yetkiner, Z. E., & Capraro, R. M. (2008). Technology in K-12 mathematics classrooms. *School Science and Mathematics*, 108(2), 80–85. <https://doi.org/10.1111/J.1949-8594.2008.TB17807.X>
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4. [https://doi.org/10.1207/S15326985EP3801\\_1](https://doi.org/10.1207/S15326985EP3801_1)
- Paas, F., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79(1), 419–430. <https://doi.org/10.2466/pms.1994.79.1.419>
- Pekrun, R. (1988). Anxiety and motivation in achievement settings: Towards a systems-theoretical approach. *International Journal of Educational Research*, 12, 307–323. [https://doi.org/10.1016/0883-0355\(88\)90008-0](https://doi.org/10.1016/0883-0355(88)90008-0)
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>

- Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, 88(5), 1653–1670. <https://doi.org/10.1111/cdev.12704>
- Pekrun, R., & Stephens, E. J. (2010). Achievement emotions: A control-value approach. *Social and Personality Psychology Compass*, 4(4), 238–255. <https://doi.org/10.1111/j.1751-9004.2010.00259.x>
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7, 233285842098621. <https://doi.org/10.1177/2332858420986211>
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. <https://doi.org/10.1080/00461520.2015.1122533>
- Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 3–24). The MIT Press.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322–338. <https://doi.org/10.1037/a0014996>
- Poropat, A. E. (2014). A meta-analysis of adult-rated child personality and academic performance in primary education. *The British Journal of Educational Psychology*, 84(2), 239–252. <https://doi.org/10.1111/bjep.12019>
- Prensky, M. (2001). *Digital natives, digital immigrants*. NCB University Press.
- Quintero, M., Hasty, L., Li, T., Song, S., & Wang, Z. (2022). A multidimensional examination of math anxiety and engagement on math achievement. *The British Journal of Educational Psychology*, 92(3), 955–973. <https://doi.org/10.1111/bjep.12482>
- Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, 53(3), 145–164. <https://doi.org/10.1080/00461520.2018.1447384>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12

- classrooms. *Journal of Educational Computing Research*, 59(1), 119–153.  
<https://doi.org/10.1177/0735633120952063>
- Reinhold, F., Leuders, T., Loibl, K., Nückles, M., Beege, M., & Boelmann, J. M. (2024). Learning mechanisms explaining learning with digital tools in educational settings: A cognitive process framework. *Educational Psychology Review*, 36(1), 14.  
<https://doi.org/10.1007/s10648-024-09845-6>
- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19(6), 551–554.  
<https://doi.org/10.1037/h0033456>
- Rodrigo, M. M. T. (2011). Dynamics of student cognitive-affective transitions during a mathematics game. *Simulation & Gaming*, 42(1), 85–99.  
<https://doi.org/10.1177/1046878110361513>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Savi, A. O., Deonovic, B., Bolsinova, M., Van der Maas, H., & Maris, G. (2018). *Tracing Systematic Errors to Personalize Recommendations in Single Digit Multiplication and Beyond*. <https://doi.org/10.31234/osf.io/53muj>
- Savi, A. O., Van Klaveren, C., & Cornelisz, I. (2023). Combating effort avoidance in computer adaptive practicing: Does a problem-skipping restriction promote learning? *Computers & Education*, 206, 104908.  
<https://doi.org/10.1016/j.compedu.2023.104908>
- Scherer, K. R. (2000). Emotions as episodes of subsystems synchronization driven by nonlinear appraisal processes. In I. Granic & M. D. Lewis (Eds.), *Emotion, Development, and Self-organization: Dynamic Systems Approaches to Emotional Development* (pp. 70–99). Cambridge University Press.
- Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research*, 89(2), 211–258. <https://doi.org/10.3102/0034654318819127>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma [Supply-use models in educational psychology. Integration of structural and process paradigms]. *Zeitschrift für Pädagogik*, 60(6), 850–866. <https://doi.org/10.25656/01:14686>

- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540–560. <https://doi.org/10.1111/j.1467-8535.2011.01197.x>
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50(1), 1–13. <https://doi.org/10.1080/00461520.2014.1002924>
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection. *Educational and Psychological Measurement*, 69(3), 493–525. <https://doi.org/10.1177/0013164408323233>
- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *The British Journal of Educational Psychology*, 87(3), 309–327. <https://doi.org/10.1111/bjep.12151>
- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich [IQB Education Trend 2021: Competence in the subjects german and mathematics at the end of year 4 in the third federal state comparison]*. Waxmann Verlag GmbH. <https://doi.org/10.31244/9783830996064>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Sulea, C., Van Beek, I., Sarbescu, P., Virga, D., & Schaufeli, W. B. (2015). Engagement, boredom, and burnout among students: Basic need satisfaction matters more than personality traits. *Learning and Individual Differences*, 42, 132–138. <https://doi.org/10.1016/j.lindif.2015.08.018>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Teachman, J. D., Paasch, K., & Carver, K. (1997). Social capital and the generation of human capital. *Social Forces*, 75(4), 1343–1359. <https://doi.org/10.1093/sf/75.4.1343>



- Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature and applications to five lines of personality dynamics research. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 199–233. <https://doi.org/10.1146/annurev-orgpsych-012420-062228>
- Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64(1), 37–54. <https://doi.org/10.3102/00346543064001037>
- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Tomaszewski, W., Xiang, N., & Western, M. (2020). Student engagement as a mediator of the effects of socio-economic status on academic performance among secondary school students in Australia. *British Educational Research Journal*, 46(3), 610–630. <https://doi.org/10.1002/berj.3599>
- Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *Psychology of Learning and Motivation*, 60, 113–189. <https://doi.org/10.1016/B978-0-12-800090-8.00004-4>
- Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology*, 99(2), 432–444. <https://doi.org/10.1037/0022-0663.99.2.432>
- Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, 97(6), 1115–1128. <https://doi.org/10.1037/a0017048>
- Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, 104(3), 763–777. <http://dx.doi.org/10.1037/a0027470>
- Van Aalderen-Smeets, S. I., Van der Walma Molen, J. H., & Xenidou-Dervou, I. (2019). Implicit STEM ability beliefs predict secondary school students' STEM self-efficacy beliefs and their intention to opt for a STEM field career. *Journal of Research in Science Teaching*, 56(4), 465–485. <https://doi.org/10.1002/tea.21506>
- Van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game.

- Educational Studies in Mathematics*, 84(3), 281–307. <https://doi.org/10.1007/s10649-013-9483-5>
- Van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized adaptive testing: Theory and practice*. Kluwer Academic Publ.
- Van der Maas, H. L. J., & Nyamsuren, E. (2017). Cognitive analysis of educational games: The number game. *Topics in Cognitive Science*, 9(2), 395–412. <https://doi.org/10.1111/tops.12231>
- Vanbecelaere, S., Demedts, F., Reynvoet, B., & Depaepe, F. (2023). Toward a framework for analyzing adaptive digital games' research effectiveness. *International Journal of Serious Games*, 10(4), 77–91. <https://doi.org/10.17083/ijsg.v10i4.618>
- Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes. *Computers & Education*, 143, 103680. <https://doi.org/10.1016/j.compedu.2019.103680>
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71, 66–76. <https://doi.org/10.1016/j.paid.2014.07.011>
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized Adaptive Testing*. Routledge. <https://doi.org/10.4324/9781410605931>
- Webber, S., & Prouse, C. (2018). The new gold standard: The rise of randomized control trials and experimental development. *Economic Geography*, 94(2), 166–187. <https://doi.org/10.1080/00130095.2017.1392235>
- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12(3), 265–310. [http://dx.doi.org/10.1016/02732297\(92\)90011-P](http://dx.doi.org/10.1016/02732297(92)90011-P)
- Wong, Z. Y., & Liem, G. A. D. (2022). Student engagement: Current state of the construct, conceptual refinement, and future research directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09628-3>
- Wong, Z. Y., Liem, G. A. D., Chan, M., & Datu, J. A. D. (2023). Student engagement and its association with academic achievement and subjective well-being: A systematic

review and meta-analysis. *Journal of Educational Psychology*. Advance online publication. <https://doi.org/10.1037/edu0000833>

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183.  
<https://doi.org/10.3102/0002831207312909>



# 3

## STUDY 1: THE RELATIONSHIP BETWEEN INDIVIDUAL CHARACTERISTICS AND PRACTICE BEHAVIOR WITHIN AN ADAPTIVE ARITHMETIC LEARNING PROGRAM

Hilz, A., Guill, K., Roloff, J., Aldrup, K., & Köller, O. (2023). The relationship between individual characteristics and practice behaviour within an adaptive arithmetic learning program. *Journal of Computer Assisted Learning*, 39(3), 970–983.  
<https://doi.org/10.1111/jcal.12780>

#### **Abstract**

**Background:** Although research on mathematics learning programs has taken off in recent years, little is known about how different person characteristics are related to practice behaviour with such programs. When implementing a mathematics learning program in the classroom, it might be important to know whether students with specific characteristics need more encouragement to make use of this learning opportunity.

**Objectives:** Using a supply-use model, we analysed the predictive power of students' cognitive characteristics (prior mathematics knowledge), personality traits (conscientiousness), motivational-affective characteristics (mathematics self-concept and mathematics anxiety), and family background characteristics (socioeconomic status and migration background) for practising with an adaptive arithmetic learning program.

**Methods:** We used longitudinal data from 203 fifth graders from seven non-academic-track schools in Germany. Practice behaviour, measured by completed tasks in the learning program, was recorded weekly for every student over a period of 22 weeks.

**Results and Conclusions:** The results of our multilevel analyses showed that mathematics anxious students practised less with the program. We did not find any relationship with the other characteristics.

**Takeaways:** Our results suggest that mathematics anxious students need more encouragement when practising with a mathematics learning program; otherwise, they do not get the chance to benefit from the use.

**Keywords:** mathematics, practice behaviour, learning behaviour, arithmetic learning programs, computer-based practising

#### 3.1 Introduction

Mathematical literacy is central to social and professional participation in modern societies (OECD, 2019). In Germany, however, about 25% of students already have fundamental deficits in basic arithmetic skills at the end of primary school; these basic skills are essential for the development of mathematical literacy in secondary school (Reinhold et al., 2019). A specific challenge for the education system, and specifically for secondary school teachers, is to provide adequate learning support for these low-achieving students in very heterogeneous classes (OECD, 2019). Given the pressure to ensure high-quality mathematics instruction, computer-based adaptive mathematics learning programs may make it possible to overcome this challenge. They easily allow teachers to provide students with supplements to regular classroom instruction that are not intended to replace instruction (Brasiel et al., 2016), and they allow students to practise at their individual skill level (Ma et al., 2014). Indeed, a large body of literature shows that various types of mathematics learning programs have the potential to positively affect achievement outcomes (Brasiel et al., 2016; Byun & Joung, 2018; Higgins et al., 2019; Ran et al., 2021; Tokac et al., 2019). However, these positive effects are mediated by how well such learning programs are used by students and may vary depending on the usage intensity (Cheung & Slavin, 2013; Ozel et al., 2008). Hence, just offering an adaptive learning program to students is not sufficient to support them; it is important that students practise with such learning opportunities regularly over a longer period of time in order for them to achieve learning success (Barzagar Nazari & Ebersbach, 2019; Vlach & Sandhofer, 2012). Therefore, the question arises as to which factors are associated with students' practice behaviour in a mathematics learning program.

Previous research suggested an association between students' characteristics, such as ability self-concept or conscientiousness, and their mathematics learning behaviour, for example, in completing homework (e.g., Ashcraft & Krause, 2007; Köller et al., 1999; Trautwein & Lüdtke, 2007). Similarly, Proulx et al. (2017) pointed out that the effectiveness of digital game-based systems in educational contexts is, for example, related to various person-oriented factors. But such person-oriented characteristics that either promote or hinder practice with a mathematics learning program have not yet been investigated. However, before implementing such a program in the classroom, it might be especially important to know whether students with specific characteristics need more encouragement in order to benefit equally from the use. Therefore, we conducted an empirical study and analysed the predictive

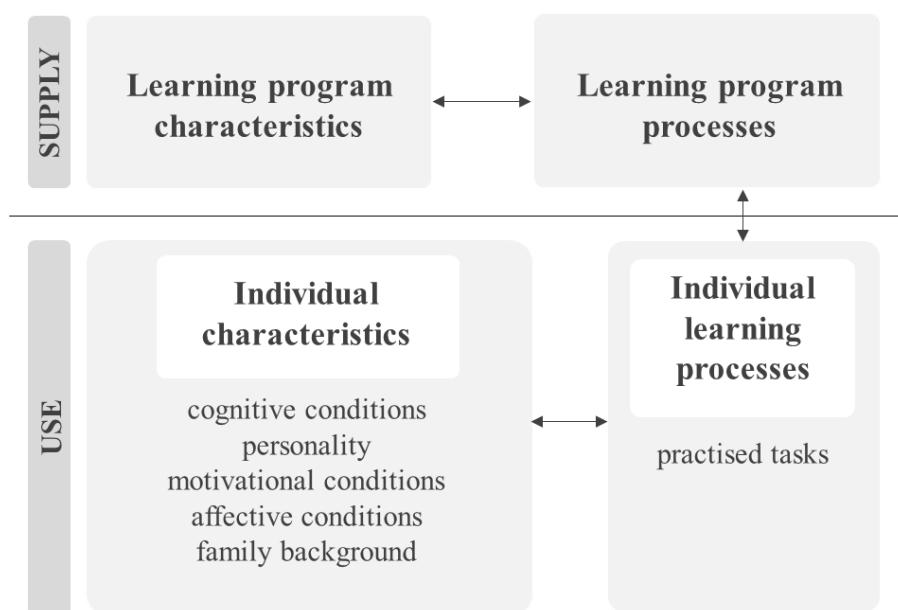
power of individual characteristics for students' practice behaviour when working with an adaptive arithmetic learning program over a period of 22 weeks.

#### 3.1.1 Individual Characteristics as Predictors of Practice Behaviour

Various researchers in the field of education have already suggested that learning behaviour is affected by different student characteristics (Imhof & Spaeth-Hilbert, 2013; Lens & Decruyenaere, 1991; Meece et al., 1990). The supply-use model, which was developed to describe the complex conditions of students' learning in classrooms, systematically integrated this idea (Helmke & Weinert, 1997; Seidel, 2014; also see Brühwiler & Blatchford, 2011). The model suggests that instruction and the design of learning environments represent a structure of *supply*. Accordingly, we conceived of arithmetic learning programs as a form of supply, which includes specific characteristics and processes. In this context, the characteristics of adaptivity and feedback are central. These characteristics then cause processes; for example, adaptivity leads to tasks that are suited to the learner's skill level, whereas feedback and related rewards make students feel that they are being valued for their practice. Importantly, even though all students receive the same supply, they may still differ in their *use* of the learning program, for example, in the extent to which they practise. One important factor that affects learning behaviour is students' characteristics, including personality; cognitive, motivational, and affective conditions; and family background. These ideas are illustrated in the adapted supply-use model presented in Figure 3.1.

**Figure 3.1**

*Adapted Supply-Use Model*





We focused on students' prior mathematics knowledge as a cognitive condition, on conscientiousness as a personality trait, on mathematics self-concept and mathematics anxiety as motivational and affective conditions, and on socioeconomic status and migration background as indicators of family background to examine the extent to which these variables predict the practice behaviour of students working on an arithmetic learning program. These variables are not only part of the theoretical model but have also always had particular relevance in educational research (Dochy et al., 1999; Nagy et al., 2007; OECD, 2013; 2019; Poropat, 2014; Seipp, 1991). Nevertheless, they have never been investigated in the context of the learning processes that occur when students use an arithmetic learning program.

#### *3.1.1.1 Prior Knowledge*

Prior knowledge is defined as a hierarchical and multidimensional entity that is dynamic and consists of different types of knowledge and skills, and it has long been considered the most important factor to influence student learning and achievement (Dochy, 1992; Tobias, 1994). In the field of research on prior knowledge and individual learning processes, Higashi et al. (2017) showed that students with higher prior computer science knowledge showed a higher willingness to participate in online computer science courses. Moreover, students with higher prior knowledge invested higher effort to complete such courses. In addition, high performance in school has been shown to significantly influence the decision to study a science, technology, engineering, or mathematics subject (Aeschlimann et al., 2015).

#### *3.1.1.2 Conscientiousness*

Conscientiousness is one dimension of the five-factor model of personality, which is used as a standard model to describe specific personality traits in contemporary research (Costa & McCrae, 1985; 1992; 2008). Conscientious people have various characteristics that favour individual learning processes, for example, striving for achievement and being self-disciplined (Costa & McCrae, 1992). Moreover, previous research has already suggested that conscientiousness is related to self-control (MacCann et al., 2009), which means that conscientiousness may reflect a better ability to self-regulate, resulting in more time spent on tasks and, consequently, more successful learning (Poropat, 2014).

#### *3.1.1.3 Academic Self-Concept*

Academic self-concept refers to the self-evaluation of one's ability in a specific domain (Marsh & Martin, 2011), and it has been suggested that it is an important factor concerning achievement-related decisions (Eccles & Wigfield, 2020). Regarding the subject of mathematics, a study by Guo et al. (2015) showed that mathematics self-concept significantly predicted mathematics coursework selection in high school, which means that students with a

high mathematics self-concept are more likely to get involved in mathematics learning processes. Furthermore, a positive mathematics self-concept was shown to be related to increased mathematics learning activity (Köller et al., 1999) and an increase in the willingness to invest effort when doing mathematics homework (Trautwein & Lüdtke, 2007).

#### *3.1.1.4 Mathematics Anxiety*

Richardson and Suinn (1972) defined mathematics anxiety as a “feeling of tension and anxiety that interferes with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations” (p. 551). Achievement emotions, such as anxiety, directly influence achievement outcomes and achievement activities (Pekrun, 2006). Orbach et al. (2019) argued that mathematics anxious students expect low success in mathematics learning situations because of missing resources to handle these situations; this is why mathematics anxious students may avoid becoming involved in learning processes in mathematics. Research from Ashcraft and Krause (2007) and Hembree (1990) supported this assumption by showing that people with mathematics anxiety avoided courses with mathematical content in school as well as professions related to mathematics.

#### *3.1.1.5 Socioeconomic Status (SES)*

Differences in SES indicate differences in access to financial, human, social, and cultural capital resources that consequently affect an individual’s access to learning opportunities (Chiu & Chow, 2015). Families with a high SES usually place high value on education in order to maintain the family’s status (Stocké, 2007) and they provide additional opportunities and support to learn at home (Hattie, 2009), which makes it more likely for students with a high SES to successfully get involved in learning processes. In the field of research on digital learning, Vekiri (2010) showed that students with a more privileged social background had better access to family sources of technological assistance and guidance. Similarly, previous research has found evidence of a digital divide, which refers to social inequality in terms of the competent use of digital media (Van Dijk, 2005). In this context, people with a low SES, in particular, are more likely to be smartphone owners and to not additionally own a desktop computer or laptop (Napoli & Obar, 2014). This is problematic because smartphones are used more often for entertainment-related or social-interactive purposes and less often for information- and learning-oriented purposes (Napoli & Obar, 2014).

#### *3.1.1.6 Migration Background*

In Germany, the number of students with a migration background is growing (OECD, 2019). These students often perceive migration as an opportunity for social advancement (Kao & Tienda, 1995) but there is a lack of family resources to translate ambitions into academic

achievement (Stanat et al., 2010). According to Portes and Rumbaut (2001), the main reason for this is a language barrier, which can make it difficult for students with a migration background to get involved in learning processes. For example, when practising with an arithmetic learning program, these students may have difficulties understanding how the login works or how to deal with the different tasks in the program. Additionally, students with a migrant background reported that their parents showed less interest in school activities (Dumont et al., 2012) and also that they perceived their parents to be less supportive with homework (Dumont et al., 2014). This indicates that not only do these students have to overcome the language barrier but also that their parents are less able to support them at home.

#### **3.1.2 The Present Study**

In summary, students may differ in their use of an adaptive arithmetic learning program, which is why the question of how individual characteristics are related to students' practice behaviour is important, for example, some students may have unfavourable individual preconditions, such as having low prior mathematics knowledge, being mathematics anxious, or having a migration background. Our study aimed to shed some light on which students need more encouragement in order to benefit equally from the provision of such a learning program. Previous research has not yet addressed this aspect; we thus aimed to fill this research gap. We conducted a 22-week study, in which we combined questionnaire data with weekly measured practice behaviour (log files) in the learning program. We aimed to answer the following research question: How do individual characteristics predict practice behaviour (number of tasks completed per week) with an adaptive arithmetic learning program?

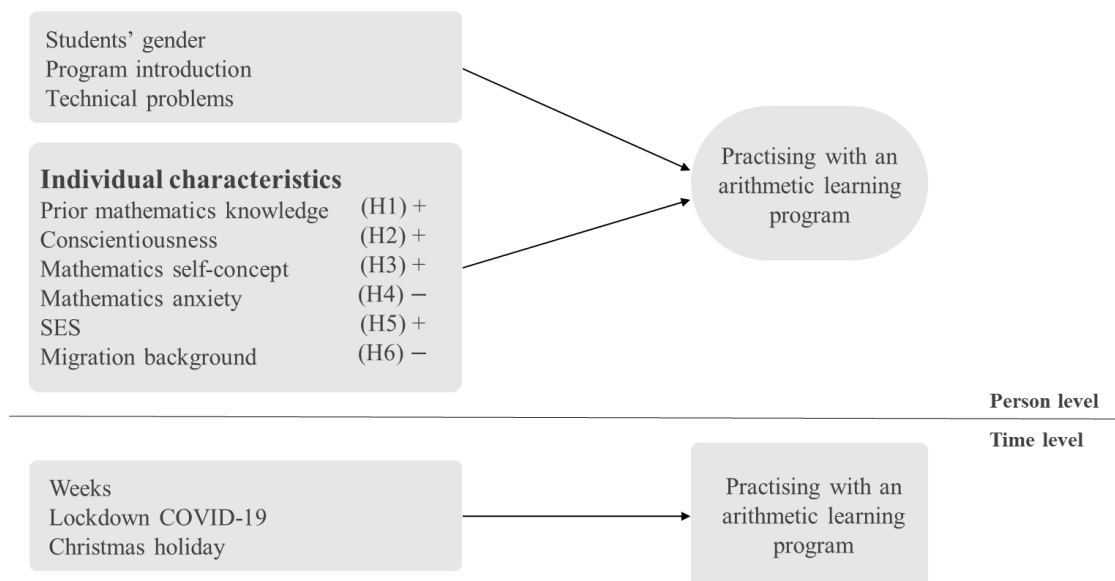
On the basis of an adapted supply-use model (Helmke & Weinert, 1997; Seidel, 2014) and previous research findings (Ashcraft & Krause, 2007; Dumont et al., 2012; 2014; Higashi et al., 2017; Köller et al., 1999; Orbach et al., 2019; Poropat, 2014; Vekiri, 2010), we expected that (H1) prior mathematics knowledge, (H2) conscientiousness, (H3) a high mathematics self-concept, and (H5) a high SES would positively predict students' practice behaviour (number of completed tasks), whereas (H4) mathematics anxiety and (H6) a migration background would negatively predict practice behaviour. The hypotheses are graphically represented in Figure 3.2.

As individual learning processes not only depend on individual characteristics but are also part of a complex system (Helmke & Weinert, 1997; Seidel, 2014), it is important to take covariates into account. To obtain an impression of students' practice behaviour over time, we considered a possible decline in practice behaviour, as research on the novelty effect suggests that students usually show high engagement when confronted with a new digital learning tool

(Jeno et al., 2019), but that this effect tends to disappear over time when students become more familiar with the tool (Clark & Sugrue, 1990). Moreover, we took two time-related events, namely, school holidays and the school closures as a consequence of the COVID-19 pandemic, into account because both were found to have effects on learning behaviour in previous research (Alexander et al., 2012; Cooper et al., 1996; Entwisle et al., 2001; Mohan et al., 2020). On the student level, we controlled our analyses for students' gender because former studies have shown significant relations between our predictors, outcome variables, and gender (Hill et al., 2016; Jansen et al., 2013; McCrae, 2002). Moreover, we controlled for technical problems when using the program (e.g., problems logging in) because such problems prevented practice and considerable time had to be invested to solve the problems. We also controlled for a program introduction because three classes did not have their personalised login data during the implementation of the program introduction. Therefore, an interactive introduction was not possible.

**Figure 3.2**

*Constructs and Hypotheses in the Current Study*



## 3.2 Methods

### 3.2.1 Design and Procedure

This study was part of a larger project on the predictors and outcomes of an arithmetic learning program. Student data were collected in schools from October 2019 to March 2020 in compliance with all data protection requirements and without any ethical concerns. Participation in the study was voluntary and parental permission was obtained before participation in the study.

The participating classes completed mathematics achievement tests and answered questionnaires assessing the student characteristics of interest for our study; a testing session lasted approximately 1.5 hours and was done with tablets (model: Samsung Galaxy Tab 4 10.1). All students were tested over a three-week period. At the end of the testing session, the classes were randomly assigned to two conditions; the condition with access to the arithmetic learning program defined the sample of the current study. The students in this condition received a 10-minute introduction to the program and accessed the program on tablets using personalised login data provided by their teachers. While practising, they got a brief oral introduction to the main principles of the program. At the end of the introductory practice session, students were told that they were allowed to, and should, practise with the program at home and in class on their own devices. Students could access the program for 22 weeks and we tracked their weekly practice behaviour. Three classes did not have personal login data due to technical problems. These students were introduced to the program in a 5-minute presentation. The teachers were asked to introduce the students to the program in more detail later. In March 2020, after 22 weeks of practice, the classes were tested again and answered questionnaires.

#### **3.2.2 Sample**

The data for this study were obtained from a sample of 203 fifth graders from 11 classes at seven urban comprehensive (non-academic-track) schools in Northern Germany. While academic-track schools prepare students primarily for higher education, non-academic-track schools offer different levels of school degrees (Maaz et al., 2008). The data from one child were excluded due to an implausibly high level of practice (more than 2,000 tasks within one week), which might have been caused by siblings also practising with the program. The final sample consisted of 202 students (49.5% male, 52.5% with a migration background, and 25.3% with parents qualified for university admission; see Table 3.1 for further sample characteristics).

#### **3.2.3 Arithmetic Learning Program**

The arithmetic learning program “Math Garden” (German version: Prowise Learn) was developed by Klinkenberg et al. (2011). Math Garden is characterised by its adaptive format, which makes it possible for students to practise at their individual arithmetic skill level. An algorithm regulates the assignment of the tasks so that the user has a pre-set average percentage of correct tasks of 75% (Klinkenberg et al., 2011). The program is structured like a garden. Wooden signs mark different tasks that target each basic arithmetic operation separately. One session consists of 10 tasks. Correct answers given within the time limit are rewarded with

virtual coins, and these coins can be used to buy virtual trophies. Teachers can check students' practice behaviour in the backend of the program.

#### **3.2.4 Measures**

##### *3.2.4.1 Practice Behaviour*

Math Garden automatically tracks the number of completed tasks per domain per week (log files). As students can quit sessions before answering all 10 tasks, we used the weekly number of completed tasks as the dependent variable for our analyses to obtain the most precise measure of practice behaviour. For each student, we estimated the total weekly practice frequency scores by adding up all of the weekly completed tasks from the following domains: addition, subtraction, multiplication, division, mix (mixture of basic arithmetic operations), counting (of objects), series (logical completion of a given number series), numerals (tasks in which numbers must be combined to obtain a desired target number), and tables (multiplication tables tasks from 1 to 10). We had a maximum observation period of 22 calendar weeks (Calendar Week 43 in 2019 to Calendar Week 12 in 2020; see Figure 3.3). The observation period varied for two reasons: First, the individual classes were tested over a period of three weeks at the beginning of the study (for logistical reasons), and second, three teachers did not provide the login data to their classes at the start of our study. On average, the program was available to the students for 20.7 weeks ( $SD = 1.61$ ,  $min = 17$ ,  $max = 22$ ). For the analyses, we coded the practice weeks for each student individually, starting with the week of the program implementation, when the students received their login data (week 0).

##### *3.2.4.2 Prior Mathematics Knowledge*

We used the addition and subtraction scales of the Heidelberg calculation test by Haffner et al. (2005) to measure prior mathematics knowledge. Each scale consists of 40 items that become increasingly difficult and students are instructed to solve as many items as possible in 120 seconds. The authors reported good retest-reliability values for the addition ( $r_{tt} = .82$ ) and subtraction scales ( $r_{tt} = .86$ ; Haffner et al., 2005).

##### *3.2.4.3 Conscientiousness*

The short scale of the Big Five Inventory for Children and Adolescents (German adaption) by Kupper et al. (2019) was used to measure conscientiousness. The scale consists of six items (e.g. "I am diligent and work fast."), which have to be assessed using a four-point Likert scale (1 = *does not apply at all* to 4 = *fully applies*). Cronbach's alpha was .66.

#### 3.2.4.4 Mathematics Self-Concept

Mathematics self-concept was measured with a scale adapted by Köller et al. (2000). The scale consists of five items (e.g. “I would prefer to do mathematics if the subject was not that difficult.”), which have to be answered on the same Likert scale as the conscientiousness scale. All items were reverse recoded, so that high scores represent a high mathematics self-concept. Cronbach’s alpha was .81.

#### 3.2.4.5 Mathematics Anxiety

Mathematics anxiety was measured with a subscale from the Mathematics Anxiety Rating Scale for Fourth to Sixth Grades by Roick et al. (2013). The scale consists of five items (e.g. “How nervous do you feel when a mathematics test is written?”), which have to be answered using a five-point Likert scale (1 = *not nervous at all* to 5 = *very nervous*). Cronbach’s alpha was .89.

#### 3.2.4.6 SES

We operationalised SES by the parental education background (Chen et al., 2018), which was requested with one item (“Does your mother or father have an Abitur [secondary school certificate qualifying students for university admission]?”), whereby the answer options were *no*, *yes*, and *I don’t know*. We were mainly interested in whether parents had an Abitur or not. However, to handle the *I don’t know* option, we split the SES into two dummies. In our analyses, we used SES–Abitur (0 = *no* and *I don’t know*, 1 = *yes*) as a predictor and SES–knowledge about parents’ Abitur (0 = *no* and *yes*, 1 = *I don’t know*) as a covariate.

#### 3.2.4.7 Migration Background

The migration background was requested with the following item “Are other languages spoken at home besides German?” with 0 = *no* and 1 = *yes* as the answer options.

#### 3.2.4.8 Covariates

At the time level, we considered a variable with the weekly measurement points (0–21) to control for a possible decline in practice behaviour. Moreover, there was a two-week Christmas holiday during our measurement period. We coded this event individually for each week (0 = *no holiday*, 1 = *holiday*). The COVID-19 lockdown affected the last week of our measurement period. We coded this event individually for each week (0 = *no school closures*, 1 = *schools were closed*).

At the students’ level, students’ gender was coded as 0 = *male* and 1 = *female*. We also controlled for the program introduction (0 = *no introduction*, 1 = *introduction*). Moreover, we controlled for technical problems. At the end of the measurement period, students were asked to answer the following item “Did you have problems getting the mathematics learning program started?” with the answer options 1 = *never*, 2 = *sometimes*, 3 = *often*, and 4 = *always*.

#### 3.2.5 Data Analysis

Due to the hierarchical structure of the data (time points within students), we used a two-level multilevel analysis (*Mplus*, Version 8.5; Muthén & Muthén, 2020). To investigate the degree to which students' practice behaviour was stable over time and, thus, could potentially be predicted by person rather than time-varying factors, we started with an unconditional model (null model) and estimated the percentages of variance between and within students. To obtain an impression of students' practice behaviour over time, we next established a Level-1 model (within person) with time (calendar weeks), Christmas holiday, and COVID-19 lockdown as predictors of students' weekly practice behaviour. To answer our research question, we tested a random intercept and fixed slope main effects model (Model 2) by adding the variables for individual characteristics as well as the covariates at the person level (Level-2 model; between person). Hence, we investigated whether students' individual characteristics predicted interindividual differences in students' average practice behaviour over the 22 weeks. We used the robust maximum likelihood (MLR) estimator to obtain standard errors that were robust to nonnormality (e.g., Bandalos, 2014; Maydeu-Olivares, 2017). Further, because students were nested within classes, we took the three-level nature of the data into account by using the `TYPE = COMPLEX TWOLEVEL` command. The `COMPLEX` command adjusts standard errors for nonindependence within classes, so nonindependence within classes was accounted for but not explicitly modelled. To ease the interpretation, all continuous independent variables were *z*-standardised ( $M = 0$ ,  $SD = 1$ ). As the dependent variable was not standardised, the beta-coefficients can be interpreted directly as effects on changes in completed tasks. Significance testing was undertaken at the 5% level.

We found missing data on both the dependent variable and the independent variables. Missing data regarding our dependent variable means that students did not have access to the program. There were two reasons for this. First, five students participated in the testing session and received login data but it was not possible to match their responses from the testing session to their practice data. Second, Math Garden was intended to be used over 22 weeks but, unfortunately, not all students were able to start in Week 1. Thus, for our analyses, the first week in which students were able to practise with Math Garden was coded as Week 0. All students practised for at least 16 weeks but, due to the delayed start of some students, missing data occurred from Week 17 (10.4%) to Week 21 (52.0%). The average missing rate (Week 0–Week 21) was 8.1%.

For the independent variables, missing data occurred for different reasons: students left out individual items or were not present during the testing session and only received access to



the learning program later and thus generated practice data ( $n = 8$ ). Lastly, the technical problems covariate was assessed after the practice period, which coincided with the COVID-19 lockdown in March 2020 in some classes. Although we tried to collect this information with an additional online survey, data for this covariate were missing for 47 students. The percentage of missing data varied from 7.4% on the Heidelberg calculation test to 24.3% on technical problems, with an average missing rate of 9.8%. We used full information maximum likelihood to account for missing data on the independent variables and on the dependent variable (Enders, 2010).

## 3.3 Results

### 3.3.1 Descriptives and Preliminary Analyses

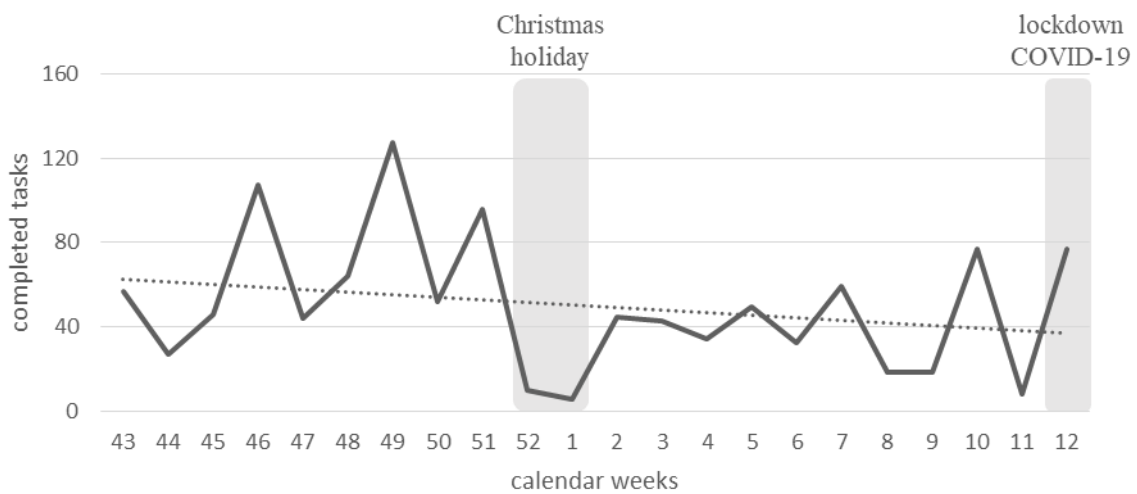
In Figure 3.3, the average number of tasks completed per calendar week by the students is displayed. On average, students solved 1,091 tasks over the whole observation period, although practice behaviour varied substantially ( $SD = 1,366.94$ ,  $min = 11$ ,  $max = 8,406$ ). The null model revealed an intraclass correlation (ICC) of .16, that is, 16% of the variance in practice with Math Garden was explained by differences between students, whereas 84% of the variance was due to differences within students over time. This result indicated that our investigation of the specific individual characteristics was worthwhile because they might explain differences in practice behaviour. Regarding possible relationships between individual characteristics and practice behaviour, we found significant correlations for mathematics anxiety and technical problems with practised tasks, each with  $r = -.22$  (see Table 3.2). These correlations can be described as being small (Cohen, 1988). The means, standard deviations, and correlations for the individual person characteristics are reported in Table 3.1.

Further, Figure 3.3 shows that there was an overall negative trend of practice over the 22 weeks and an obvious drop in practice during the Christmas holiday (Calendar Weeks 52 and 1). We found significant correlations between Christmas holiday, the COVID-19 lockdown, and time (calendar weeks) and practice behaviour, ranging between  $-.05 \leq r \leq -.10$ .

**Table 3.1***Descriptive Statistics and Correlation Matrix for Individual Characteristics*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9
<i>Individual characteristics</i>											
1. Prior mathematics knowledge	36.22	8.50									
2. Conscientiousness	2.79	0.57	.09								
3. Mathematics self-concept	2.72	0.81	<b>.26</b>	<b>.37</b>							
4. Mathematics anxiety	3.44	1.21	<b>-.15</b>	<b>-.25</b>	<b>-.47</b>						
5. SES–Abitur <sup>a</sup>	74.7%		-.05	<b>.15</b>	.12	-.06					
6. Migration background <sup>b</sup>	47.5%		.01	.12	.00	-.07	.08				
<i>Covariates</i>											
7. Gender <sup>c</sup>	49.5%		-.01	.02	<b>-.32</b>	<b>.19</b>	-.11	-.01			
8. Program introduction <sup>d</sup>	25.2%		.06	.00	.04	.07	.11	.06	-.07		
9. Technical problems	1.78	0.94	-.08	-.01	-.06	.03	.12	.02	-.03	.01	
10. SES–knowledge about parents' Abitur <sup>e</sup>	33.0%		.05	-.14	-.11	.00	<b>-.83</b>	-.03	.11	-.09	-.12

*Note.* Variables measured on a metric scale are presented as mean values, while reference categories of categorical variables are presented as percentages. Significant correlations are printed in bold ( $p \leq .05$ ). Reference categories: <sup>a</sup>no parent qualified for university/I don't know. <sup>b</sup>no other languages spoken at home besides German. <sup>c</sup>male. <sup>d</sup>no introduction. <sup>e</sup>no parent qualified for university/one parent qualified for university.

**Figure 3.3***Completed Tasks per Calendar Week*

### 3.3.2 Relationship Between Individual Characteristics and Practice Behaviour

Adding the individual characteristics as Level-2 predictors, the main effect we found was that of mathematics anxiety on practising with Math Garden (see Table 3.2, Model 2) after gender, program introduction, technical problems, and SES—knowledge about parents' Abitur had been controlled for. As expected (H5), students with mathematics anxiety had a lower level of practice with Math Garden ( $b = -14.17$ ,  $p = .006$ ). Concerning our hypotheses H1–H4 and H6, we did not find any relationships. To handle the high variance of the dependent variable, we additionally ran two robustness analyses, in which we successively excluded outlying values of practice behaviour. We provide the results of these additional analyses as Supporting Information. As the main results remained the same, with just one exception (the effect of mathematics anxiety was only marginally significant), we consider them to be robust.

**Table 3.2***Results of Multilevel Analyses for Practice Behaviour*

	Correlation				
	with		Model 1		Model 2
	practising				
	<i>r</i>	<i>b</i>	SE	<i>b</i>	SE
<i>Intercept</i>				114.54	26.30
<b>Within students (Level 1)</b>					
Calendar Weeks	<b>-.10</b>	<b>-3.00</b>	1.46	<b>-2.98</b>	1.49
Christmas Holiday <sup>a</sup>	<b>-.10</b>	<b>-51.94</b>	9.70	<b>-51.96</b>	9.72
Lockdown COVID-19 <sup>b</sup>	.05	54.78	48.55	54.55	49.06
<b>Between students (Level 2)</b>					
Prior mathematics knowledge	-.03			-4.15	8.11
Conscientiousness	.12			5.10	5.44
Mathematics self-concept	.08			-4.47	4.44
Mathematics anxiety	<b>-.22</b>			<b>-14.17</b>	5.12
SES–Abitur <sup>c</sup>	-.00			-19.37	14.53
Migration background <sup>d</sup>	-.01			-3.28	6.71
<i>Covariates II</i>					
Gender <sup>e</sup>	-.06			-5.28	10.68
Program introduction <sup>f</sup>	-.08			-7.20	28.27
Technical problems	<b>-.22</b>			<b>-13.33</b>	4.73
SES–knowledge about parents' Abitur <sup>g</sup>	-.05			-23.11	15.51
$R^2_{\text{within}}$			.03		.03
$R^2_{\text{between}}$					.13

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardised ( $M = 0$ ;  $SD = 1$ ).  $n_{\text{time}} = 22$ ;  $n_{\text{students}} = 202$ . Reference categories: <sup>a</sup>no holiday. <sup>b</sup>no lockdown. <sup>c</sup>no parent qualified for university/I don't know. <sup>d</sup>no other languages spoken at home besides German. <sup>e</sup>male. <sup>f</sup>no introduction. <sup>g</sup>no parent qualified for university/one parent qualified for university.

#### **3.4 Discussion**

In line with an adapted supply-use model (Helmke & Weinert, 1997; Seidel, 2014), the aim of this study was to investigate the predictive power of individual characteristics for practice behaviour with an adaptive arithmetic learning program. The results support only one of our expectations. While we were able to confirm our hypothesis about the relationship between mathematics anxiety and practice behaviour, our other hypotheses were not confirmed.

##### **3.4.1 Individual Characteristics as Predictors of Practising With an Arithmetic Learning Program**

Our results indicated that, as expected, mathematics anxiety negatively predicted the level of practice with the arithmetic learning program. Thus, highly mathematics anxious students were less likely to practise with such a program; this highlights the importance of encouraging these students to do so. As these students primarily focus on their unpleasant feelings when being confronted with mathematical content (Ashcraft & Moore, 2009), one support strategy when implementing such a program might be to provide a much longer guided introductory period to make students feel more comfortable with the program. In addition, for the first weeks of use, teachers could establish regular training units in mathematics lessons to familiarize the students with everyday practice. During such regular practice sessions, mathematics anxious students might perceive tasks as being controllable, which—in turn—may reduce anxiety (Pekrun, 2006) and result in more practice. Moreover, the assumptions of Orbach et al. (2019) indicate that, in particular, the success expectations of mathematics anxious students need to be fostered. For Math Garden, students have the option to manually increase their pre-set success probability from 75% to 90%, which leads to higher success rates and might then reduce the inhibition threshold to practise.

Contrary to our expectations, we did not find any relationship of prior knowledge or mathematics self-concept with practice behaviour. Keeping in mind that low-achieving students have low prior knowledge and mostly unfavourable motivational prerequisites (Chiu & Klassen, 2010; Van der Beek et al., 2017), the absence of a relationship between these variables is a positive finding as it might indicate that low-achieving students, who are, after all, most dependent on the use of supportive learning opportunities, are not disadvantaged in terms of practice behaviour. The reason for this finding (i.e., no statistically significant relationship) might be due to our decision to set the entry level of tasks in Math Garden to Grade-4 level in order to especially motivate low-achieving students to practise with the program. Thus, despite high expectations of success, students with high prior mathematics knowledge and a high

mathematics self-concept might not have seen the program as valuable because of the very easy entry tasks and, consequently, might not have practised with it more often than their classmates (Eccles & Wigfield, 2020). However, this assumption needs to be experimentally tested in future research and the hypothesised relationship needs to be explored as the results could be very helpful to practitioners. As the level of entry tasks can only be adjusted on the class level and not individually, teachers have to decide for whom practising with the program should be made more attractive, that is, for high- or low-achieving students. In line with previous research results from Riazzy et al. (2020), which showed that students with higher mathematics grades were more likely to persist until the end of a virtual learning environment even though the difficulty did not increase, we recommend choosing a lower entry level in order to avoid widening the knowledge gap. Moreover, keeping in mind the adaptive nature of the program, the missing relations suggest that the program algorithm works appropriately on different performance levels, as students with high prior knowledge or with a high mathematics self-concept did not complete fewer tasks over time. This means that students were not underchallenged during practice.

Also contrary to our expectations, we did not find any association between conscientiousness and practice behaviour even though evidence from homework research suggests that conscientiousness has an impact on learning behaviour (Trautwein et al., 2006; Trautwein & Lüdtke, 2007). In our study, we assume that sample characteristics may be the reason for this unexpected finding, for example, the younger age group compared to the studies of Trautwein and colleagues. As research on homework behaviour suggests that parental involvement is particularly high at a young age (Hoover-Dempsey & Sandler, 1997), parents of less conscientious children might especially try to compensate for their assumed lower practice behaviour by constantly reminding them to use the learning program provided. Therefore, parental involvement might explain more variance in practice behaviour than conscientiousness does for younger children; this assumption needs to be tested in further studies.

Concerning SES and migration background, we also did not find any relation with practice behaviour. Although we were not able to confirm our hypotheses, our results suggest that neither a lower SES nor a migration background has an educationally disadvantageous effect. This means that the use of learning programs at least does not create further inequality; this is a positive finding. A reason for the lack of a relationship with migration background could be that the program is mostly language-free and students with a migration background are therefore not at a disadvantage (due to a language barrier) in terms of practice behaviour.

Hence, one might further expect that there will be no disadvantages regarding the effects of the program but this needs to be tested in future research. The reason for the lack of statistical effects of SES could be due to an unfavourable operationalization of the construct, which is discussed in more detail in the next section.

#### **3.4.2 Strengths and Limitations**

A major strength of this study is the collection of practice behaviour with weekly log data, which were combined with test and questionnaire data. The long data collection period of 22 weeks is also noteworthy, as research in this field usually involves interventions with shorter time periods (Hillmayr et al., 2020). In addition, our study extends previous research on learning programs in mathematics by examining the relationship between different variables and practice behaviour with such learning programs; this has not previously been considered.

Although our study provides some new insights into students' practice behaviour with adaptive arithmetic learning programs, our analyses have some limitations that should be considered when interpreting the results. As most high-achieving students attend academic-track schools in Germany, our sample, which included only non-academic-track schools, consisted of, on average, low-achieving students. Thus, the results cannot be generalised for all types of schools in Germany. Further, the results cannot easily be generalised to different countries. Digital education has taken off in Germany in recent years but it is still behind the times, while in many other countries (e.g., Sweden, Denmark, or Singapore), digital solutions in education are well established (Ikeda, 2020). We expect that the fundamentally better digital infrastructure at schools in these countries might also have different effects on practice behaviour.

Moreover, previous research has already suggested that teachers vary in their competencies in and attitudes towards digital media and also that the digital infrastructure in schools varies (Eickelmann et al., 2019). Due to the design of our study, we were only able to control for a three-level structure, which accounted for students being clustered in classes. Unfortunately, we were not able to also consider teacher variables at Level 3 to investigate variance between classes.

The descriptive analyses for SES showed that 67.6% of students did not know whether or not their parents had an Abitur. This result suggests that the operationalization of SES as a student self-report measure of parents' educational background was unfavourable for the fifth graders in our sample. This was also indicated by the nonsignificant correlation between SES–Abitur and migration background. Numerous previous studies have found significant

correlations between these constructs (e.g., Eickelmann et al., 2019; OECD, 2013; 2019), which is why our results regarding SES must be interpreted with caution. We recommend that future research uses parent reports (e.g., occupation and cultural and economic capital) or other forms of student ratings, such as asking students about the number of books at home (Eriksson et al., 2019), to assess SES in a more valid way.

Moreover, a possible migration status was determined by the language(s) spoken at home. Even though many studies have assessed migration background in such a way (Kast et al., 2021; Konrad-Ristau & Burghardt, 2021), it might also be important to know for how long families have lived in Germany and which family members were born abroad. It would then be possible to distinguish between first- and second-generation migrants, which would be valuable information as previous research has suggested that first-generation migrants are more disadvantaged (OECD, 2019; Wendt et al., 2020). We recommend that future studies take this into account.

Further, we found high variance in our dependent variable, which is in line with earlier research that measured students' behaviour with log data in online learning environments (e.g., Hofman et al., 2018; Jansen et al., 2013). Although these log data are a major strength of our study, as there is no more objective way to measure behaviour than by tracking students' activities online, there are indeed some challenges when analysing the data as outliers may affect the results. We attempted to address this by using the MLR estimator and performing robustness analyses in which outliers were successively removed. Nonetheless, researchers should keep this in mind when interpreting the results.

#### **3.4.3 Theoretical and Practical Implications**

Even though we were able to confirm only one of the assumptions of our theoretical model, our study shows that adapting a supply-use model to the context of providing a mathematics learning program can offer new perspectives. Previous research on mathematics learning programs has focused mainly on performance as an outcome (Higgins et al., 2019; Pellegrini et al., 2021; Tokac et al., 2019) and has largely neglected the factors that determine the use of these programs. Thus, our study investigated a new and innovative theoretical assumption by considering individual student characteristics as predictors of practice behaviour. Yet, other determinants of students' use of learning programs may also be worth considering. Regarding the supply level, it is probably a little too simplistic to assume that only program processes have an influence on individual learning processes. In many cases, teachers are part of the provision of such a program (Ozel et al., 2008). Hence, a successful program implementation is



determined by teachers' competencies in and attitudes towards digital media (Eickelmann et al., 2019) and also by them having access to technical devices and support to enable them to confidently use the program (Ozel et al., 2008). Therefore, future research needs to take such teacher-related characteristics into account. However, the responsibility for (mathematics) education does not lie solely with teachers in schools; it also lies with parents at home (Collins & Halverson, 2010). Hence, with regard to the use of mathematics learning programs at home, students' practice behaviour might also be influenced by parental characteristics. Research on parental homework involvement, for example, showed positive effects for responsiveness and structure but negative effects for control (Dumont et al., 2014). These forms of parental involvement may well have an impact on the use of a learning program. Moreover, the supply-use model suggests that individual learning processes may also affect individual characteristics. Further research could test this bidirectional model assumption, as this has not been suggested in research on learning programs so far. For example, positive feedback on performance in the program, which a student gets especially if they practise a lot, might increase self-concept (Marsh & Martin, 2011) and decrease anxiety (Pekrun, 2006).

The results of our study have some practical implications for implementing a mathematics learning program in the classroom. For example, teachers should take into account that especially mathematics anxious students need to get extra support; otherwise, these students might practise less and, consequently, not have the chance to benefit from the implementation. Researchers have already suggested different psychological strategies to reduce mathematics anxiety. For instance, students' mathematics anxiety decreases when they perceive high teacher awareness of and responsiveness to their emotional needs (Authors, 2020 [details removed for peer review]). Thus, fostering teachers' sensitivity to mathematics anxious students' needs might help students to overcome their inhibited practice behaviour. Teachers can be sensitive to students' needs in a variety of ways: by actively and consciously looking for indicators that students need support or help, for instance, in students' behaviour, facial expressions, or gestures; by responding consistently to the needs of students, for example, by giving them enough time to think about tasks; and by solving problems effectively, which can be checked by having students confirm that their questions or problems have been resolved (Pianta et al., 2012). Furthermore, strengthening students' autonomy, providing positive feedback, and ensuring a positive error culture can also reduce mathematics anxiety (Brandenberger & Moser, 2018). In the context of practising with arithmetic learning programs, this could mean that it is beneficial if mathematics anxious students are initially allowed to decide for themselves which

domains they would like to work on. In addition, teachers should focus on appreciating students' effort in practice sessions and not on their results.

As regular practice is necessary in order for students to benefit from the program (Barzagar Nazari & Ebersbach, 2019), teachers should keep in mind that after the initial enthusiasm, practice with the program may decrease over time. We therefore recommend that practising with the program is repeatedly integrated into regular school lessons, for example, in the form of short units at school or as a homework assignment. Especially after school holidays, when practice behaviour declines dramatically, these strategies should be considered. Until now, such a holiday effect has been shown predominantly for the summer holidays (Alexander et al., 2012; Cooper et al., 1996). As we found a similar effect for a much shorter period (i.e., the Christmas holiday: two weeks), this indicates that such declines in learning behaviour occur in general during school holidays independent of the duration of the holiday. Thus, it seems important for teachers to motivate students to start practising more intensively after every holiday period.

### **3.5 Conclusions**

Although research on game-based learning and practice programs has recently intensified (Hillmayr et al., 2020; Pellegrini et al., 2021; Sailer & Homner, 2020), the relationship between individual characteristics and practice behaviour has not been addressed so far. This study thus adds a new perspective to the current state of research in this area. Regarding the question of which individual characteristics predict practice with an arithmetic learning program, our results suggest that, in particular, students with mathematics anxiety need more encouragement in order to at least get the chance to benefit from arithmetic learning programs. However, future research should specifically investigate further determinants of practice behaviour with mathematics learning programs, such as parental or teacher characteristics. In addition, research is needed on how different program components (i.e., adaptivity or feedback) may particularly foster the practice behaviour of students at risk, such as students with a migration background or students with low prior mathematics knowledge. Such research could gain valuable information on how to potentially close existing knowledge gaps.

### 3.6 References

- Aeschlimann, B., Herzog, W., & Makarova, E. (2015). Studienpräferenzen von Gymnasiastinnen und Gymnasiasten: Wer entscheidet sich aus welchen Gründen für ein MINT-Studium [Study preferences of high school students: who chooses STEM studies and for what reasons]? *Schweizerische Zeitschrift für Bildungswissenschaften*, 37(2), 285–300. <https://doi.org/10.25656/01:12646>
- Aldrup, K., Klusmann, U., & Lüdtke, O. (2020). Reciprocal associations between students' mathematics anxiety and achievement: Can teacher sensitivity make a difference? *Journal of Educational Psychology*, 112(4), 735–750. <https://doi.org/10.1037/edu0000398>
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2012). Schools, achievement and inequality: A seasonal perspective. In G.D. Borman & M. Boulay (Eds.), *Summer learning: Research, policy, and programs* (pp. 25–51). Routledge.
- Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, 14(2), 243–248. <https://doi.org/10.3758/BF03194059>
- Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational assessment*, 27(3), 197–205. <https://doi.org/10.1177/10734282908330580>
- Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(1), 102–116. <https://doi.org/10.1080/10705511.2014.859510>
- Barzagar Nazari, K., & Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology*, 33(2), 288–298. <https://doi.org/10.1002/acp.3485>
- Brandenberger, C. C., & Moser, N. (2018). Förderung der Lernfreude und Reduzierung der Angst im Mathematikunterricht in der Sekundarstufe 1 [Promoting the joy of learning and reducing anxiety in mathematics at secondary level]. In G. Hagenauer & T. Hascher (Eds.), *Emotionen und Emotionsregulation in Schule und Hochschule* (pp. 323–337). Waxmann.

- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and instruction*, 21(1), 95–108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics*, 118(3-4), 113–126. <https://doi.org/10.1111/ssm.12271>
- Chen, Q., Kong, Y., Gao, W., & Mo, L. (2018). Effects of socioeconomic status, parent-child relationship, and learning motivation on reading ability. *Frontiers in Psychology*, 9, 1297. <https://doi.org/10.3389/fpsyg.2018.01297>
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88-113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Chiu, M. M., & Chow, B. W.-Y. (2015). Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family socioeconomic status, educational resources, and attitudes toward reading. *Journal of Educational Psychology*, 107(1), 152–169. <https://doi.org/10.1037/a0036897>
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20(1), 2-17.
- Clark, R. E., & Sugrue, B. M. (1990). North American disputes about research on learning from media. *International Journal of Educational Research*, 14(6), 507–520. [https://doi.org/10.1016/0883-0355\(90\)90021-Y](https://doi.org/10.1016/0883-0355(90)90021-Y)
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.
- Collins, A., & Halverson, R. (2010). The second educational revolution: rethinking education in the age of technology. *Journal of Computer Assisted Learning*, 26(1), 18–27. <https://doi.org/10.1111/j.1365-2729.2009.00339.x>
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer holiday on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227–268. <https://doi.org/10.3102/00346543066003227>
- Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Psychological Assessment Resources.

- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1), 5–13.  
<https://doi.org/10.1037/1040-3590.4.1.5>
- Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In D. H. Saklofske, G. J. Boyle, & G. Matthews (Eds.), *The SAGE handbook of personality theory and assessment: Vol. 2: Personality measurement and testing* (pp. 179–198). SAGE. <https://doi.org/10.4135/9781849200479.n9>
- Dochy, F. J. R. C. (1992). Assessment of prior knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles. Lemma BV.
- Dochy, F. J. R. C., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, 69(2), 145–186. <https://doi.org/10.3102/00346543069002145>
- Dumont, H., Trautwein, U., Lüdtke, O., Neumann, M., Niggli, A., & Schnyder, I. (2012). Does parental homework involvement mediate the relationship between family background and educational outcomes? *Contemporary Educational Psychology*, 37(1), 55–69. <https://doi.org/10.1016/j.cedpsych.2011.09.004>
- Dumont, H., Trautwein, U., Nagy, G., & Nagengast, B. (2014). Quality of parental homework involvement: predictors and reciprocal relations with academic functioning in the reading domain. *Journal of Educational Psychology*, 106(1), 144–161.  
<https://doi.org/10.1037/a0034100>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61(4), 101859.  
<https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (2019). *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking [ICILS 2018 #Germany. Computer and information-related competencies of students in the second international comparison and competencies in the field of Computational Thinking]*. Münster: Waxmann.
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. Guilford Press.

- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2001). Keep the faucet flowing: Summer learning and home environment. *American Educator*, 25(3), 10–15, 47.
- Eriksson, K., Helenius, O., & Ryve, A. (2019). Using TIMSS items to evaluate the effectiveness of different instructional practices. *Instructional Science*, 47(1), 1–18. <https://doi.org/10.1007/s11251-018-9473-1>
- Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, 51(8), 1163–1176. <https://doi.org/10.1037/a0039440>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4: Heidelberger Rechentest – Erfassung mathematischer Basiskompetenzen im Grundschulalter [HRT 1-4: Heidelberger calculation test – Assessment of basic mathematical skills at primary school age]*. Hogrefe.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen [Conditional factors of academic performance]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie, Bd. 3: Psychologie des Unterrichts und der Schule* (pp. 71 – 176). Göttingen: Hogrefe.
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33–46. <https://doi.org/10.2307/749455>
- Higashi, R. M., Schunn, C. D., & Flot, J. B. (2017). Different underlying motivations and abilities predict student versus teacher persistence in an online course. *Educational Technology Research and Development*, 65(6), 1471–1493. <https://doi.org/10.1007/s11423-017-9528-z>
- Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research*, 57(2), 283–319. <https://doi.org/10.1177/0735633117748416>
- Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, 48(3), 45–53. <https://doi.org/10.1016/j.lindif.2016.02.006>

- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.  
<https://doi.org/10.1016/j.compedu.2020.103897>
- Hofman, A. D., Kievit, R., Stevenson, C., Molenaar, D., Visser, I., & Van der Maas, H. (2018). The dynamics of the development of mathematics skills: A comparison of theories of developing intelligence. Advance online publication.  
<https://doi.org/10.31219/osf.io/xa2ft>
- Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education?. *Review of educational research*, 67(1), 3–42.  
<https://doi.org/10.3102%2F00346543067001003>
- Ikeda, M. (2020). "Were schools equipped to teach – and were students ready to learn – remotely?". *PISA in Focus*, 108. OECD Publishing.  
<https://doi.org/10.1787/4bcd7938-en>
- Imhof, M., & Spaeth-Hilbert, T. (2013). The role of motivation, cognition, and conscientiousness for academic achievement. *International Journal of Higher Education*, 2(3), 69–80.
- Jansen, M., Schroeders, U., & Stanat, P. (2013). Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften [Motivational student characteristics in mathematics and science]. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle, & C. Pöhlmann (Eds.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (pp. 348–365). Waxmann.
- Jeno, L. M., Vandvik, V., Eliassen, S., & Grytnes, J.-A. (2019). Testing the novelty effect of an m-learning tool on internalization and achievement: A self-determination theory approach. *Computers & Education*, 128(3), 398–413.  
<https://doi.org/10.1016/j.compedu.2018.10.008>
- Kao, G., & Tienda, M. (1995). Optimism and achievement. The educational performance of immigrant youth. *Social Science Quarterly*, 76(1), 1–19.
- Kast, J., Lehofer, M., Tanzer, N., Hagn, J., & Schwab, S. (2021). Parents' attitudes towards students with a background of migration—Does the background of migration matter?. *Studies in Educational Evaluation*, 70, 101014.  
<https://doi.org/10.1016/j.stueduc.2021.101014>

- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824.  
<https://doi.org/10.1016/j.compedu.2011.02.003>
- Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse [Course choices of girls and boys in mathematics: On the role of subject-specific self-concept and interest]. *Zeitschrift für Pädagogische Psychologie*, 14(1), 26–37.
- Köller, O., Klemmert, H., Möller, J., & Baumert, J. (1999). Eine längsschnittliche Überprüfung des Modells des Internal/External Frame of Reference [A longitudinal test of the Internal/External Frame of Reference Model]. *Zeitschrift für Pädagogische Psychologie*, 13(3), 128–134.
- Konrad-Ristau, K., & Burghardt, L. (2021). Differences in children’s social development: How migration background impacts the effect of early institutional childcare upon children’s prosocial behaviour and peer problems. *Frontiers in Psychology*, 12, 614844. <https://doi.org/10.3389/fpsyg.2021.614844>
- Kupper, K., Krampen, D., Rammstedt, B., & Rohrmann, S. (2019). Kurzversion des Big Five Inventory für Kinder und Jugendliche [Short version of the Big Five Inventory for children and adolescence] (BFI-K KJ). *Diagnostica*, 65(2), 86–96.
- Lens, W., & Decruyenaere, M. (1991). Motivation and demotivation in secondary education: Student characteristics. *Learning and instruction*, 1(2), 145–159.  
[https://doi.org/10.1016/0959-4752\(91\)90024-3](https://doi.org/10.1016/0959-4752(91)90024-3)
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, 2(2), 99–106. <https://doi.org/10.1111/j.1750-8606.2008.00048.x>
- MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of Conscientiousness. *Learning and Individual Differences*, 19(4), 451–458. <https://doi.org/10.1016/j.lindif.2009.03.007>



- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 383–394. <https://doi.org/10.1080/10705511.2016.126960>
- McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 105–125). Kluwer Academic/Plenum Publishers. [https://doi.org/10.1007/978-1-4615-0763-5\\_6](https://doi.org/10.1007/978-1-4615-0763-5_6)
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, 82(1), 60. <https://doi.org/10.1037/0022-0663.82.1.60>
- Mohan, G., McCoy, S., Carroll, E., Mihut, G., Lyons, S., & MacDomhnaill, C. (2020). *Learning for all? Second level education during COVID-19 in Ireland*. ESRI. <https://doi.org/10.26504/sustat92.pdf>
- Muthén, B. O., & Muthén, L. K. (2008-2020). *Mplus computer software (version 8.5)*. Muthén & Muthén.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2007). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12(4), 323–345. <https://doi.org/10.1080/13803610600765687>
- Napoli, P. M., & Obar, J. A. (2014). The emerging mobile internet underclass: A critique of mobile internet access. *The Information Society*, 30(5), 323–334.
- OECD (2013). *PISA 2012 results in focus: What fifteen-year-olds know and what they can do with what they know*. OECD Publishing.
- OECD (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing.
- Orbach, L., Herzog, M., & Fritz, A. (2019). State- und Trait-Mathematikängste. Hemmende Prädiktoren mathematischer Leistungsfähigkeit [State and trait math anxiety. Performance-inhibiting predictors of mathematics achievement]? *Empirische Sonderpädagogik*, 11, 3–30.

- Ozel, S., Ebrar Yetkiner, Z., & Capraro, R. M. (2008). Technology in K-12 mathematics classrooms. *School Science & Mathematics, 108*, 80-85.  
<https://doi.org/10.1111/J.1949-8594.2008.TB17807.X>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open, 7*(1), 1–29.  
<https://doi.org/10.1177/2332858420986211>
- Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom Assessment Scoring System (CLASS)*. Teachstone.
- Poropat, A. E. (2014). A meta-analysis of adult-rated child personality and academic performance in primary education. *British Journal of Educational Psychology, 84*(2), 239–252. <https://doi.org/10.1111/bjep.12019>
- Portes, A., & Rumbaut, R. G. (2001). *Legacies. The story of the immigrant second generation*. University of California Press.
- Proulx, J. N., Romero, M., & Arnab, S. (2017). Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. *Simulation & Gaming, 48*(1), 81–97.  
<https://doi.org/10.1177/1046878116674399>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12 classrooms. *Journal of Educational Computing Research, 59*(1), 119–153.  
<https://doi.org/10.1177/0735633120952063>
- Reinhold, F., Reiss, K., Diedrich, J., Hofer, S., & Heinze, A. (2019). Mathematische Kompetenzen in PISA 2018 – aktueller Stand und Entwicklung [Mathematical competence in PISA 2018 – latest status and development]. In K. Reiss, M. Weis, E. Klieme, & O. Köller (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (pp. 187–210). Waxmann.
- Riazy, S., Simbeck, K., Woestenfeld, R., & Traeger, M. (2020). Prior knowledge as a predictor for persistence. In *Proceedings of the 12th International Conference on Computer Supported Education* (pp. 137–144). Science and Technology Publications.  
<https://doi.org/10.5220/0009324201370144>

- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19(6), 551–554. <https://doi.org/10.1037/h0033456>
- Roick, T., Göllitz, D., & Hasselhorn, M. (2013). Affektive Komponenten der Mathematikkompetenz. Die Mathematikangst-Ratingskala für vierte bis sechste Klassen (MARS 4-6) [Affective components of mathematics competence. The Mathematics Anxiety Rating Scale for 4th to 6th grade (MARS 4-6)]. In M. Hasselhorn, A. Heinze, W. Schneider & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen* (pp. 205–219). Hogrefe.
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma [Supply-use models in instructional psychology. Integration of structure and process paradigm]. *Zeitschrift für Pädagogik*, 60(6), 850–866. <https://doi.org/10.25656/01:14686>
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, 4(1), 27–41. <https://doi.org/10.1080/08917779108248762>
- Stanat, P., Segeritz, M., & Christensen, G. (2010). Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund [School-related motivation and aspiration of students with a migration background]. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung: Festschrift für Jürgen Baumert* (pp. 31–58). Waxmann.
- Stocké, V. (2007). Explaining educational decision and effects of families' social class position: An empirical test of the breen goldthorpe model of educational attainment. *European Sociological Review*, 23(4), 505–519. <https://doi.org/10.1093/esr/jcm014>
- Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64(1), 37–54. <https://doi.org/10.3102/00346543064001037>
- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35, 407–420. <https://doi.org/10.1111/jcal.12347>
- Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology*, 99(2), 432–444. <https://doi.org/10.1037/0022-0663.99.2.432>

- Trautwein, U., Lüdtke, O., Kastens, C., & Köller, O. (2006). Effort on homework in grades 5-9: Development, motivational antecedents, and the association with effort on classwork. *Child Development*, 77(4), 1094–1111. <https://doi.org/10.1111/j.1467-8624.2006.00921.x>
- Van der Beek, J. P., Van der Ven, S. H., Kroesbergen, E. H., & Leseman, P. P. (2017). Self-concept mediates the relation between achievement and emotions in mathematics. *British Journal of Educational Psychology*, 87(3), 478-495. <https://doi.org/10.1111/bjep.12160>
- Van Dijk, J. A. G. M. (2005). *The deepening divide: Inequality in the information society*. SAGE Publication.
- Vekiri, I. (2010). Socioeconomic differences in elementary students' ICT beliefs and out-of-school experiences. *Computers & Education*, 54(4), 941–950. <https://doi.org/10.1016/j.compedu.2009.09.029>
- Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, 83(4), 1137–1144. <https://doi.org/10.1111/j.1467-8624.2012.01781.x1>
- Wendt, H., Schwippert, K., Stubbe, T.C., & Jusufi, D. (2020). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund [Mathematical and scientific competencies of students with and without a migration background]. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky, & H. Wendt (Eds.), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 291–314). Waxmann.

## 3.7 Supplementary Tables

**Table S3.1***Results of Multilevel Analyses for Practice Behaviour*

	Correlation				
	with		Model 1		Model 2
	practising				
	<i>r</i>	<i>b</i>	SE	<i>b</i>	SE
<i>Intercept</i>				102.16	24.55
<b>Within students (Level 1)</b>					
Calendar Weeks	<b>-.11</b>	<b>-2.86</b>	1.30	<b>-2.88</b>	1.33
Christmas Holiday <sup>a</sup>	<b>-.12</b>	<b>-51.46</b>	9.70	<b>-51.42</b>	9.67
Lockdown COVID-19 <sup>b</sup>	.06	56.65	46.23	56.92	46.63
<b>Between students (Level 2)</b>					
Prior mathematics knowledge	.12			3.36	2.50
Conscientiousness	.08			0.84	2.78
Mathematics self-concept	.10			-3.05	3.32
Mathematics anxiety	<b>-.24</b>			<b>-11.03</b>	4.02
SES - Abitur <sup>c</sup>	.04			-24.86	14.02
Migration background <sup>d</sup>	-.04			-3.70	4.12
<i>Covariates II</i>					
Gender <sup>e</sup>	-.03			0.90	6.56
Program introduction <sup>f</sup>	.04			5.35	18.15
Technical problems	<b>-.22</b>			<b>-8.75</b>	2.07
SES - knowledge of parents' Abitur <sup>g</sup>	-.15			<b>-33.47<sup>†</sup></b>	14.34
$R^2_{\text{within}}$		.04		.04	
$R^2_{\text{between}}$				.17	

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Marginally significant coefficients are printed in italics ( $p \leq .10$ ). Continuous predictors were standardised ( $M = 0$ ;  $SD = 1$ ).  $n_{\text{time}} = 22$ ;  $n_{\text{students}} = 197$ . Reference categories: <sup>a</sup>no holiday. <sup>b</sup>no lockdown. <sup>c</sup>no parent qualified for university/I don't know. <sup>d</sup>no other languages spoken at home besides German. <sup>e</sup>male. <sup>f</sup>no introduction. <sup>g</sup>no parent qualified for university/one parent qualified for university. <sup>†</sup>Coefficient differ in significance level ( $p \leq .05$ ) in contrast to original analysis.

**Table S3.2***Results of Multilevel Analyses for Practice Behaviour*

	Correlation				
	with		Model 1		Model 2
	practising				
	<i>r</i>	<i>b</i>	SE	<i>b</i>	SE
<i>Intercept</i>				103.76	23.29
<b>Within students (Level 1)</b>					
Calendar Weeks	<b>-.14</b>	<b>-3.11</b>	1.25	<b>-3.12</b>	1.28
Christmas Holiday <sup>a</sup>	<b>-.11</b>	<b>-47.98</b>	8.83	<b>-47.93</b>	8.93
Lockdown COVID-19 <sup>b</sup>	.04	45.23	34.40	45.50	34.76
<b>Between students (Level 2)</b>					
Prior mathematics knowledge	<b>.19</b>			<b>4.98<sup>†</sup></b>	2.44
Conscientiousness	.10			1.38	2.76
Mathematics self-concept	.12			-2.24	3.00
Mathematics anxiety	<b>-.21</b>			<b>-7.49<sup>†</sup></b>	4.08
SES - Abitur <sup>c</sup>	.06			<b>-26.30<sup>†</sup></b>	12.08
Migration background <sup>d</sup>	.00			-1.02	4.01
<i>Covariates II</i>					
Gender <sup>e</sup>	-.07			-1.92	6.13
Program introduction <sup>f</sup>	.05			3.47	15.67
Technical problems	<b>-.22</b>			<b>-7.14</b>	1.99
SES - knowledge of parents' Abitur <sup>g</sup>	<b>-.20</b>			<b>-35.44<sup>†</sup></b>	11.39
$R^2_{\text{within}}$		.04		.04	
$R^2_{\text{between}}$				.20	

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Marginally significant coefficients are printed in italics ( $p \leq .10$ ). Continuous predictors were standardised ( $M = 0$ ;  $SD = 1$ ).  $n_{\text{time}} = 22$ ;  $n_{\text{students}} = 192$ . Reference categories: <sup>a</sup>no holiday. <sup>b</sup>no lockdown. <sup>c</sup>no parent qualified for university/I don't know. <sup>d</sup>no other languages spoken at home besides German. <sup>e</sup>male. <sup>f</sup>no introduction. <sup>g</sup>no parent qualified for university/one parent qualified for university. <sup>†</sup>Coefficient differ in significance level ( $p \leq .05$ ) in contrast to original analysis.

# 4

## ADDITIONAL REPLICATION ANALYSIS STUDY 1

### 4.1 Introduction

Replicating research is a key activity in scientific endeavor as it serves as a de facto reliability check (Plucker & Makel, 2021). Successful replications of previous results underpin their relevance and justify (sometimes costly) related actions. In contrast, establishing that results cannot be replicated provides stakeholders with information that can avoid resources and costs being consumed unnecessarily (Plucker & Makel, 2021). Replications can occur in two different forms, namely, direct replications and conceptual replications (Plucker & Makel, 2021). The goal of a direct replication is to follow the original study's methods as closely as possible, whereas a conceptual replication focuses more on investigating the theoretical mechanism behind a particular finding and less on the original study's methods (Schmidt, 2009; Simons, 2014), for example, investigating the suggested link in another context or, with regard to learning program research, by using a different program. Regardless of the form, replications are rare. In the field of educational psychology, only 0.13% to 1.07% of published results have been replicated (Makel & Plucker, 2014). Thus, researchers urgently need to increase replication rates. In light of this, I aimed to replicate Study 1 in terms of a direct replication, whereby I focused on using the same constructs (with slightly different operationalizations) as well as the same statistical methods but with a different and even larger sample.

### 4.2 Research Question and Hypotheses

Based on the research question addressed in Study 1, this replication also aimed to answer the question of how individual characteristics predict practice behavior (operationalized as the number of tasks completed per week) with the adaptive arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014). The results of Study 1 only found support for the hypothesis that math anxious students practiced less. Regarding the hypothesis that students with higher prior math knowledge, higher levels of conscientiousness or math self-concept, and a higher socioeconomic status (SES) would practice more, as well as the hypothesis that students with a migration background would practice less, no supporting empirical evidence was found. However, the unsupported hypotheses should not be directly rejected as they still have strong theoretical foundations. Thus, on the basis of the theoretical assumptions and the empirical results of Study 1, I again expected that (H1) prior math knowledge, (H2) conscientiousness, (H3) a high math self-concept, and (H5) a high SES would positively predict students' practice behavior (i.e., number of tasks completed), whereas (H4) math anxiety and (H6) a migration background would negatively predict practice behavior.



### 4.3 Methods

#### 4.3.1 Design and Procedure

The data used in this replication analysis come from a larger project investigating the effectiveness of different approaches to supporting students in math (i.e., Implementation Period 2; Carstensen et al., 2023). Within this project, students were randomly assigned at the school level to either an experimental group or a waitlist control group. For this replication, I considered the experimental group who practiced with Math Garden and their data, which were collected from October 2021 to January 2022; this group practiced for the same average number of weeks as the experimental group in Study 1, that is, 22 weeks. The implementation of Math Garden and the testing of students followed the same procedure as that used in Study 1.

#### 4.3.2 Sample

The sample comprised 890 fifth-grade students from 53 classes at 29 urban comprehensive (non-academic-track) schools across Germany, who participated in an intervention with Math Garden. Thus, the sample size was more than three times larger than that of Study 1 and also differed in the federal states included (Schleswig-Holstein in Study 1 vs. Berlin, Brandenburg, Hamburg, North Rhine-Westphalia, Rhineland-Palatinate, and Schleswig-Holstein in the replication). The data of two students were excluded due to an implausibly high level of practice (more than 2,000 tasks within one week), as were the data of 11 other students because of missing data on all relevant variables, such as questionnaire, test, and practice data. Moreover, four classes were excluded because they stopped participating in the study after the pretest ( $n = 48$ ). The final sample consisted of 829 students (48.7% female, 49.5% with a migration background, and 17.7% with parents who attended an academic-track school; see Table 4.1 for further sample characteristics).

#### 4.3.3 Measures

##### 4.3.3.1 Practice Behavior

In line with the original study, I used the number of math-related tasks completed in Math Garden per week as the dependent variable for the analyses, as well as a maximum observation period of 22 calendar weeks (Calendar Week 35 in 2021 to Calendar Week 10 in 2022; see Figure 4.1). On average, the program was available to the students for 18.70 weeks ( $SD = 1.4$ ,  $min = 14$ ,  $max = 22$ ). As in Study 1, I coded the number of weeks each student spent practicing individually for the analyses, starting with the week of the program implementation, which is when the students received their login data (Week 0).

#### 4.3.3.2 Individual Characteristics

As in Study 1, the addition and subtraction scales of the Heidelberg calculation test by (Haffner et al., 2005) were used to measure prior math knowledge. Similarly, the short scale of the Big Five Inventory for Children and Adolescents (German adaption) by Kupper et al. (2019) was used to measure conscientiousness. Cronbach's alpha was .62. Math anxiety was also measured with the subscale of the Math Anxiety Rating Scale for Fourth to Sixth Grades by Roick et al. (2013). Cronbach's alpha was .91.

In contrast, the assessment of math self-concept, SES, and students' migration background differed slightly to the assessment in Study 1. Regarding math self-concept, the German adaption of the math self-concept scale from the Self-Description Questionnaire was used (Arens et al., 2011; Marsh, 1990). The scale consists of eight items (e.g. "I like math."), which have to be answered on a five-point Likert scale (1 = *wrong* to 5 = *true*). Cronbach's alpha was .91.

SES was operationalized by the parents' educational background (Chen et al., 2018), but instead of asking about the Abitur (i.e., secondary school certificate qualifying students for university admission), students were asked to provide information about whether their parents had attended an academic-track school (i.e., Gymnasium in Germany; "Did your mother or father attend a Gymnasium?"). The answer options were *no*, *yes*, and *I don't know*. I was mainly interested in whether parents attended a Gymnasium or not. However, to handle the *I don't know* option, I split the SES into two dummies, which corresponded to the procedure in Study 1. In the analyses, I used SES–Gymnasium (0 = *no* and *I don't know*, 1 = *yes*) as a predictor and SES–knowledge Gymnasium (0 = *no* and *yes*, 1 = *I don't know*) as a covariate. As the results of Study 1 showed that only a few students knew about their parents' educational background, a second proxy to measure SES was used, which asked students about the number of books at home (Eriksson et al., 2019; "Approximately how many books are in your home?"). The answer options were 1 = *none or very few books (0–10 books)*, 2 = *enough to fill a bookshelf board (11–25 books)*, 3 = *enough to fill one bookshelf (26–100 books)*, 4 = *enough to fill two bookshelves (101–200 books)*, 5 = *enough to fill three or more bookshelves (over 200 books)*. Next to these answer options, the quantity of books was displayed as pictures.

Instead of assessing a possible migration background with a dummy-coded item (Study 1: "Are other languages spoken at home besides German?" with 0 = *no* and 1 = *yes* as the answer options), migration background in the replication was measured with the item "How often do you speak German at home?" and the following answer options: 1 = *I always speak German at home*, 2 = *I almost always speak German at home*, 3 = *I sometimes speak German*

*at home and sometimes another language*, 4 = *I never speak German at home*. For the analyses, I included the continuously operationalized migration background variable, but I also built a dummy-coded version in terms of the direct replication approach (0 = *I always speak German at home*, 1 = *I almost always speak German at home* to *I never speak German at home*).

#### 4.3.3.3 Covariates

In terms of keeping the replication as similar as possible to Study 1, I also considered the development of students' practice behavior over time in order to account for a possible decline. Thus, at the time level, I considered the weekly measurement points (Week 0–21). Further, I controlled for time-related events such as school holidays, as there was a two-week autumn holiday and a two-week Christmas holiday during the measurement period. I coded these events individually for each week (0 = *no holiday*, 1 = *holiday*). However, I did not take school closures as a consequence of the COVID-19 pandemic into account, as the data were collected in 2021/2022, which was after the peak of the pandemic.

On the student level, I controlled for students' gender (0 = *male* and 1 = *female*) and for a program introduction (0 = *no introduction*, 1 = *introduction*). As students were not asked about technical program problems, this covariate could not be included, in contrast to Study 1.

#### 4.3.4 Data Analysis

The statistical analysis followed the approach taken in Study 1, which means that, due to the hierarchical structure of the data (time points within students), I used a two-level multilevel analysis (*Mplus*, Version 8.5; Muthén & Muthén, 2020). I started with an unconditional model (null model) and estimated the percentages of variance between and within students, followed by a Level-1 model (within person) with time (calendar weeks) and holidays as predictors of students' weekly practice behavior. Finally, I tested three random intercept and fixed-slope main effects models by adding the variables for individual characteristics as well as the covariates at the person level (Level-2 model; between person). Hereby, Model A had the closest agreement with the Level-2 model from Study 1. Additionally, I varied measures regarding the migration background (Model B–migration background continuously) and measures regarding SES (Model C–SES books). I used the robust maximum likelihood estimator, and I also took the three-level nature of the data into account by using the `TYPE = COMPLEX TWOLEVEL` command. All continuous independent variables were *z*-standardized ( $M = 0$ ,  $SD = 1$ ), but the dependent variable was not standardized. Significance testing was undertaken at the 5% level.

I found missing data on both the dependent variable and the independent variables. Missing data regarding the dependent variable means that students did not have access to the program. There were two reasons for this. First, students participated in the testing session and received login data but it was not possible to match their responses from the testing session to their practice data ( $n = 161$ ). Second, Math Garden was intended to be used over 22 weeks but, unfortunately, not all students were able to start in Week 1. Thus, for the analyses, the first week in which students were able to practice with Math Garden was coded as Week 0. All students practiced for at least 14 weeks but, due to the delayed start of some students, missing data occurred from Week 15 (1.5%) to Week 21 (95.5%). The average missing rate (Week 0–Week 21) was 15.1%.

For the independent variables, students left out individual items or were not present during the testing session. Thus, they received access to the learning program later but still generated practice data. The percentage of missing data varied from 0% on students' gender to 11.7% on SES–books, with an average missing rate of 6.3%. I used full information maximum likelihood to account for missing data on the independent variables and on the dependent variable (Enders, 2010).

## **4.4 Results**

### **4.4.1 Descriptives and Preliminary Analyses**

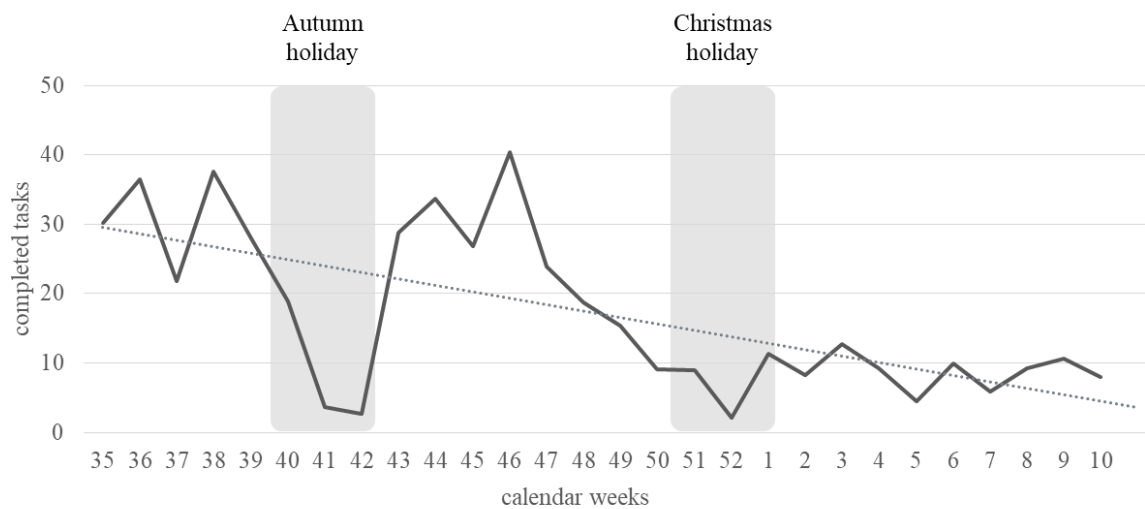
In Figure 4.1, the average number of tasks completed per calendar week by the students is displayed. On average, students completed 313 tasks over the whole observation period, although practice behavior varied substantially ( $SD = 625.04$ ,  $min = 0$ ,  $max = 7,771$ ). Thus, students, on average, engaged less with Math Garden than they did in Study 1. The null model revealed an intraclass correlation (ICC) of .17, which is very similar to the ICC in Study 1.

Regarding the possible relationships between individual characteristics and practice behavior, I found significant correlations for math anxiety ( $r = -.09$ ; see Table 4.2) and a migration background (dummy:  $r = -.07$ ; continuous:  $r = -.11$ ). These correlations can be described as being small (Cohen, 1988). The means, standard deviations, and correlations for the individual student characteristics are reported in Table 4.1. Further, Figure 4.1 shows that there was an overall negative trend of practice over the 22 weeks, with drops in practice during both holiday periods (Calendar Weeks 40 to 42, Calendar Weeks 51 to 1). I found significant but small negative correlations between the holiday periods and practice behavior ( $r = -.04$ ; see Table 4.2), as well as between time (calendar weeks) and practice behavior ( $r = -.03$ ).

**Table 4.1***Descriptive Statistics and Correlation Matrix for Individual Student Characteristics*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10
<i>Individual student characteristics</i>												
1. Prior math knowledge	40.47	9.66										
2. Conscientiousness	2.75	0.54	<b>.21</b>									
3. Math self-concept	3.43	0.93	<b>.40</b>	<b>.49</b>								
4. Math anxiety	3.19	1.14	<b>-.12</b>	<b>-.13</b>	<b>-.23</b>							
5. SES–Gymnasium <sup>a</sup>	17.7%		.02	<b>.12</b>	<b>.09</b>	-.01						
6. SES–books	2.69	1.24	.07	.05	<b>.08</b>	-.06	<b>.10</b>					
7. Migration background dummy <sup>b</sup>	49.5%		.03	.01	.04	<b>.12</b>	-.02	<b>-.11</b>				
8. Migration background continuous	1.95	0.96	-.03	.00	.01	<b>.16</b>	.02	<b>-.17</b>	<b>.88</b>			
<i>Covariates</i>												
9. Gender <sup>c</sup>	48.7%		<b>-.20</b>	-.04	<b>-.19</b>	<b>.19</b>	-.01	.02	<b>.10</b>	<b>.11</b>		
10. Program introduction <sup>d</sup>	93.2%		<b>.25</b>	.06	.00	.04	.02	-.01	<b>.26</b>	.04	-.02	
11. SES–knowledge Gymnasium <sup>e</sup>	53.6%		-.03	-.05	-.03	-.01	<b>-.61</b>	<b>-.08</b>	.04	.04	-.03	.05

*Note.* Variables measured on a metric scale are presented as mean values; reference categories of categorical variables are presented as percentages. Significant correlations are printed in bold ( $p \leq .05$ ). Reference categories: <sup>a</sup>no parent attended Gymnasium/I don't know. <sup>b</sup>no other languages spoken at home besides German. <sup>c</sup>male. <sup>d</sup>no introduction. <sup>e</sup>no parent attended academic-track school/one parent attended academic-track school.

**Figure 4.1***Number of Tasks Completed per Calendar Week*

#### 4.4.2 Relationship Between Individual Characteristics and Practice Behavior

Adding the individual characteristics as Level-2 predictors, I found a negative effect of students' migration background on practice behavior (see Table 4.2, Model A) after gender, program introduction, and SES—knowledge about parents' Gymnasium attendance were controlled for. The result was robust for both a dummy-coded variable and a continuous operationalization (see Model B). Thus, as expected (H6), students with a migration background displayed lower levels of practice with Math Garden (dummy:  $b_A = -3.52, p = .042$ ; continuous:  $b_B = -2.18, p = .010$ ). Concerning H1–H5, I did not find any relationships. Thus, with regard to Study 1, the pattern of findings differed slightly for migration background and math anxiety, but the results regarding prior math knowledge, conscientiousness, math self-concept, and SES were consistent. This also applied to the Level-1 variables, which means that students practiced less during holiday periods ( $b_A = -10.48, p < .001$ ) and that practice behavior declined over the 22 weeks ( $b_A = -0.50, p < .001$ ).

**Table 4.2***Results of Multilevel Analyses for Practice Behavior*

	Completed tasks	Model A		Model B		Model C	
	<i>r</i>	<i>b<sub>A</sub></i>	<i>SE</i>	<i>b<sub>B</sub></i>	<i>SE</i>	<i>b<sub>C</sub></i>	<i>SE</i>
<i>Intercept</i>		15.65	6.01	13.12	6.09	15.62	5.96
<b>Within students (Level 1)</b>							
Weeks	<b>−.03</b>	<b>−0.50</b>	0.15	<b>−0.49</b>	0.15	<b>−0.49</b>	0.15
Holidays <sup>a</sup>	<b>−.04</b>	<b>−10.48</b>	2.14	<b>−10.48</b>	2.14	<b>−10.47</b>	2.14
<b>Between students (Level 2)</b>							
Prior math knowledge	.04	0.31	1.04	0.24	1.04	0.33	1.03
Conscientiousness	.02	−0.10	1.04	−0.08	1.04	−0.14	1.04
Math self-concept	.04	0.64	0.99	0.62	0.97	0.64	0.99
Math anxiety	<b>−.09</b>	−1.44	1.00	−1.33	0.97	−1.46	1.01
SES–Gymnasium <sup>c</sup>	−.01	−1.05	1.75	−0.73	1.77		
SES–books	−.00					−0.24	0.71
Migration background dummy <sup>d</sup>	<b>−.07</b>	<b>−3.52</b>	1.74			<b>−3.57</b>	1.78
Migration background continuous	<b>−.11</b>			<b>−2.18</b>	0.84		
<i>Covariates</i>							
Gender <sup>e</sup>	−.05	−0.98	1.25	−0.93	1.23	−0.94	1.27
Program introduction <sup>f</sup>	.03	5.97	5.52	6.45	5.54	5.72	5.65
SES–knowledge Gymnasium <sup>g</sup>	.01	−0.11	1.35	0.04	1.34		
$R^2_{\text{within}}$		.02		.02		.02	
$R^2_{\text{between}}$		.10		.11		.10	

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors are standardized ( $M = 0$ ;  $SD = 1$ ).  $n_{\text{time}} = 22$ ;  $n_{\text{students}} = 829$ . Reference categories: <sup>a</sup>no holidays. <sup>b</sup>no lockdown. <sup>c</sup>no parent attended academic-track school/I don't know. <sup>d</sup>no other languages spoken at home besides German. <sup>e</sup>male. <sup>f</sup>no introduction. <sup>g</sup>no parent attended academic-track school/one parent attended academic-track school.

### 4.5 Discussion

With these additional analyses, I aimed to directly replicate Study 1, as replications in the field of educational psychology are needed to provide evidence for the robustness of the results. In line with Study 1, I expected that higher levels of prior math knowledge, conscientiousness, and math self-concept, as well as a higher SES, would be associated with more practice in Math Garden, whereas I expected math anxiety and a migration background to be related with less practice. In Study 1, empirical evidence was found only for the assumption that math anxious students practiced less, whereas no statistical links emerged for prior math knowledge, conscientiousness, math self-concept, SES, or a migration background with students' practice behavior.

Overall, the analysis of this replication yielded two results that differed from those of Study 1 but, otherwise, the same patterns of findings was found as that obtained in Study 1. The first difference from Study 1 was that students with a migration background practiced less with Math Garden in the replication, which was, however, in line with my hypothesis. The second difference was due to the fact that I could not replicate the negative link between math anxiety and practice behavior found in Study 1, which was, however, contrary to the theoretical assumptions that students with math anxiety avoid math-related tasks and situations (Eysenck et al., 2007). Regarding similar patterns of findings, the nonsignificant results regarding the effects of prior math knowledge, conscientiousness, math self-concept, and SES on students' practice behavior were in line with those from Study 1. However, these nonsignificant results did not support my proposed hypotheses.

Regarding math anxiety and migration background, interestingly, both student characteristics showed significant but also rather small negative associations with students' practice behavior. For math anxiety, the correlation was considerably smaller than that found in Study 1, which might be explained by the, on average, lower math anxiety level within the sample of the replication analyses. Regarding students' migration background, the operationalization of the variable might play a role in the negative relationship that emerged; instead of asking *if* other languages were spoken at home besides German, students provided information on *how often* they spoke German. Unlike in Study 1, this form of assessment provided students with more options and, therefore, might have created more meaningful variance which might have led to the small but expected effect found within the statistical analyses. Thus, after controlling for math anxiety, migration background was the more powerful



predictor in the statistical model of the additional analyses. This varying pattern of findings is addressed in more detail in the General Discussion (see Chapter 8).

For practitioners, this negative effect shows that students with a migration background might need more initial support when practicing with math learning programs in order to achieve a similar level of practice as their peers without a migration background. As a language barrier and less parental support at home might be the underlying mechanisms of this effect (Portes & Rumbaut, 2001; Dumont et al., 2012, 2014), teachers especially need to ensure that students with a migration background can follow the instructions when a math learning program is implemented. However, the tasks in the Math Garden program do not contain that much language. Thus, practitioners using programs with a greater emphasis on language, for instance, programs including word problems, might need to be even more aware of supporting students with a migration background. Otherwise, there is a risk that these students will be even more disadvantaged than is already the case (OECD, 2023). Even though the results do not provide evidence that math anxious students practice less, I would not conclude that math anxious students are not at risk, and I would continue to presume that they might need further support due to their lower achievement gains, which have clearly been shown in previous research (Barroso et al., 2021; Hembree, 1990; Ma, 1999; see Study 1 for suggested strategies).

The absence of a relationship between prior math knowledge, conscientiousness, math self-concept, SES, and practice behavior in Math Garden is again a positive finding. As especially low-achieving students (i.e., students with low prior math knowledge, a low math self-concept, and a low SES; Chiu & Klassen, 2010; Van der Beek et al., 2017) are most dependent on the use of supportive learning opportunities, the results of Study 1 and the replication analyses showed that students with these characteristics at least did not practice less. The replication of this pattern of findings might be a first indication that there might be no associations between these student characteristics and students' practice behavior with math learning programs. Still, it is necessary to keep in mind that the findings might only apply to the younger age group, in which parents and teachers often strongly structure students' learning environment (Eccles & Midgley, 1989). Such contextual factors might therefore be more decisive at this age than individual student characteristics are, and they could, in turn, become less important for older students. Furthermore, the results might only apply to practice with Math Garden, a program that works adaptively and, thus, provides students with higher chances of solving tasks successfully than of failing (Klinkenberg et al., 2011). Therefore, future studies might replicate this research in terms of a conceptual replication using a differently functioning learning program in order to get a better understanding of the underlying mechanisms. The

assumption about the role of adaptivity, and possible approaches to investigate the mechanisms behind adaptivity, will also be discussed in the General Discussion (see Chapter 8).

### 4.6 References

- Arens, A. K., Trautwein, U., & Hasselhorn, M. (2011). Erfassung des Selbstkonzepts im mittleren Kindesalter: Validierung einer deutschen Version des SDQ I [Assessment of self-concept in middle childhood: Validation of a German version of the SDQ I.]. *Zeitschrift Für Pädagogische Psychologie*, 25(2), 131–144.  
<https://doi.org/10.1024/1010-0652/a000030>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Carstensen, B., Hilz, A., Benckwitz, L., Asbury, K., Roloff, J., & Guill, K. Intervention in mathematics to professionally support low-achieving students. (2023). Retrieved from [osf.io/kdbcr](https://osf.io/kdbcr). Advance online publication. <https://doi.org/10.17605/OSF.IO/KDBCR>
- Chen, Q., Kong, Y., Gao, W., & Mo, L. (2018). Effects of socioeconomic status, parent-child relationship, and learning motivation on reading ability. *Frontiers in Psychology*, 9, 1297. <https://doi.org/10.3389/fpsyg.2018.01297>
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20(1), 2-17.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dumont, H., Trautwein, U., Lüdtke, O., Neumann, M., Niggli, A., & Schnyder, I. (2012). Does parental homework involvement mediate the relationship between family background and educational outcomes? *Contemporary Educational Psychology*, 37(1), 55–69. <https://doi.org/10.1016/j.cedpsych.2011.09.004>
- Dumont, H., Trautwein, U., Nagy, G., & Nagengast, B. (2014). Quality of parental homework involvement: predictors and reciprocal relations with academic functioning in the reading domain. *Journal of Educational Psychology*, 106(1), 144–161.  
<https://doi.org/10.1037/a0034100>
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York: Guilford Press.

- Eccles, J. S., & Midgley, C. (1989). Stage– environment fit: Developmentally appropriate classrooms for young adolescents. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 3. Goals and cognitions* (pp. 13– 44). Academic Press.
- Eriksson, K., Helenius, O., & Ryve, A. (2019). Using TIMSS items to evaluate the effectiveness of different instructional practices. *Instructional Science*, 47(1), 1–18. <https://doi.org/10.1007/s11251-018-9473-1>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion (Washington, D.C.)*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4: Heidelberger Rechentest – Erfassung mathematischer Basiskompetenzen im Grundschulalter [HRT 1-4: Heidelberger calculation test – Assessment of basic mathematical skills at primary school age]*. Göttingen: Hogrefe.
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33. <https://doi.org/10.2307/749455>
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>
- Kupper, K., Krampen, D., Rammstedt, B., & Rohrmann, S. (2019). Kurzversion des Big Five Inventory für Kinder und Jugendliche [Short version of the Big Five Inventory for children and adolescence] (BFI-K KJ). *Diagnostica*, 65(2), 86–96.
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 30(5), 520-540. <https://doi.org/10.2307/749772>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty. *Educational Researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189X14545513>
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82(4), 623–636. <https://doi.org/10.1037/0022-0663.82.4.623>
- Muthén, B. O., & Muthén, L. K. (2008-2020). *Mplus computer software (version 8.5)*. Los Angeles, CA: Muthén & Muthén.
- OECD (2013). *PISA 2012 results in focus: What fifteen-year-olds know and what they can do with what they know*. OECD Publishing.

- OECD. (2023). *PISA 2022 results: The state of learning and equity in education (Vol. I)*. OECD Publishing.
- Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, 56(2), 90–100.  
<https://doi.org/10.1080/00461520.2021.1895796>
- Portes, A., & Rumbaut, R. G. (2001). *Legacies. The story of the immigrant second generation*. Berkeley, CA: University of California Press.
- Roick, T., Göllitz, D., & Hasselhorn, M. (2013). Affektive Komponenten der Mathematikkompetenz. Die Mathematikangst-Ratingskala für vierte bis sechste Klassen (MARS 4-6) [Affective components of mathematics competence. The Mathematics Anxiety Rating Scale for 4th to 6th grade (MARS 4-6)]. In M. Hasselhorn, A. Heinze, W. Schneider & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen* (pp. 205–219). Göttingen: Hogrefe.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100.  
<https://doi.org/10.1037/a0015108>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 9(1), 76–80.  
<https://doi.org/10.1177/1745691613514755>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Van der Beek, J. P., Van der Ven, S. H., Kroesbergen, E. H., & Leseman, P. P. (2017). Self-concept mediates the relation between achievement and emotions in mathematics. *British Journal of Educational Psychology*, 87(3), 478–495.  
<https://doi.org/10.1111/bjep.12160>

# 5

## STUDY 2: HOW TO CONTINUE? NEW APPROACHES TO INVESTIGATING THE EFFECTS OF ADAPTIVE MATH LEARNING PROGRAMS ON STUDENTS' PERFORMANCE, SELF-CONCEPT, AND ANXIETY

Hilz, A., Guill, K., Roloff, J., Sommerhoff, D., & Aldrup, K. (2023). How to continue? New approaches to investigating the effects of adaptive math learning programs on students' performance, self-concept, and anxiety. *Journal of Intelligence*, 11(6), 108.

<https://doi.org/10.3390/jintelligence11060108>

### Abstract

Math learning programs were expected to revolutionize students' learning, but their effects so far have mostly been disappointing. Following the debate about *why* to continue research on math learning programs, we aimed to reformulate this question into *how* to continue this research. Investigations to date have neither considered a sufficiently wide set of outcome variables nor differentiated between performance measures (e.g., measuring addition and subtraction performance separately) and affective-motivational variables. Moreover, as students can only benefit from a program if they use it, researchers need to take practice behavior into account. Thus, we investigated whether the adaptive arithmetic learning program Math Garden fostered students' addition and subtraction performance, their math self-concept, and a reduction of their math anxiety. We also investigated how practice behavior (practiced tasks/weeks) affected these outcomes. We used a randomized pretest-posttest control group design with 376 fifth-grade students in Germany. Students in the experimental condition practiced with Math Garden for 20.7 weeks and had an increase in math self-concept. The more subtraction tasks the students practiced, the more they improved their subtraction performance. We found no effects on math anxiety. The results are discussed in terms of providing a starting point for new directions in future research.

**Keywords:** digital game-based learning; math learning program; intervention study; practice behavior; math performance

## 5.1 Introduction

It has always been a challenge for math teachers to provide instruction that meets the needs of all of their students, as students substantially differ in their prior math knowledge (Reinhold et al., 2019). This applies particularly to the transition from elementary school to secondary school (i.e., Grade 5 in most parts of Germany) when students from different elementary schools are regrouped into new secondary school classes. As the performance pressure in secondary schools increases, math teachers often do not have the time to address prior deficits and unlock students' full potential regarding basic arithmetic operations. Hence, especially students who have problems understanding basic arithmetic operations are at risk of falling behind because basic arithmetic skills form the basis of understanding more complex mathematics (Andersson, 2010; Bailey et al., 2014; Hansen et al., 2017; Jordan et al., 2013) and are a strong predictor of later achievement (Bailey et al., 2012; Barbieri et al., 2021; Duncan et al., 2007; Siegler et al., 2012). Math learning programs may help to meet this challenge due to their beneficial features, such as adaptivity and corrective feedback. They have, therefore, moved into the focus of research in recent decades (Higgins et al., 2019; Hillmayr et al., 2020; Pellegrini et al., 2021; Sailer & Homner, 2020).

Unfortunately, the results obtained up until now regarding achievement outcomes are disappointing. In their meta-analysis examining the overall effect size of digital game-based learning, Byun and Joung (2018) found a statistically significant but only small effect size ( $d = 0.37$ ). Hence, they concluded that there may be other ways for students to learn math more effectively (Byun & Joung 2018). Similarly, in their meta-analysis, Tokac et al. (2019) found only a small, marginally significant overall effect ( $d = 0.13$ ) for game-based interventions in math compared to traditional nonvideo game-based classroom instruction, and the authors also pointed out the high heterogeneity regarding the effectiveness of such game-based interventions. Pellegrini et al. (2021), who also found only a very small nonsignificant effect size ( $d = 0.05$ ) concerning the implementation of digital technology in math instruction, sharpened the topic by pointing out that the use of technology in education has been expected to have revolutionary impacts on learning outcomes but that, no matter how beneficial and promising program mechanisms sound, they have not yet improved students' performance substantially. But what is the educational implication of these disappointing results? Should learning programs no longer be recommended for use in the classroom because research is not able to show that their implementation is effective? Rather than assuming that math learning programs are ineffective and potentially rejecting their use, one could take a step back and ask

whether the methodological approaches of previous studies were fully capable of detecting potential effects.

Thus, regarding the question of *why* to continue research on math learning programs, we suggest this question is reformulated into *how* to continue research on math learning programs. Posing this question is highly relevant because prematurely rejecting the use of math learning programs might lead to these interventions being abandoned, even though they are easily applicable and cost-effective and can thus reach large numbers of students. Thus, the present study identified current desiderata in research on math learning programs, addressed them conceptually, and then analyzed the effects of an adaptive computer-based arithmetic learning program with a randomized pretest-posttest control group design in a sample of fifth-grade students in Germany ( $N = 376$ ). The approaches postulated in this study might be a starting point for future research on math learning programs.

### 5.1.1 Promising Mechanisms of Math Learning Programs

The central promising characteristics of math learning programs are their adaptivity and corrective feedback (Hillmayr et al., 2020). Adaptivity in this context means that the task difficulty adapts to the individual ability level of each learner (Hillmayr et al., 2020), while corrective feedback highlights correct solutions and corrects incorrect solutions (Shute, 2008). There is a wide variety of interactive math learning programs that include these desirable mechanisms and that have also already been used for scientific purposes, for example, Bettermarks from Germany, Mindsteps from Switzerland, or Carnegie Learning from the U.S., but they all differ in their conceptual functioning (Hillmayr et al., 2020).

In this study, we used the adaptively working arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014). Following the characterization by Nattland and Kerres (2009), Math Garden can be classified as a drill-and-practice program. Drill-and-practice programs serve to strengthen previously acquired content knowledge by allowing the learner to practice at their own pace and to repeat specific types of exercises as often as necessary (Nattland & Kerres, 2009). In line with this definition, Math Garden allows students to strengthen basic arithmetic operations that are central to elementary school curricula on individual performance levels because an algorithm regulates the assignment of the tasks so that students have an average percentage of correct tasks of 75% (Klinkenberg et al., 2011). Moreover, after each task, direct corrective feedback is given. The interplay of these promising mechanisms might affect not only students' performance but also their affective-motivational outcomes, such as math self-concept and math anxiety, as further elaborated below.



### *5.1.1.1 Effect Mechanisms of Math Learning Programs on Math Performance*

The cognitive theory of multimedia learning (Mayer, 2014) describes why using digital learning programs such as Math Garden can be beneficial for students' performance: Students need to actively engage with the learning content in order to understand new information (Mayer, 2014). Hence, interactive digital learning programs help them to directly influence their own learning processes, as the defining feature of interactivity is responsiveness to the learner's actions during learning (Moreno & Mayer, 2007). Thus, programs such as Math Garden respond to students' ability level by providing appropriate tasks, and immediate program feedback helps students to instantly reflect on their performance and, if necessary, to rethink inefficient calculation strategies (Shute, 2008). Besides performance, similar effect mechanisms might also affect math self-concept.

### *5.1.1.2 Effect Mechanisms of Math Learning Programs on Math Self-Concept*

Math self-concept refers to the self-evaluation of one's ability in math and is considered to be an important factor concerning performance and achievement-related choices (Eccles & Wigfield, 2020). According to the reciprocal effects model (Marsh, 1990; Marsh & Martin, 2011), performance and math self-concept enhance each other. This has been confirmed by several studies (Arens et al., 2022; Jiang et al., 2020). As Math Garden offers high success rates independent of students' actual skill level because problems are adapted to the individual skill level during practice (Klinkenberg et al., 2011), all students can be expected to experience positive feedback and rewards, which then should imply good performance. Thus, they might perceive their own abilities as being high, generalize this positive feedback to an entire subject area (e.g., math), and internalize it into a positive self-concept for that subject area (Craven et al., 1991). Hence, receiving feedback that they have solved math tasks successfully should lead to an increase in students' math self-concept. In addition to math self-concept, math anxiety is another prominent affective-motivational construct that has been shown to have a direct relation to students' performance (Ashcraft & Kirk, 2001; Caviola et al., 2022; Devine et al., 2018; Ma, 1999; Pekrun, 2006).

### *5.1.1.3 Effect Mechanisms of Math Learning Programs on Math Anxiety*

Richardson and Suinn (1972) defined math anxiety as a "feeling of tension and apprehension that interferes with the manipulation of numbers and the solving of math problems in a wide variety of ordinary life and academic situations" (p. 551). The control-value theory (Pekrun, 2006) postulates that (math) anxiety occurs when the value of an achievement-related task is perceived as high but solving the task is considered to be uncontrollable due to missing resources, such as a lack of knowledge and skills. However, as adaptive programs such as Math

Garden adjust the task difficulty to the learner's skill level, students with math anxiety should perceive the tasks as being (more) controllable in terms of solving the tasks correctly, which should reduce their anxiety. Positive program feedback should help students to recognize the controllability of tasks even further. Moreover, programs such as Math Garden offer students the chance to make mistakes in private. As negative experiences with math in public appear to contribute to math anxiety (Bekdemir, 2010), students should feel safer making mistakes during practice with a math learning program than during oral practice sessions with the whole class. Hence, if the number of public, potentially negative, or embarrassing math-related experiences is reduced, students might experience even further control.

Nonetheless, even though the use of arithmetic learning programs like Math Garden in school as an additional instructional feature to strengthen arithmetic skills and to foster affective-motivational outcomes is promising in theory, researchers often detected no effects or only small effects on math performance (Bai et al., 2012; Byun & Joung, 2018; Hung et al., 2014; Jansen et al., 2013; Pareto et al., 2012; Pellegrini et al., 2021; Tokac et al., 2019). In contrast to performance effects, affective-motivational effects have not been investigated much so far. Hence, we considered whether a change in research on math learning programs might be necessary to improve the quality of studies in this field. First studies have already started to investigate math learning programs in a more differentiated way. For instance, Hallstedt et al. (2018) investigated addition and subtraction performance separately, Jansen et al. (2013) considered not only performance as an outcome but also affective-motivational variables to examine the effectiveness of math learning programs while also taking practice behavior into account, and Vanbecelaere et al. (2022) also asked for a systematic and more standardized assessment framework of digital games for learning. Our aim was to build upon these studies and combine their methodologies in one approach.

### **5.1.2 New Approaches to Investigating the Effectiveness of Math Learning Programs**

When considering how future studies on math learning programs can be improved, and after reviewing the literature on math learning program evaluations, in our view, three different approaches can be taken to investigate the effectiveness of math learning programs in new ways.

#### *5.1.2.1 Measuring Distinct Subdomain Performance*

As a key learning outcome, researchers mostly aim to measure students' math performance improvement after implementing a math learning program. To do so, they usually assess performance with a broad status diagnostic score including domains such as measurement, geometry, or arithmetic (Bai et al., 2012; Beserra et al., 2014; Ke, 2008; Shin et al., 2012).

However, such an overall performance score may not be able to reveal improvements in more specific math subdomains, such as different arithmetic operations (e.g., addition or subtraction). Hence, Ran et al. (2021) called for more specific outcomes when investigating math learning programs, arguing that an overall score might lead to flawed conclusions about the effectiveness of the intervention. For instance, a significant performance improvement as shown in an overall arithmetic score does not necessarily imply an equally large performance improvement in addition and subtraction, whereas, in turn, no performance in an overall arithmetic score does not necessarily mean that there is no improvement in addition and subtraction. To the best of our knowledge, so far, only Hassler Hallstedt et al. (2018) differentiated between single performance measures for addition and subtraction skills in their analyses. They were able to show that a math tablet intervention had effects on low-achieving second graders. They found a performance increase in both addition and subtraction (Hassler Hallstedt et al., 2018). This underlines the importance of using distinct performance measures, as the study clearly showed that the intervention was effective for both addition and subtraction. Hence, more research that differentiates between performance measures is needed to gain a more detailed insight into which exact math abilities are fostered with the learning program being implemented. Such research would probably reveal hidden effects.

### *5.1.2.2 Affective-Motivational Outcomes*

In addition to the differentiated effects on performance that have not been sufficiently researched thus far, the effects on affective-motivational outcomes have also been mostly neglected. Given that math self-concept and math anxiety are essentially related to students' math performance and their engagement in math (Ashcraft & Krause, 2007; Chinn, 2009; Guay et al., 2003; Marsh & Martin, 2011), a learning program, even if it does not increase performance in the first step, can be effective in the long term by improving math self-concept or reducing math anxiety. Indeed, the results of a meta-analysis by Fadda et al. (2022), which is, to the best of our knowledge, the only existing meta-analysis regarding the motivational effects of math learning programs, showed that providing math learning programs can foster students' motivation. The authors pointed out that the highest effects were found in studies that operationalized motivation as expectancy of success and related beliefs about competence (i.e., self-concept) in contrast to value components (Fadda et al., 2022). However, as only three studies were of high methodological quality and met the criteria to be included in the meta-analysis regarding the effects of math self-concept, more research is needed here. Further, only very few studies so far have focused on the effects of math learning programs on the reduction of math anxiety. In their intervention study, Huang et al. (2014) showed a decrease in math

anxiety from pretest to posttest, whereas Hung et al. (2014) did not find any effects of their intervention on math anxiety. However, both studies only had a small experimental sample, with 25 students each per condition. Moreover, the studies differed substantially in their number of treatments during the intervention: Huang et al. (2014) had 40-min sessions twice a week over six weeks, therefore, this was a more distributed intervention, whereas, in contrast, Hung et al. (2014) only had one 240-min session. Hence, this was a more massed intervention. These results might indicate that a more distributed intervention over a longer period of time is needed to reduce math anxiety, as students probably need some time to perceive tasks as being controllable. However, the authors only supervised the practice sessions and did not control for the number of tasks that were actually practiced during their intervention. This lack of data brings us to the last desideratum.

### *5.1.2.3 Considering Practice Behavior*

Many studies only provided students with a learning program and then evaluated whether such an intervention made a difference in the performance of this group compared to a wait-list control group (Bai et al., 2012; Chang et al., 2015; Huang et al. 2014; Hung et al. 2014; Shin et al. 2012). This approach neglects the actual practice behavior of the students. However, as researchers have already shown in meta-analyses that short-term math learning program interventions lead to higher effectiveness, due to higher student engagement at the beginning of the intervention (novelty effect) than long-term interventions (Hillmayr et al., 2020; Sung et al., 2017), practice behavior needs to be considered.

The body of literature on practice behavior shows that students need to practice regularly over a longer period of time to achieve the best performance outcomes (Barzagar Nazari & Ebersbach, 2019; Carpenter et al., 2012; Vlach & Sandhofer, 2012). Thus, it is reasonable to assume that a math learning program may fail to have effects because students simply practiced too little with the program. This assumption cannot be tested simply by investigating the effects of the provision of a learning program. Indeed, some experimental studies supervised students' practice sessions with a math learning program in classrooms to ensure that they practiced a given number of tasks (Huang et al., 2014; Rodrigo, 2011; Shin et al., 2012; Van den Heuvel-Panhuizen et al., 2013), and all were able to show student performance gains. However, this approach can only provide limited information about how students interact with the program when they are not supervised, not only in school but, for instance, also at home. With regard to increasing digitalization and the possibility of using log and trace data, researchers should take advantage of this potential to uncover hidden behavioral patterns in a more naturalistic setting (Baker & Inventado, 2014; Haleva et al., 2021). Unfortunately, the number of studies that took

log and trace data into account is still rather modest. The few studies referring to such data mostly considered the time spent practicing with math learning programs (Haleva et al., 2021; Hassler Hallstedt et al., 2018; Louw et al., 2008; Spitzer, 2022), and they all were able to show that time spent practicing improved student performance. Another approach that can be employed to take practice behavior into account is to consider the number of tasks practiced by the students in the program. This is more of an in-depth approach in contrast to the time spent practicing as, for example, students can solve 10 tasks in 10 min if they concentrate but can solve only two tasks in the same 10 min if they are distracted. Jansen et al. (2013) showed that performance improvement in the program Math Garden was mediated by the number of tasks attempted. The authors also considered practice behavior in terms of practiced tasks when investigating effects on motivation-related variables (i.e., perceived math competence) and math anxiety, but they did not find any effects (Jansen et al., 2013). This might be due to practiced tasks not being the optimal indicator to show the effects of practice behavior on affective-motivational variables. Another alternative, which has not yet been considered in research, would be to include the number of weeks over which students practiced with the program as practice behavior (referred to from here on as “practiced weeks”). As the practiced tasks are more a measure of quantity, the practiced weeks better reflect how regularly students practiced with a program (i.e., distributed practice behavior). As previous studies have already shown that students benefit most from math learning opportunities if their practice behavior is distributed (Barzagar Nazari & Ebersbach, 2019; Schutte et al., 2015), including the number of practiced weeks could represent a new and valuable facet of practice behavior.

### **5.1.3 The Present Study**

This study addresses three research desiderata on math learning programs and provides empirical data that can be used for future investigations. First, we used distinct arithmetic performance measures (i.e., addition and subtraction) to investigate potential performance increases on a detailed level. Second, we considered affective-motivational outcomes (i.e., math self-concept and math anxiety) to investigate in which different ways students may benefit from the implementation of Math Garden. Third, we took practice behavior in terms of practiced tasks and practiced weeks in Math Garden for all outcomes into account (i.e., performance and affective-motivational variables), as missing effects might occur because students practiced too little. This resulted in three research questions regarding the provision of Math Garden and students’ actual practice behavior in terms of practiced tasks and practiced weeks:

How does the provision of Math Garden affect students' addition and subtraction performance, their math self-concept, and their math anxiety? On the basis of the theoretically assumed promising program features, we expected that the provision of Math Garden would improve (H1.1) students' addition and (H1.2) subtraction performance. We assumed that the program would additionally (H1.3) increase students' math self-concept and (H1.4) decrease students' math anxiety.

How does the number of addition and subtraction tasks practiced affect students' addition and subtraction performance, respectively? How does the number of all math-related tasks (including addition, subtraction, and other types of math tasks, see Section 2.3.4) practiced in Math Garden affect students' math self-concept and their math anxiety? We expected that (H2.1) the more students practiced addition tasks during the intervention, the higher their addition performance after the intervention would be. We expected that (H2.2) the more students practiced subtraction tasks, the higher their subtraction performance would be. Moreover, we expected that students who practiced more math-related tasks overall would have (H2.3) a higher math self-concept and (H2.4) lower math anxiety after the intervention because they received more positive feedback due to the adaptive algorithm.

How does the number of practiced weeks spent on addition and subtraction tasks (referred to from here on as “practiced weeks of addition or of subtraction”) affect students' addition and subtraction performance, respectively? How does the total number of practiced weeks (including weeks spent practicing addition, subtraction, and other types of math tasks, see Section 2.3.4) in Math Garden affect students' math self-concept and their math anxiety? We expected that (H3.1) students who spent more weeks practicing addition would have higher addition performance. We expected that (H3.2) students who spent more weeks practicing subtraction would have higher subtraction performance. Moreover, we expected that students who spent more weeks practicing with Math Garden overall would have (H3.3) a higher math self-concept and (H3.4) lower math anxiety after the intervention.

We addressed our research questions and hypotheses with a pretest-posttest control group design with a random assignment of students on the class level, assessing the suggested student outcomes with objective measures outside of Math Garden, which is a major strength of this study. The time period spent practicing with the program in our study was substantial ( $M = 20.7$  weeks), and we had a large sample ( $N = 376$ ) with similar numbers of students in both groups. We chose to use Math Garden because it is a low-cost program that was designed to be easily implemented in schools and to track students' practice behavior. To ensure the robustness of our findings, we considered several important control variables in our analyses. We controlled

for students' pretest scores to account for differences in their baseline levels. Moreover, we controlled for students' gender and a possible migration background because previous studies showed that these variables had effects on our outcome variables (Liu & Wilson, 2009; Marsh & Yeung, 1998; Van Mier et al., 2018). As addition and subtraction performance was measured on tablets in a speed test, we also checked for typing speed (Hassler Hallstedt et al., 2018).

### **5.2 Materials and Methods**

#### **5.2.1 Design and Procedure**

Student data were collected in schools from October 2019 (T<sub>1</sub>) to March 2020 (T<sub>2</sub>) in compliance with data protection requirements, and the study was approved by the institutional ethics committee. Participation in the study was voluntary and parental permission was obtained before students participated in the study.

The participating classes completed math performance tests and answered questionnaires assessing the relevant variables of interest for our study. A testing session lasted approximately 1.5 h and was done with tablets. All students were tested within a three-week period. At the end of the initial testing session, the classes were randomly assigned either to an experimental group or to a wait-list control group. Hence, variance differences between schools were controlled for. The students in the experimental group received a 10-min introduction to Math Garden and accessed the program on tablets using personalized login data provided by their teachers. While practicing, they got a brief oral introduction to the main program principles. At the end of the introductory practice session, students were told that they were allowed to, and should, practice with the program at home and in class on available digital devices (e.g., desktop computer, tablet, or smartphone). Their teacher also received a short written introduction to Math Garden and was asked to integrate the program into regular math lessons, to assign it for homework, or to use it during supervision classes, which provided him or her with many opportunities to individually adapt the use of Math Garden to meet all students' needs. Students could access the program for, on average, 20.7 weeks. Their weekly practice behavior was tracked. Three classes did not have personal login data during the initial testing session due to technical problems. These students were introduced to the program in a 5-min presentation. The teachers were asked to introduce the students to the program in more detail later. In March 2020, after 22 weeks, the classes were tested again, and the students answered questionnaires. After this testing session, the wait-list control group received access to Math Garden.

### 5.2.2 Sample

The data for this study were obtained from a sample of 376 fifth graders from 21 classes at seven urban comprehensive (non-academic-track) schools in northern Germany. While academic-track schools in Germany prepare students primarily for higher education, non-academic-track schools offer different levels of school degrees (Maaz et al., 2008).

We excluded one student due to an implausibly high level of practice (more than 2000 tasks within one week), which might have been caused by siblings also practicing with the program, and we excluded another five students because they had missing data on all relevant variables, such as questionnaire data and practice data. The final sample then consisted of 200 students in the experimental condition (50.0% male, 53.4% with a migration background) and 170 students in the wait-list control condition (52.7% male, 43.2% with a migration background). The two conditions did not differ significantly in their pretest mean values (see Table 5.1). A priori power analyses (Faul et al., 2007; 2009) showed that for between-group analyses (Research Question 1), setting the estimated effect size  $d$  to 0.30 and assuming a power of at least .80 with an  $\alpha$ -level of .05, a sample size of 176 students per group would be needed. Further, for within-group analyses (Research Questions 2 and 3), setting the estimated effect of  $R^2$  to .08, assuming a power of at least .80 with an  $\alpha$ -level of .05, a sample size of 154 students would be needed. Unfortunately, the posttest coincided with the COVID-19 lockdown in March 2020 in eight classes, which led to missing data on all posttest scores for 66 (33%) students in the experimental condition and for 50 (29%) students in the wait-list control condition. We did not exclude these students from our analytic sample (see Section 2.4 for missing data handling).



**Table 5.1***Descriptives of Self-Report and Performance Measures*

	Condition					
	All Students		Wait-List Control		Experimental	
	<i>N</i> = 370		<i>n</i> = 170		<i>n</i> = 200	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Math addition performance T <sub>1</sub>	5.61	3.12	5.74	3.23	5.49	3.05
Math addition performance T <sub>2</sub>	7.90	2.77	7.83	2.91	7.97	2.66
Math subtraction performance T <sub>1</sub>	5.20	3.63	5.41	3.58	5.01	3.67
Math subtraction performance T <sub>2</sub>	6.62	3.69	6.71	3.53	6.56	3.85
Math self-concept T <sub>1</sub>	2.70	0.82	2.67	0.84	2.72	0.81
Math self-concept T <sub>2</sub>	2.73	0.83	2.59	0.86	2.85	0.78
Math anxiety T <sub>1</sub>	3.36	1.21	3.28	1.22	3.44	1.21
Math anxiety T <sub>2</sub>	3.20	1.26	3.17	1.22	3.23	1.30
Gender <sup>a</sup>	0.49	0.50	0.47	0.50	0.50	0.50
Migration background T <sub>1</sub> <sup>b</sup>	0.49	0.50	0.43	0.50	0.53	0.50
Tablet typing speed T <sub>1</sub>	7.88	2.36	8.12	2.27	7.66	2.43

*Note.* Reference categories: <sup>a</sup> male. <sup>b</sup> no other languages spoken at home besides German.

### 5.2.3 Measures

#### 5.2.3.1 Addition and Subtraction Performance

We used the addition and subtraction scales of the Heidelberg calculation test by Haffner et al. (2005) to measure math performance. Each scale consists of 40 items that become increasingly difficult. Students are instructed to solve as many items as possible in 120 s. Thus, this test also assesses quick mental arithmetic, which is exactly what students were asked to practice in Math Garden. As our goal was to measure changes in student performance, we expected most performance improvement to occur on items with high item-scale correlations. However, due to the conceptual design of the Heidelberg calculation test, including all items might systematically underestimate the predictive power of the change diagnostics because the first items may cause ceiling effects, and a speed component may affect the last items, which could result in these items then having low item-scale correlations. Therefore, we decided to make an item selection, and we only considered items that had an item-scale correlation higher than .50 (Item 14 to 24 for the addition scale and Item 15 to 25 for the subtraction scale) for performance

scores. We found good retest reliability values for the addition ( $r_{tt} = .52$ ) and subtraction scales ( $r_{tt} = .62$ ). To ensure transparency, we additionally ran our analyses with items that had an item-scale correlation higher than .30 and with all items that the students completed.

### 5.2.3.2 Math Self-Concept

Math self-concept was measured with a scale adapted by Köller et al. (2000). The scale consists of five items (e.g., “I would prefer to do math if the subject was not that difficult.”), which have to be answered using a four-point Likert scale (1 = does not apply at all to 4 = fully applies). All items were reverse coded so that high scores represent a high math self-concept. Cronbach’s alpha was .83 in the pretest and .84 in the posttest. Further, the pre- and posttest scores correlated significantly ( $r_{tt} = .64$ ).

### 5.2.3.3 Math Anxiety

Math anxiety was measured with a subscale from the Math Anxiety Rating Scale for Fourth to Sixth Grades by Roick et al. (2013). The scale consists of five items (e.g., “How nervous do you feel when a math test is written?”), which have to be answered using a five-point Likert scale (1 = not nervous at all to 5 = very nervous). Cronbach’s alpha was .90 in the pretest and .92 in the posttest. The pre- and posttest scores here were also significantly correlated ( $r_{tt} = .37$ ).

### 5.2.3.4 Practice Behavior with Math Garden

Math Garden is an adaptively working web-based learning program. The adaptive algorithm is based on the Elo (1978) rating system, which allows for on-the-fly estimation of item difficulty and person ability parameters (for further information, see Klinkenberg et al. 2011). After solving a task correctly, students receive positive feedback and are presented with a new task with a similar or slightly higher difficulty level. After solving a task incorrectly, students receive corrective feedback and are presented with a new task with a similar or slightly lower difficulty level. Students cannot repeat previous tasks.

Math Garden automatically tracks the number of tasks completed per domain per week (log files). Thus, for each student, we estimated six practice frequency scores: We calculated the number of tasks completed in the domains of addition and subtraction, respectively. We calculated the number of weeks for which students practiced addition tasks and subtraction tasks, respectively. We calculated an overall practice frequency score by calculating the total number of math-related tasks completed in Math Garden, which included tasks on addition, subtraction, multiplication, division, mix (mixture of basic arithmetic operations), counting (of objects), series (logical completion of a given number series), numerals (tasks in which numbers must be combined to obtain a desired target number), and tables (multiplication tables tasks

from 1 to 10), and we calculated the total number of practiced weeks for all math-related tasks for each student. The descriptives for these six different scores are reported in Table 5.2.

**Table 5.2**

*Descriptives of Practice Behavior in Math Garden for the Experimental Condition*

	<i>M</i>	<i>SD</i>	Range
Addition tasks practiced	177.81	268.08	0–2412
Subtraction tasks practiced	54.63	120.78	0–1369
Overall tasks practiced	1090.91	1366.94	11–8406
Practiced weeks of addition	2.74	2.12	0–10
Practiced weeks of subtraction	1.72	1.67	0–12
Overall weeks practiced	4.62	2.96	1–13

*Note.* Experimental condition only.  $n = 200$ .

We had a maximum observation period of 22 calendar weeks (Calendar Week 43 in 2019 to Calendar Week 12 in 2020). The observation period varied for two reasons: First, single classes were tested (i.e., pretest and questionnaire) over a period of three weeks at the beginning of the study (for logistical reasons), and second, three teachers did not provide the login data to their classes at the start of our study. On average, the program was available to the students for 20.7 weeks ( $SD = 1.61$ ,  $min = 17$ ,  $max = 22$ ).

#### 5.2.3.5 Covariates

Students' gender was coded as 0 = male and 1 = female. A migration background was assessed with the following item: "Are other languages spoken at home besides German?" with 0 = no and 1 = yes as the answer options (Happ et al., 2021). The speed scale of the Heidelberg calculation test (Haffner et al., 2005) was used to control for how fast students were at tapping on tablets.

#### 5.2.4 Data Analysis

To answer our research questions, we ran a total of 12 multiple regression analyses in *Mplus* (Version 8.5; Muthén & Muthén, 2008-2020). Addressing the question of how the provision of Math Garden affects addition and subtraction performance, math self-concept, and math anxiety, the predictor was the group assignment (0 = wait-list control group, 1 = experimental group, four multiple regression models). This dummy-coded variable makes it possible to statistically test mean differences between the different groups (Cohen et al., 2003). The dependent variables were the posttest scores when controlling for pretest scores and the other

covariates. We additionally calculated effect sizes (Cohen's  $d_z$ ) to estimate the size of the mean differences.

To investigate the degree to which practice behavior with Math Garden affected our outcomes, we only analyzed the experimental group, as the wait-list control group did not generate any practice data in Math Garden. The dependent variables were the posttest scores when controlling for pretest scores and other covariates, but the predictors were either the number of practiced tasks (Research Question 2, four multiple regression models) or the number of practiced weeks (Research Question 3, four multiple regression models). However, due to our two distinct subdomain performance measures (i.e., addition and subtraction), we made further differentiations: We used the number of addition tasks practiced/practiced weeks of addition to predict addition performance and the number of subtraction tasks practiced/practiced weeks of subtraction to predict subtraction performance. As we assumed that the effects on math self-concept and math anxiety would be less related to the practicing of specific tasks (e.g., addition or subtraction) and more related to feedback mechanisms such as the adaptive algorithm, we used the number of all tasks practiced/all weeks practiced in Math Garden to predict math self-concept and math anxiety, respectively.

As we made an item selection (item-scale correlation  $> .50$ ) for the performance measures, we additionally ran robustness analyses for all of the models that included performance measures. For this, we calculated the effects on performance again with items that had item-scale correlations higher than .30 (Item 9 to 28 for the addition scale and Item 11 to 30 for the subtraction scale) and with all items answered by the students. These results are provided in the online material (see Tables S5.1–S5.4).

Because students were nested within classes, we took the two-level nature of the data into account by using the `TYPE = COMPLEX` command. The `COMPLEX` command adjusts standard errors for nonindependence within classes. Therefore, nonindependence within classes was accounted for but not explicitly modeled. Further, we used the robust maximum likelihood estimator to obtain standard errors that were robust to nonnormality (e.g., Bandalos, 2014; Maydeu-Olivares, 2017). To ease the interpretation, all continuous independent variables were  $z$ -standardized ( $M = 0$ ,  $SD = 1$ ).

In many empirical studies, missing data are a potential methodological problem. We found missing data on all variables. Regarding the self-reported and test measures, this was because some students in our final analytic sample were absent during at least one of both testing sessions ( $N = 156$ ), for different reasons. On the one hand, dropouts resulted in the following missing data pattern: Four students participated in neither the pretest nor the posttest,

but they provided Math Garden data, 10 students did not participate in the pretest but did participate in the posttest, and 26 students participated in the pretest but did not participate in the posttest. On the other hand, the posttest coincided with the COVID-19 lockdown in March 2020 in eight classes, which led to further missing data on all posttest scores for 66 (33%) students in the experimental condition and for 50 (29%) students in the wait-list control condition, although we tried to collect information on the self-reported measures with an additional online survey. Statistical comparisons between the group of students with missing data and the group of students without missing data are provided in the online material (see Table S5.5). The results show that the two groups did not differ regarding their self-reported measures and performance test, but the group of students with missing data had lower practice behavior scores, except for the addition domain. Moreover, three students participated in at least one testing session and received login data, but it was not possible to match their responses from the testing session to their practice data, which then resulted in missing practice data. Overall, the percentage of missing data varied from 0% on group assignment to 41.6% on math self-concept ( $T_2$ ), with an average missing rate of 11.6%. To handle the missing data, we used the full information maximum likelihood approach, which means that missing values were not imputed or filled in, but model parameters and standard errors were directly estimated using all available raw data while no data points were excluded (Enders, 2001; 2022). Thus, we avoided listwise deletion of students with missing data on single measurement occasions.

### 5.3 Results

#### 5.3.1 Descriptives and Preliminary Analyses

The means and standard deviations for the self-report and performance measures are reported in Table 5.1.  $t$ -tests revealed that pretest scores and the measured covariates did not significantly differ between the two conditions, which suggests that the randomization on the class level worked well. Regarding our relevant study variables, we noticed that, on average, math self-concept scores descriptively dropped from pre- to posttest in the wait-list control condition, but we found an increase in the experimental condition. The mean values for the performance measures descriptively increased in both conditions, and anxiety scores slightly decreased on average in both conditions from pre- to posttest.

Table 5.2 displays the descriptives of practice behavior in Math Garden. The high variance of practiced tasks over the 22-week period is noticeable. Overall, students practiced, on average, for 4.62 weeks ( $SD = 2.96$ ) with Math Garden, with a minimum of one week and a

maximum of 13 weeks. Tables 5.3 and 5.4 show the correlations between the independent variables and the dependent variables for all students and for the experimental group only. Regarding the possible effects of Math Garden on our outcome variables, we found significant correlations for group assignment with math self-concept at T<sub>2</sub> ( $r = .12, p = .031$ ; see Table 5.3).

**Table 5.3***Correlation Matrix of the Study Variables*

	1	2	3	4	5	6	7	8	9	10	11
1. Math Learning Program <sup>a</sup>											
2. Math addition performance T <sub>1</sub>	-.05										
3. Math subtraction performance T <sub>1</sub>	-.07	<b>.69</b>									
4. Math self-concept T <sub>1</sub>	.03	<b>.20</b>	<b>.32</b>								
5. Math anxiety T <sub>1</sub>	.07	<b>-.18</b>	<b>-.26</b>	<b>-.45</b>							
6. Gender <sup>b</sup>	.03	-.00	-.08	<b>-.27</b>	<b>.22</b>						
7. Migration background T <sub>1</sub> <sup>c</sup>	.10	-.04	-.04	-.04	.01	.06					
8. Tablet typing speed	-.10	<b>.51</b>	<b>.36</b>	<b>.21</b>	-.06	.02	-.05				
9. Math addition performance T <sub>2</sub>	-.03	<b>.52</b>	<b>.57</b>	<b>.29</b>	<b>-.27</b>	-.07	-.01	<b>.31</b>			
10. Math subtraction performance T <sub>2</sub>	-.05	<b>.46</b>	<b>.62</b>	<b>.25</b>	<b>-.23</b>	<b>-.14</b>	-.01	<b>.21</b>	<b>.67</b>		
11. Math self-concept T <sub>2</sub>	<b>.12</b>	<b>.22</b>	<b>.34</b>	<b>.64</b>	<b>-.36</b>	<b>-.25</b>	-.01	<b>.14</b>	<b>.36</b>	<b>.35</b>	
12. Math anxiety T <sub>2</sub>	.05	-.11	<b>-.21</b>	<b>-.32</b>	<b>.37</b>	<b>.27</b>	<b>.16</b>	<b>.03</b>	<b>-.19</b>	<b>-.19</b>	<b>-.45</b>

*Note.* Correlations refer to all students,  $N = 370$ . Significant correlations are printed in bold ( $p \leq .05$ ). Reference

categories: <sup>a</sup> wait-list control condition. <sup>b</sup> male. <sup>c</sup> no other languages spoken at home besides German.

**Table 5.4***Correlation Matrix for Practice Behavior and Outcomes*

	1	2	3	4	5	6	7	8	9
1. Addition tasks practiced									
2. Subtraction tasks practiced	<b>.36</b>								
3. Overall tasks practiced	<b>.69</b>	<b>.40</b>							
4. Practiced weeks of addition	<b>.71</b>	<b>.49</b>	<b>.63</b>						
5. Practiced weeks of subtraction	<b>.52</b>	<b>.77</b>	<b>.52</b>	<b>.77</b>					
6. Overall weeks practiced	<b>.48</b>	<b>.46</b>	<b>.72</b>	<b>.76</b>	<b>.65</b>				
7. Math addition performance T <sub>2</sub>	-.12	-.00	-.01	.01	-.06	.09			
8. Math subtraction performance T <sub>2</sub>	-.07	.02	-.03	.03	-.00	.07	<b>.71</b>		
9. Math self-concept T <sub>2</sub>	.02	.05	-.01	-.05	-.01	-.02	<b>.20</b>	<b>.29</b>	
10. Math anxiety T <sub>2</sub>	-.07	-.07	-.01	-.01	.03	.04	-.08	-.14	<b>-.52</b>

*Note.* Correlations refer to the experimental condition only,  $n = 200$ . Significant correlations are printed in bold ( $p \leq .05$ ).



### 5.3.2 Effects of Providing Math Garden

Our first research question concerned differences between pre- and posttest scores concerning addition and subtraction performance, math self-concept, and math anxiety between the wait-list control group and the experimental group. After controlling for gender, a possible migration background, and tablet typing speed, we found a statistically significant effect of providing Math Garden on students' math self-concept (see Table 5.5). As expected (H1.3), providing students with Math Garden fostered their math self-concept in contrast to the wait-list control condition ( $\beta = .12$ ,  $p = .002$ ;  $d_z = 0.26$ ). According to Cohen (1988), this effect size can be described as small. Concerning our Hypotheses H1.1, H1.2, and H1.4, we did not find any significant effects.

### 5.3.3 Effects of Practiced Tasks

To answer our second research question, we examined how the number of tasks practiced in Math Garden affected addition and subtraction performance, math self-concept, and math anxiety in the experimental group (see Table 5.6). After controlling for gender, a possible migration background, and tablet typing speed, we found a statistically significant effect of the number of subtraction tasks practiced on students' subtraction performance. As expected (H2.2), the more subtraction tasks students practiced, the higher their subtraction performance was ( $\beta = .10$ ,  $p = .008$ ). Regarding Hypotheses H2.1, H2.3, and H2.4, we did not find any statistically significant effects.

### 5.3.4 Effects of Practiced Weeks

Addressing our third research question, we examined how the number of practiced weeks in Math Garden predicted addition and subtraction performance, math self-concept, and math anxiety in the experimental group (see Table 5.6). We did not find any statistically significant effects regarding our hypotheses (H3.1–H3.4) in the original analyses. However, in further analyses (the results of which are provided in the online material, see Table S5.4), we found—as expected—that the more weeks students spent practicing addition in Math Garden, the higher their addition performance in the posttest was ( $\beta = .13$ ,  $p = .015$ ). In these analyses, math addition performance was measured with items that only had item-scale correlations higher than .30. In the original analysis, this effect was marginally significant. We discuss the implications of this result in the following section.

**Table 5.5**  
*Results on Effects of Providing Math Garden*

	Addition			Subtraction			Math		
	Performance			Performance			Self-Concept		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<i>Predictors</i>									
Outcome at T <sub>1</sub>	<b>.49</b>	.10	.000	<b>.61</b>	.06	.000	<b>.61</b>	.05	.000
Math Learning Program <sup>a</sup>	-.00	.07	.960	-.02	.07	.727	<b>.12</b>	.04	.002
<i>Covariates</i>									
Gender <sup>b</sup>	-.07	.05	.255	-.09	.07	.184	-.08	.07	.212
Migration background T <sub>1</sub> <sup>c</sup>	-.01	.06	.907	.02	.06	.744	-.00	.05	.957
Tablet typing speed T <sub>1</sub>	.08	.15	.469	-.02	.07	.813	.04	.08	.633
<i>R</i> <sup>2</sup>	.29			.38			.43		
							.21		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 370$ .

Reference categories: <sup>a</sup> wait-list control condition. <sup>b</sup> male. <sup>c</sup> no other languages spoken at home besides German

**Table 5.6***Results on Effects of Practice Behavior*

	Addition			Subtraction			Math			Math		
	Performance			Performance			Self-Concept			Anxiety		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
PRACTICED TASKS												
Predictors												
Outcome at T <sub>1</sub>	.42	.13	.001	.65	.07	.000	.54	.07	.000	.37	.08	.000
Practiced tasks	.03	.09	.767	.10	.04	.008	-.01	.10	.931	.08	.11	.455
Covariates												
Gender <sup>a</sup>	-.10	.08	.205	-.14	.10	.152	-.20	.11	.070	.21	.13	.114
Migration background T <sub>1</sub> <sup>b</sup>	-.03	.06	.593	.08	.07	.278	-.09	.08	.242	.25	.11	.027
Tablet typing speed T <sub>1</sub>	.23	.15	.128	.03	.09	.746	.00	.12	.990	.01	.07	.865
R <sup>2</sup>	.31			.48			.41			.25		
PRACTICED WEEKS												
Predictors												
Outcome at T <sub>1</sub>	.42	.12	.001	.63	.08	.000	.54	.07	.000	.37	.07	.000
Practiced weeks	.11	.06	.081	.05	.05	.363	-.01	.06	.815	.11	.07	.113
Covariates												
Gender <sup>a</sup>	-.11	.07	.123	-.15	.09	.107	-.20	.11	.068	.21	.13	.105
Migration background T <sub>1</sub> <sup>b</sup>	-.03	.06	.589	-.08	.08	.316	-.09	.08	.282	.25	.12	.038
Tablet typing speed T <sub>1</sub>	.24	.16	.118	.04	.09	.635	.00	.13	.988	.01	.07	.935
R <sup>2</sup>	.33			.47			.41			.26		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 200$ . Reference categories: <sup>a</sup> male. <sup>b</sup> no other languages spoken at home besides German.

## 5.4 Discussion

By addressing new approaches that can be taken to continue investigating math learning programs, the aim of this study was to explore the effects of the arithmetic learning program Math Garden on distinct subdomain performance measures and on affective-motivational outcomes (i.e., math self-concept and math anxiety), while also taking practice behavior in Math Garden into account. The results supported only some of our hypotheses. We were able to show that providing Math Garden fostered students' math self-concept, in contrast to a wait-

list control condition, and we also showed that the more subtraction tasks students practiced in Math Garden, the higher their subtraction performance in the posttest was. While the items selected for analysis did not show an effect of practiced weeks of addition on students' addition performance in the posttest, an extension to include items with item-scale correlations higher than .30 indicated that the number of practiced weeks of addition was related to students' addition performance. Our other hypotheses were not confirmed.

### **5.4.1 Effects of Math Garden on Math Performance**

We did not find any effects of the mere provision of Math Garden on addition or subtraction performance. Given the ongoing debate about the effectiveness of math learning programs (Byun & Joung, 2018; Pellegrini et al., 2021), this lack of performance effects is not entirely unexpected and again highlights the problem of heterogeneous research findings in this area (Tokac et al., 2019). However, by only investigating the effects of providing the program, many studies are limited at this point, and one can only speculate about why missing effects occur. One possible explanation for the missing results might be that Math Garden does not help students improve their addition and subtraction skills. However, as we would expect Math Garden to have only small effects on students' learning because the program does not intend to teach students new content but rather to repeat what has already been learned, these expected small effects might be especially difficult to reveal if students have not practiced much with the program.

Thus, an alternative possible explanation would be that students did not practice enough with the program to benefit from it, but to make any further suggestions about this, practice behavior needs to be taken into account. In our study, we considered two different types of practice behavior, namely, the number of practiced tasks and the number of practiced weeks. Whereas, for this study, the number of practiced tasks was primarily a measure of quantity, the number of practiced weeks might better reflect the quality of the practice, that is, how regularly students practiced with the program. We found an expected effect of the number of subtraction tasks practiced on subtraction performance, but we did not find an effect of practicing addition tasks on addition performance. In terms of the number of practiced weeks, we found an inverse pattern, that is, we did not find an effect of the number of practiced weeks of subtraction on subtraction performance, but we found an expected positive effect of the number of practiced weeks of addition on addition performance. However, this second effect only occurred when performance was measured with items that had item-scale correlations higher than .30. In contrast, the main analysis was conducted with items that had item-scale correlations higher

than .50. Thus, this result needs to be interpreted with caution and can only be generalized to a very limited extent. The effects of subtraction tasks practiced on subtraction performance seem to be more robust because we found this effect in all of our analyses.

In general, according to our G\*Power analyses (Faul et al., 2007; 2009), the missing effects of practice behavior might be explained by the reduction of sample size due to the COVID-19 school closures and, thus, a loss of power. Hence, for future research, we recommend that this study is replicated in a bigger sample. Further, inactivity and dropout rates are serious issues when investigating the effects of learning programs, and they might explain missing effects as well (Bacca-Acosta & Avila-Garzon, 2021; Spitzer & Musslick, 2021). Students might have shown high engagement with Math Garden right after receiving it, which, in turn, might have resulted in short-term performance increases, but they might then have become inactive in the long-term so that effects were no longer detectable, especially at the end of a longer intervention such as ours. Thus, on the one hand, we would like to emphasize the importance of investigating how to put students in a mindset where they continuously and systematically engage with such learning programs (Spitzer & Musslick, 2021). On the other hand, future studies could assess performance not just once in a posttest but could instead measure it in small steps, that is, several times within the intervention period, in order to get more insight into what really happens during the use of such a learning program. Still, the finding that practice behavior had no effects on addition performance but it did on subtraction performance might also be due to subtraction being more difficult than addition (Anderson et al., 2022; Kamii et al., 2001). Possibly, this is why practicing with Math Garden was particularly helpful for the subtraction domain.

However, surprisingly, we found an effect only for the number of subtraction tasks practiced and not for the number of practiced weeks of subtraction. Given the assumptions about distributed learning behavior, which means that practice is most effective when practice sessions are spaced out over time with breaks in between rather than when tasks are repeated in immediate succession (i.e., massed practicing; Carpenter et al., 2012; Gerbier & Toppino, 2015; Toppino & Gerbier, 2014), it is reasonable to assume that the practiced weeks are a relevant indicator. Missing effects might be because most students practiced only for a small number of weeks and, thus, only little—and probably not meaningful—variance was created within the variables for the practiced weeks. Therefore, future research might systematically vary the practiced weeks, for instance, by reminding a subpopulation of students every day to practice for five minutes. Further, future research might combine the practiced weeks with the practiced tasks measure because, with the current approach, for instance, a student who

practiced 100 tasks in four weeks has the same score as a student who practiced 10 tasks in four weeks. Thus, in terms of challenging the hypotheses on massed and distributed practice behavior, future studies might investigate a moderation between the practiced weeks and the practiced tasks. As showing such an interaction effect would require even higher statistical power (e.g., through a sufficiently large sample size; Aiken & West, 1991), we concentrated on considering the practiced tasks and practiced weeks separately as direct effects. Additionally, our results indicate that both measures were highly correlated. Hence, suppressor effects might become a problem for future studies investigating such an approach. Thus, we recommend ensuring not only that there are students who practice few tasks in few weeks and many tasks in many weeks but also that there are students who practice few tasks in many weeks and many tasks in few weeks. Still, as no previous study ever took the measure of practiced weeks into account, our approach provides initial, valuable information on how to employ a new practice behavior measure regarding math learning programs.

### **5.4.2 Effects of Math Garden on Math Self-Concept**

Our results suggest that, as expected, providing students with Math Garden fostered their math self-concept in contrast to a wait-list control condition. Considering the fact that math self-concept is one of the most important predictors of later performance (Eccles & Wigfield, 2020; Guay et al., 2003; Marsh & Martin, 2011), this is a very promising result. Interestingly, the provision of Math Garden in our study seemed to have a buffering effect. While we descriptively observed that math self-concept in the wait-list control condition declined over the intervention period, which is in line with previous research that suggested an overall decrease in self-concept over a school year (Eshel & Klein, 1981; Fraine et al., 2007), the provision of Math Garden appeared to counteract this decline and also led to a significant increase in math self-concept. Thus, students seemed to benefit from Math Garden regarding their math self-concept, which is why we can recommend implementing the program in classrooms. Further, this result is also in line with earlier research: In a Dutch sample, Jansen et al. (2013) showed a small but positive effect of providing Math Garden on perceived math competence, which is a construct that is closely related to math self-concept (Lee, 2009). However, they only had an intervention duration of 11.1 weeks on average, and their effect can thus be assumed to be modest because of the short intervention period (Jansen et al., 2013). By replicating this result in a German sample with an intervention duration almost twice as long, our finding supports its robustness.

Further, keeping in mind that the students in our sample all came from non-academic-track schools, it is reasonable to assume that they were mostly lower achievers. Although research results have already shown that math learning program interventions can be especially effective for low-achieving students regarding performance (Hassler Hallstedt et al., 2018; Ran et al., 2021), effects on affective-motivational outcomes have not yet been considered. However, as low-achieving students, in particular, might have unfavorable motivational prerequisites after continuous experiences of failure, future research should also focus on investigating such effects in this subpopulation. The results of our study might be the first indication that such interventions have the potential to foster low-achieving students' math self-concept, which is promising and thus needs further investigation.

Contrary to our expectations, we did not find an effect of practice behavior on math self-concept. This result is in line with research by Jansen et al. (2013), who were not able to show any effects on the construct of perceived math competence, which is related to math self-concept when considering the practiced tasks in Math Garden. Given the positive effects of providing Math Garden on students' math self-concept, one might conclude that the amount of feedback students receive when practicing with the program does not influence the extent to which the program fosters their math self-concept. An explanation for this missing effect of practice behavior might be found in the study results of Bernacki et al. (2015), who investigated the effect of feedback on the motivational construct of self-efficacy. They showed that feedback on the correctness of tasks affected self-efficacy, especially at the beginning of the training session. However, over time, the previous self-efficacy became more and more predictive of further self-efficacy, while the number of correct answers, and thus positive feedback, became less important. The authors postulated that positive feedback has a positive effect, especially at the start of a learning activity, but that self-efficacy then stabilizes (Bernacki et al. 2015). This might be the same with math self-concept. To test such an assumption for practice behavior with Math Garden, future research that assesses math self-concept multiple times during the intervention period is needed.

### **5.4.3 Effects of Math Garden on Math Anxiety**

We did not find any effects of providing Math Garden or of practice behavior with Math Garden on math anxiety. These missing findings are in line with research from Jansen et al. (2013), who were also not able to show any effects on math anxiety. Regarding these missing effects, research by Hilz et al. (2023) showed that math-anxious students generally practice less with math learning programs. Thus, math anxious students might not have benefited from the

adaptive algorithm as they practiced too little to notice that the program adjusted tasks to their ability level. In turn, they might not have perceived tasks in Math Garden to be controllable, which is why math anxiety did not decrease. Furthermore, many authors have pointed out that math anxiety is a phenomenon that begins at a young age and is, therefore, deeply rooted in students (Ashcraft & Krause, 2007; Bekdemir, 2010; Ramirez et al., 2013). Therefore, although students' experiences with Math Garden in the present investigation might have been positive due to the adaptive algorithm and, therefore, a sufficient amount of positive feedback might have been received, these experiences may not outweigh negative math experiences in the past (Jansen et al. 2013).

Additionally, our math anxiety items mainly focused on math test anxiety, which is only one dimension of math anxiety. Research has already shown that math anxiety has other facets and components, such as an affective and a cognitive component (Wigfield & Meece 1988), or a trait and a state component (Orbach et al., 2019). Again, to get a more detailed view of the underlying effects of math learning programs, future investigations might include these different components in their investigations. Future research might also experimentally investigate in more detail whether the manipulation of specific program elements has the potential to reduce math anxiety. For example, the developers of Math Garden recently implemented the option of hiding coins when students solve the tasks. Normally, while a specific task is being presented, coins spilling out under the task symbolize the remaining time to give an answer. The more quickly students give the correct answer, the more remaining coins they receive as a reward. However, these coins might put pressure on math-anxious students in particular because working memory capacities might be even more reduced (Ashcraft & Kirk, 2001; Hunt & Sandhu, 2017). Hence, turning off this visibly expiring time might have a positive effect on students' math anxiety.

### **5.4.4 Limitations**

Even though our study provides new insights into how math learning programs like Math Garden potentially affect students' performance, their math self-concept, and their math anxiety, our analyses have limitations that need to be considered when interpreting the results. As already mentioned, the posttest sample size was reduced due to school closures because of the COVID-19 lockdown. Even though this sample dropout appeared to be unsystematic, as we did not suggest testing dates to schools with any specific strategy, and the two subsamples did not differ in their pretest scores, the consequence was a loss of power according to our G\*Power



analyses (Faul et al., 2007; 2009). Hence, for replicating the robustness of our desirable finding on math self-concept, we recommend replicating our findings in a larger sample.

Due to the fact that most high-achieving students in Germany attend academic-track schools, our sample, which only included non-academic-track schools, consisted of, on average, lower-achieving students. Therefore, the results cannot be generalized to all types of German schools. In turn, as we also did not explicitly consider low achievement, our results can also not be fully generalized to all low-achieving students.

Further, we did not collect information about the environment (i.e., at home or in school) in which students practiced with Math Garden. We assume that using Math Garden voluntarily at home might be more effective regarding performance than using Math Garden during math lessons at school because using Math Garden outside of school means extra learning time. For instance, Bakker et al. (2015) showed that their math learning program intervention was most effective when students practiced at home after being introduced to the program at school. Future research might investigate whether students benefit more from extra practice time at home in contrast to practice time during math lessons.

Moreover, school, teacher, and parental characteristics might have affected our outcome variables. For instance, schools vary in their digital infrastructure and, thus, in the availability of digital devices (Eickelmann et al., 2019; Stanat et al., 2022). Because we randomized students on the class level, the variance between schools should have been controlled for. However, because of the design of our study, we were only able to control for variance differences between classes. Hence, we were unable to consider teacher variables at Level 2 to investigate variance between classes. However, as teachers vary in their competencies and attitudes towards the use of digital media in the classroom (Eickelmann et al., 2019), it is reasonable to assume that such characteristics might have affected our outcomes as well. Further, it would be interesting to know the extent to which teachers used Math Garden during their math instruction. Moreover, we were also not able to take parental characteristics into account. The results of previous research on the quality dimensions of parental homework involvement have already shown that, for instance, high parental control inhibits students' academic functioning and achievement (Guill et al., 2020; Moroni et al., 2015). Further, we did not know what type of digital devices students used to practice with Math Garden at home. As research on the motives for using digital devices has already shown that, for instance, smartphones are often used for entertainment-related and social-interactive purposes, whereas desktop computers offer better possibilities for information and learning-orientated use (Napoli & Obar, 2014), we assume that students might be more distracted when using Math Garden on

smartphones instead of on desktop computers. This might have affected their performance. Future research might take such variables into account.

Finally, this study had two design limitations. First, we only investigated the effects of Math Garden compared to a “business as usual” group (i.e., students receiving regular math instruction) that did not receive any support in the form of additional learning opportunities. Thus, the effects of Math Garden can only be interpreted compared to the effects of not having any additional learning opportunities. Thus, the current study rather provides evidence for the effects of Math Garden than for the effectiveness of Math Garden. For the latter, a comparison with another support strategy or other additional learning opportunities with the same invested learning time would be helpful to examine if Math Garden is more effective in supporting learners than other support strategies. Second, randomization took place on the class level rather than on the individual level. This is standard practice when implementing a treatment in educational research to ensure the acceptance and usefulness of the study (Plewis & Hurry, 1998), but it should, nonetheless, be kept in mind when interpreting the results.

### **5.4.5 Practical Implications**

Although only some of our theoretical assumptions were confirmed, this study has implications for practitioners who plan to implement a math learning program in the classroom. The distinct performance measure approach could help teachers to decide which program to use in the classroom or educational administrations to decide which specific programs to recommend for classroom use because this approach can clarify which program is effective in fostering a particular math ability (Ran et al., 2021). Thus, with the results of our study, we can recommend that teachers use Math Garden as it might foster their students’ subtraction skills more than if no additional learning opportunities are provided. Further investigations might also test how Math Garden affects students’ multiplication and division skills.

However, our study also showed that simply providing the program did not affect students’ performance. Thus, teachers need to actively encourage their students to practice with Math Garden in order for them to benefit regarding their performance. This seems to apply in particular to math-anxious students (Hilz et al., 2023). For instance, teachers might use Math Garden at the end of every math lesson to get their students used to practicing, or they might use it in supervision classes or assign practice as homework. Such strategies should be followed, especially after a vacation period, as students tend to stop practicing during vacations (Hilz et al., 2023; Hofman et al., 2018). Moreover, as Math Garden has the potential to foster students’ math self-concept, the use of Math Garden in the classroom might motivate students to further

deal with math and math-related topics, which, in turn, might have long-term effects on future performance (Eccles & Wigfield, 2020; Guay et al., 2003; Marsh & Martin, 2011).

### 5.5 Conclusions

Even though research on math learning programs has taken off in recent years (Higgins et al., 2019; Hillmayr et al., 2020; Pellegrini et al., 2021; Sailer & Homner, 2020), the results of their effectiveness are heterogeneous, and the programs seem to largely fail to meet the expectation that they can fundamentally revolutionize the school system (Pellegrini et al., 2021). However, instead of concluding that they are ineffective and potentially rejecting their use, this study offers new ways in which future research might address math learning program evaluations and thus adds a new perspective to the current state of research in this field. Our results suggest that the learning program Math Garden might foster students' subtraction skills more than when an additional learning program is not used, but only when practice behavior is considered. Further, we were able to show that providing Math Garden fosters students' math self-concept. For future investigations, we would like to encourage researchers to evaluate (math) learning programs on a more detailed level to uncover potentially hidden effects. Thus, we recommend that researchers take the following points into account:

- Focus on measuring distinct subdomains of performance: This can help practitioners, in particular, to make decisions about the target group for whom the program might be most beneficial.
- Take affective-motivational variables into account: Even if a program has no effect on performance shortly after the intervention, performance might increase in the long term if affective-motivational variables, which are predictors of performance, are affected by the intervention. Again, it is also important to investigate single dimensions of these variables in more detail, for instance, to differentiate between the cognitive and the affective component of math anxiety.
- Consider practice behavior with log and trace data: This can provide a deeper insight into the optimal amount of practice with (math) learning programs and which types of students' behavior might benefit the most from the implementation (e.g., distributed practicing).

## 5.6 References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting Interactions*. SAGE.
- Anderson, S., Sommerhoff, D., Schurig, M., Ufer, S., & Gebhardt, M. (2022). Developing learning progress monitoring tests using difficulty-generating item characteristics. An example for basic arithmetic operations in primary schools. *Journal for Educational Research Online*, 14(1), 122–145. <https://doi.org/10.25656/01:24875>
- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology*, 102(1), 115–134. <https://doi.org/10.1037/a0016838>
- Arens, A. K., Frenzel, A. C., & Goetz, T. (2022). Self-concept and self-efficacy in math: Longitudinal interrelations and reciprocal linkages with achievement. *The Journal of Experimental Education*, 90(3), 615–633. <https://doi.org/10.1080/00220973.2020.1786347>
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology*, 130(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, 14(2), 243–248. <https://doi.org/10.3758/BF03194059>
- Bacca-Acosta, J., & Avila-Garzon, C. (2021). Student engagement with mobile-based assessment systems: A survival analysis. *Journal of Computer Assisted Learning*, 37(1), 158–171. <https://doi.org/10.1111/jcal.12475>
- Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology*, 43(6), 993–1003. <https://doi.org/10.1111/j.1467-8535.2011.01269.x>
- Bailey, D. H., Hoard, M. K., Nugent, L., & Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. *Journal of Experimental Child Psychology*, 113(3), 447–455. <https://doi.org/10.1016/j.jecp.2012.06.004>

- Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). *Early predictors of middle school fraction knowledge. Developmental Science, 17*(5), 775–785.  
<https://doi.org/10.1111/desc.12155>
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics* (pp. 61–75). Springer New York.  
[https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)
- Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2015). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability. *Contemporary Educational Psychology, 40*, 55–71.  
<https://doi.org/10.1016/j.cedpsych.2014.09.001>
- Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(1), 102–116.  
<https://doi.org/10.1080/10705511.2014.859510>
- Barbieri, C. A., Young, L. K., Newton, K. J., & Booth, J. L. (2021). Predicting middle school profiles of algebra performance using fraction knowledge. *Child Development, 92*(5), 1984–2005. <https://doi.org/10.1111/cdev.13568>
- Barzagar Nazari, K., & Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology, 33*(2), 288–298. <https://doi.org/10.1002/acp.3485>
- Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. *Educational Studies in Mathematics, 75*(3), 311–328. <https://doi.org/10.1007/s10649-010-9260-7>
- Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: variability and relations to behavior, performance, and learning. *Metacognition and Learning, 10*(1), 99–117. <https://doi.org/10.1007/s11409-014-9127-x>
- Beserra, V., Nussbaum, M., Zeni, R., Rodriguez, W., & Wurman, G. (2014). Practising arithmetic using educational video games with an interpersonal computer. *Journal of Educational Technology & Society, 17*(3), 343–358.
- Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics, 118*(3-4), 113–126.  
<https://doi.org/10.1111/ssm.12271>

- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369–378. <https://doi.org/10.1007/s10648-012-9205-z>
- Caviola, S., Toffalini, E., Giofrè, D., Ruiz, J. M., Szűcs, D., & Mammarella, I. C. (2022). Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants. *Educational Psychology Review*, 34(1), 363–399. <https://doi.org/10.1007/s10648-021-09618-5>
- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International*, 52(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Chinn, S. (2009). Mathematics anxiety in secondary students in England. *Dyslexia*, 15(1), 61–68. <https://doi.org/10.1002/dys.381>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge Taylor & Francis Group.
- Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, 83(1), 17–27. <https://doi.org/10.1037/0022-0663.83.1.17>
- Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, 32(1), 132–150. <https://doi.org/10.1016/j.cedpsych.2006.10.005>
- Devine, A., Hill, F., Carey, E., & Szűcs, D. (2018). Cognitive and emotional math problems largely dissociate: Prevalence of developmental dyscalculia and mathematics anxiety. *Journal of Educational Psychology*, 110(3), 431–444. <https://doi.org/10.1037/edu0000222>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on

- motivation. *Contemporary Educational Psychology*, 61, 101859.  
<https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Eds.). (2019). *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking [ICILS 2018 #Germany: Computer and information-related competencies of students in the second international comparison and competencies in computational thinking]*. Waxmann.
- Elo, A. (1978). *The rating of chessplayers, past and present*. Arco Publishers.
- Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5), 713–740. <https://doi.org/10.1177/00131640121971482>
- Enders, C. K. (2022). *Applied missing data analysis*. The Guilford Press.
- Eshel, Y., & Klein, Z. (1981). Development of academic self-concept of lower-class and middle-class primary school children. *Journal of Educational Psychology*, 73(2), 287–293. <https://doi.org/10.1037/0022-0663.73.2.287>
- Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2022). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning*, 38(1), 304–325. <https://doi.org/10.1111/jcal.12618>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, 4(3), 49–59. <https://doi.org/10.1016/j.tine.2015.01.001>
- Fraine, B. de, Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, 32(1), 132–150. <https://doi.org/10.1016/j.cedpsych.2006.10.005>

- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95(1), 124–136. <https://doi.org/10.1037/0022-0663.95.1.124>
- Guill, K., Bahr, I., & Ömerogullari, M. (2019). Empirische Arbeit: Qualität der Hausaufgabenhilfe in Elternhaus und Nachhilfeunterricht als Prädiktoren des hausaufgabenbezogenen Arbeitsverhaltens [Empirical work: Quality of homework help at home and tutoring as predictors of homework-related work behavior]. *Psychologie in Erziehung und Unterricht*. Advance online publication. <https://doi.org/10.2378/peu2020.art08d>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1–4: Heidelberger Rechentest – Erfassung mathematischer Basiskompetenzen im Grundschulalter [HRT 1–4: Heidelberg Arithmetic Test - Assessment of basic mathematical skills at primary school age]*. Hogrefe.
- Haleva, L., HersHKovitz, A., & Tabach, M. (2021). Students’ activity in an online learning environment for mathematics: The role of thinking levels. *Journal of Educational Computing Research*, 59(4), 686–712. <https://doi.org/10.1177/0735633120972057>
- Hansen, N., Jordan, N. C., & Rodrigues, J. (2017). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology*, 50, 45–59. <https://doi.org/10.1016/j.cedpsych.2015.11.002>
- Happ, R., Nagel, M.-T., Zlatkin-Troitschanskaia, O., & Schmidt, S. (2021). How migration background affects master degree students’ knowledge of business and economics. *Studies in Higher Education*, 46(3), 457–472. <https://doi.org/10.1080/03075079.2019.1640670>
- Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology*, 110(8), 1127–1148. <https://doi.org/10.1037/edu0000264>
- Higgins, K., Huscroft-D’Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research*, 57(2), 283–319. <https://doi.org/10.1177/0735633117748416>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A



- context-specific meta-analysis. *Computers & Education*, 153, 103897.  
<https://doi.org/10.1016/j.compedu.2020.103897>
- Hilz, A., Guill, K., Roloff, J., Aldrup, K., & Köller, O. (2023). The relationship between individual characteristics and practice behaviour within an adaptive arithmetic learning program. *Journal of Computer Assisted Learning*, 39(3), 970–983.  
<https://doi.org/10.1111/jcal.12780>
- Hofman, A. D., Kievit, R., Stevenson, C., Molenaar, D., Visser, I., & Van der Maas, H. L. J. (2018). The dynamics of the development of mathematics skills: A comparison of theories of developing intelligence. <https://doi.org/10.31219/osf.io/xa2ft>
- Huang, Y.-M., Huang, S.-H., & Wu, T.-T. (2014). Embedding diagnostic mechanisms in a digital game for learning mathematics. *Educational Technology Research and Development*, 62(2), 187–207. <https://doi.org/10.1007/s11423-013-9315-4>
- Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1(2-3), 151–166.  
<https://doi.org/10.1007/s40692-014-0008-8>
- Hunt, T. E., & Sandhu, K. K. (2017). Endogenous and exogenous time pressure: Interactions with mathematics anxiety in explaining arithmetic performance. *International Journal of Educational Research*, 82, 91–98. <https://doi.org/10.1016/j.ijer.2017.01.005>
- Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, 24, 190–197. <https://doi.org/10.1016/j.lindif.2012.12.014>
- Jiang, S., Simpkins, S. D., & Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Developmental Psychology*, 56(11), 2137–2151. <https://doi.org/10.1037/dev0001110>
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, 116(1), 45–58.  
<https://doi.org/10.1016/j.jecp.2013.02.001>
- Kamii, C., Lewis, B. A., & Kirkland, L. D. (2001). Fluency in subtraction compared with addition. *The Journal of Mathematical Behavior*, 20(1), 33–42.  
[https://doi.org/10.1016/S0732-3123\(01\)00060-8](https://doi.org/10.1016/S0732-3123(01)00060-8)

- Ke, F. (2008). Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development*, 56(5-6), 539–556. <https://doi.org/10.1007/s11423-008-9086-5>
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>
- Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse [Course choices of girls and boys in mathematics: The role of subject-specific self-concept and interest]. *Zeitschrift Für Pädagogische Psychologie*, 14(1), 26–37. <https://doi.org/10.1024//1010-0652.14.1.26>
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education*, 22(2), 164–184. <https://doi.org/10.1080/08957340902754635>
- Louw, J., Muller, J., & Tredoux, C. (2008). Time-on-task, technology and mathematics achievement. *Evaluation and Program Planning*, 31(1), 41–50. <https://doi.org/10.1016/j.evalprogplan.2007.11.001>
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520. <https://doi.org/10.2307/749772>
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, 2(2), 99–106. <https://doi.org/10.1111/j.1750-8606.2008.00048.x>
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82(4), 623–636. <https://doi.org/10.1037/0022-0663.82.4.623>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>

- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and english constructs. *American Educational Research Journal*, 35(4), 705–738.  
<https://doi.org/10.3102/00028312035004705>
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 383–394.  
<https://doi.org/10.1080/10705511.2016.1269606>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43–71). Cambridge University Press.
- Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9047-2>
- Moroni, S., Dumont, H., Trautwein, U., Niggli, A., & Baeriswyl, F. (2015). The need to distinguish between quantity and quality in research on parental involvement: The example of parental help with homework. *The Journal of Educational Research*, 108(5), 417–431. <https://doi.org/10.1080/00220671.2014.901283>
- Muthén, B. O., & Muthén, L. K. (2008-2020). *Mplus computer software (version 8.5)*. Muthén & Muthén.
- Napoli, P. M., & Obar, J. A. (2014). The emerging mobile internet underclass: A critique of mobile internet access. *The Information Society*, 30(5), 323–334.  
<https://doi.org/10.1080/01972243.2014.944726>
- Nattland, A., & Kerres, M. (2009). Computerbasierte Methoden im Unterricht [Computer-based methods in the classroom]. In K.-H. Arnold, U. Sandfuchs, & J. Wiechmann (Eds.), *UTB Schulpädagogik. Handbuch Unterricht* (pp. 317–324). Klinkhardt.
- Orbach, L., Herzog, M., & Fritz, A. (2019). Relation of state- and trait-math anxiety to intelligence, math achievement and learning motivation. *Journal of Numerical Cognition*, 5(3), 371–399. <https://doi.org/10.5964/jnc.v5i3.204>
- Pareto, L., Haake, M., Lindström, P., Sjöden, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation. *Educational Technology Research and Development*, 60(5), 723–751.  
<https://doi.org/10.1007/s11423-012-9246-5>

- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7, 233285842098621. <https://doi.org/10.1177/2332858420986211>
- Plewis, I., & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation*, 4(1), 13–26. <https://doi.org/10.1076/edre.4.1.13.13014>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, 14(2), 187–202. <https://doi.org/10.1080/15248372.2012.664593>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12 classrooms. *Journal of Educational Computing Research*, 59(1), 119–153. <https://doi.org/10.1177/0735633120952063>
- Reinhold, F., Reiss, K., Diedrich, J., Hofer, S., & Heinze, A. (2019). Mathematische Kompetenz in PISA 2018—Aktueller Stand und Entwicklung [Mathematical competence in PISA 2018—current status and development]. In K. Reiss, M. Weis, E. Klieme, & O. Köller (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (pp. 187–210). Waxmann.
- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19(6), 551–554. <https://doi.org/10.1037/h0033456>
- Rodrigo, M. M. T. (2011). Dynamics of student cognitive-affective transitions during a mathematics game. *Simulation & Gaming*, 42(1), 85–99. <https://doi.org/10.1177/1046878110361513>
- Roick, T., Göllitz, D., & Hasselhorn, M. (2013). Affektive Komponenten der Mathematikkompetenz. Die Mathematik-Angst-Ratingskala für vierte bis sechste Klassen [Affective components of math competence. The Mathematics Anxiety Rating Scale for fourth to sixth grades] (MARS 4-6). In M. Hasselhorn, A. Heinze, W. Schneider, & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen* (pp. 205–219). Hogrefe.

- Sailer, M., & Homner, L. (2020). The gamification of learning: a meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology*, 53(2), 149–159. <https://doi.org/10.1016/j.jsp.2014.12.003>
- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540–560. <https://doi.org/10.1111/j.1467-8535.2011.01197.x>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, 23(7), 691–697. <https://doi.org/10.1177/0956797612440101>
- Spitzer, M. W. H., & Musslick, S. (2021). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic. *PloS One*, 16(8), e0255629. <https://doi.org/10.1371/journal.pone.0255629>
- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). IQB-Bildungstrend 2021. *Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich [Competencies in the subjects German and mathematics at the end of year 4 in the third country comparison]*. Waxmann.
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Sung, Y.-T., Yang, J.-M., & Lee, H.-Y. (2017). The effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis. *Review of Educational Research*, 87(4), 768–805. <https://doi.org/10.3102/0034654317704307>

- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *Psychology of Learning and Motivation*, 60, 113–189. <https://doi.org/10.1016/B978-0-12-800090-8.00004-4>
- Van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game. *Educational Studies in Mathematics*, 84(3), 281–307. <https://doi.org/10.1007/s10649-013-9483-5>
- Van Mier, H. I., Schleepen, T. M. J., & Van den Berg, F. C. G. (2018). Gender differences regarding the impact of math anxiety on arithmetic performance in second and fourth graders. *Frontiers in Psychology*, 9, 2690. <https://doi.org/10.3389/fpsyg.2018.02690>
- Vanbecelaere, S., Demedts, F., Reynvoet, B., & Depaepe, F. (2022). The effectiveness of adaptive digital games for learning: Calling for a broader view on assessment. In K. Kiili, K. Antti, F. de Rosa, M. Dindar, M. Kickmeier-Rust, & F. Bellotti (Eds.), *Games and Learning Alliance* (pp. 269–278). Springer International Publishing,
- Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, 83(4), 1137–1144. <https://doi.org/10.1111/j.1467-8624.2012.01781.x>
- Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80(2), 210–216. <https://doi.org/10.1037/0022-0663.80.2.210>

## 5.7 Supplementary Tables

**Table S5.1***Effects of Providing Math Garden on Performance With Items > .30 Item-Scale Correlation*

	Addition performance			Subtraction performance		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.55</b>	.10	.000	<b>.67</b>	.06	.000
Math Learning Program <sup>a</sup>	−.00	.07	.973	−.01	.07	.941
<i>Covariates</i>						
Gender <sup>b</sup>	−.08	.05	.162	−.09	.05	.117
Migration background T <sub>1</sub> <sup>c</sup>	.02	.05	.709	.04	.05	.425
Tablet typing speed T <sub>1</sub>	.05	.09	.631	−.02	.06	.769
<i>R</i> <sup>2</sup>	.33			.46		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 370$ . Reference categories: <sup>a</sup>wait-list control condition. <sup>b</sup>male. <sup>c</sup>no other languages spoken at home besides German.

**Table S5.2***Effects of Providing Math Garden on Performance With all Performance Items*

	Addition performance			Subtraction performance		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.55</b>	.10	.000	<b>.65</b>	.06	.000
Math Learning Program <sup>a</sup>	.03	.07	.686	-.01	.07	.918
<i>Covariates</i>						
Gender <sup>b</sup>	-.07	.05	.148	-.08	.06	.134
Migration background T <sub>1</sub> <sup>c</sup>	-.05	.05	.296	.05	.05	.249
Tablet typing speed T <sub>1</sub>	.02	.08	.788	.01	.06	.912
<i>R</i> <sup>2</sup>	.32			.45		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 370$ . Reference categories: <sup>a</sup>wait-list control condition. <sup>b</sup>male. <sup>c</sup>no other languages spoken at home besides German.



**Table S5.3***Effects of Practice Behavior on Performance With Items > .30 Item-Scale Correlation*

	Addition performance			Subtraction Performance		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<b>PRACTICED TASKS</b>						
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.49</b>	.14	.000	<b>.70</b>	.06	.000
Practiced tasks	.07	.07	.316	<b>.12</b>	.05	.015
<i>Covariates</i>						
Gender <sup>a</sup>	−.11	.06	.056	−.14	.10	.095
Migration background T <sub>1</sub> <sup>b</sup>	−.00	.05	.976	−.04	.06	.515
Tablet typing speed T <sub>1</sub>	.16	.12	.185	−.01	.08	.885
<i>R</i> <sup>2</sup>	.34			.53		
<b>PRACTICED WEEKS</b>						
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.49</b>	.14	.000	<b>.68</b>	.07	.000
Practiced weeks	<b>.13<sup>†</sup></b>	.05	.015	.06	.06	.282
<i>Covariates</i>						
Gender <sup>a</sup>	<b>−.13<sup>†</sup></b>	.05	.014	<b>−.15<sup>†</sup></b>	.07	.045
Migration background T <sub>1</sub> <sup>b</sup>	−.00	.05	.953	−.04	.06	.589
Tablet typing speed T <sub>1</sub>	.18	.12	.154	−.01	.09	.928
<i>R</i> <sup>2</sup>	.36			.52		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 200$ . Reference categories: <sup>a</sup> male. <sup>b</sup> no other languages spoken at home besides German. <sup>†</sup>Coefficient differed in significance level ( $p \leq .05$ ) in contrast to original analysis with item-scale correlation > .5.

**Table S5.4***Effects of Practice Behavior on Performance With all Performance Items*

	Addition performance			Subtraction Performance		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<b>PRACTICED TASKS</b>						
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.47</b>	.14	.001	<b>.70</b>	.07	.000
Practiced tasks	.09	.08	.237	<b>.14</b>	.04	.001
<i>Covariates</i>						
Gender <sup>a</sup>	<b>−.11<sup>†</sup></b>	.05	.042	−.13	.08	.093
Migration background T1 <sup>b</sup>	.03	.05	.507	−.01	.06	.894
Tablet typing speed T <sub>1</sub>	.14	.11	.199	.01	.07	.939
<i>R</i> <sup>2</sup>	.30			.54		
<b>PRACTICED WEEKS</b>						
<i>Predictors</i>						
Outcome at T <sub>1</sub>	<b>.47</b>	.14	.001	<b>.68</b>	.07	.000
Practiced weeks	.11	.06	.070	.09	.05	.087
<i>Covariates</i>						
Gender <sup>a</sup>	<b>−.12<sup>†</sup></b>	.05	.012	<b>−.15<sup>†</sup></b>	.07	.041
Migration background T1 <sup>b</sup>	.03	.06	.568	−.01	.06	.933
Tablet typing speed T <sub>1</sub>	.15	.11	.167	.03	.08	.692
<i>R</i> <sup>2</sup>	.31			.52		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 200$ . Reference categories: <sup>a</sup> male. <sup>b</sup> no other languages spoken at home besides German. <sup>†</sup>Coefficient differed in significance level ( $p \leq .05$ ) in contrast to original analysis with item-scale correlation  $> .5$ .

**Table S5.5***Comparing Study Variables With and Without Missing Values*

	Without Missing Values		With Missing Values	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>ALL STUDENTS</b>				
Math addition performance T <sub>1</sub>	5.76	3.15	5.38	3.11
Math addition performance T <sub>2</sub>	7.88	2.76	8.40	2.84
Math subtraction performance T <sub>1</sub>	5.24	3.74	5.13	3.47
Math subtraction performance T <sub>2</sub>	6.62	3.73	6.90	3.00
Math self-concept T <sub>1</sub>	2.73	0.86	2.65	0.76
Math self-concept T <sub>2</sub>	2.74	0.83	2.36	0.82
Math anxiety T <sub>1</sub>	3.39	1.24	3.32	1.12
Math anxiety T <sub>2</sub>	3.21	1.28	3.02	1.00
Gender <sup>a</sup>	0.48	0.50	0.50	0.50
Migration background T <sub>1</sub> <sup>b</sup>	0.52	0.50	0.44	0.50
Tablet typing speed T <sub>1</sub>	7.97	2.20	7.75	2.59
<b>EXPERIMENTAL CONDITION</b>				
Practiced addition tasks	193.36	246.17	159.70	291.88
Practiced subtraction tasks	<b>69.61</b>	150.24	<b>37.19</b>	69.72
Overall practiced tasks	<b>1280.06</b>	1467.16	<b>870.58</b>	1211.08
Practiced weeks addition	<b>3.02</b>	2.42	<b>2.42</b>	1.66
Practiced weeks subtraction	<b>1.97</b>	2.01	<b>1.43</b>	1.08
Overall practiced weeks	<b>5.58</b>	3.31	<b>3.51</b>	2.00

*Note.* All students:  $n_{\text{without}} = 214$ ,  $n_{\text{with}} = 156$ . Experimental condition:  $n_{\text{without}} = 116$ ,  $n_{\text{with}} = 84$ . Bold print indicates values that were statistically significantly different at  $p \leq .05$ . Reference categories: <sup>a</sup> male. <sup>b</sup> no other languages spoken at home besides German.



# 6

## ADDITIONAL REPLICATION ANALYSIS STUDY 2

### 6.1 Introduction

As already emphasized in Chapter 4, replicating research is extremely important. This especially applies to intervention research, as interventions are often cost-intensive and require human resources (Plucker & Makel, 2021). Thus, successful replications might provide stakeholders with a robust basis on which decisions can be made to further invest in an intervention or even to scale it up, whereas, in contrast, unsuccessful replications of previous desirable results might indicate that the assumed underlying effect mechanisms or contextual influences might need further investigation. In this light, the effects of math learning program interventions and, more specifically, the effects of Math Garden and practice behavior with Math Garden on performance and affective-motivational outcomes need to be put to the test again. Thus, I focused on replicating Study 2 by using the same constructs (with slightly different operationalizations) as well as the same statistical methods but with a different and even larger sample.

### 6.2 Research Question and Hypotheses

Based on the research question addressed in Study 2, this replication also aimed to answer the question of how the provision of and practice behavior with Math Garden was related to students' addition and subtraction performance, their math self-concept, and their math anxiety. The results of Study 2 provided support for the hypothesis that the more subtraction tasks students completed, the higher their subtraction performance improvement was, and, furthermore, that the provision of Math Garden was associated with an increase in students' math self-concept in the experimental group in contrast to the control group. Regarding all other expected effects, no supporting empirical evidence was found. However, based on the results of Study 2, the unsupported hypotheses should not be directly rejected as they still have strong theoretical foundations and are supported by previous findings stated in the literature. In addition, the original study had a lack of power due to the reduced sample size in the posttest as a result of COVID-19 school closures. Thus, for the replication analysis, I again expected that the provision of Math Garden in the experimental group would foster students' addition and subtraction performance, as well as their math self-concept, and would decrease students' math anxiety in contrast to the control group (H1.1–H1.4, respectively). Regarding the experimental group, I expected that the more tasks students completed (addition and subtraction), the higher their addition and subtraction performance and their math self-concept and the lower their math anxiety after the intervention would be (H2.1–H2.4, respectively). I

expected the same regarding the number of weeks spent practicing (addition and subtraction) tasks (H3.1–H3.4, respectively) for students in the experimental group. As in Study 2, I controlled the analyses for students' pretest scores to account for differences in their baseline values, their gender, a possible migration background, and tablet typing speed.

### 6.3 Methods

#### 6.3.1 Design and Procedure

The data used in this replication analysis come from a larger project investigating the effectiveness of different approaches to supporting students in math (i.e., Implementation Period 2; Carstensen et al., 2023). Within this project, students were randomly assigned at the school level to either an experimental group that practiced with Math Garden or a waitlist control group. For this replication, I used the student data collected from October 2021 (pretest,  $T_1$ ) to February 2022 (posttest,  $T_2$ ). The implementation of Math Garden and the testing of students followed the same procedure as in Study 2, but the length of the practice period differed slightly, consisting of 28 weeks instead of 22 weeks. On average, the program was available to the students for 24.7 weeks ( $SD = 1.4$ ,  $min = 20$ ,  $max = 28$ ).

#### 6.3.2 Sample

The sample comprised 1,093 fifth-grade students from 65 classes at 35 urban comprehensive (non-academic-track) schools across Germany, who participated in the aforementioned research project. Thus, the sample size was more than three times larger than that of Study 2 and differed in the federal states included (Schleswig-Holstein in Study 2 vs. Berlin, Brandenburg, Hamburg, North Rhine-Westphalia, Rhineland-Palatinate, and Schleswig-Holstein in the replication). In the experimental group, the data of two children were excluded due to an implausibly high level of practice (more than 2,000 tasks within one week), as were the data of 11 students because of missing data on all relevant variables, that is, questionnaire, test, and practice data. Furthermore, four classes were excluded because they stopped participating in the study after the pretest ( $n = 48$ ). The final analytic sample consisted of 829 students in the experimental group (41.2% female, 48.3% with a migration background; see Table 6.1 for further sample characteristics) and 203 students in the waitlist control group (49.4% female, 49.1% with a migration background).

### 6.3.3 Measures

#### 6.3.3.1 Performance and Affective-Motivational Outcomes

As in Study 2, the addition and subtraction scales of the Heidelberg calculation test by Haffner et al. (2005) were used to measure math performance. Similar to Study 2, I made an item selection (Items 14 to 24 for the addition scale and Items 15 to 25 for the subtraction scale), and I found good retest reliability values for both scales ( $r_{ttAdd} = .64$ ,  $r_{ttSub} = .70$ ).

The assessment of math self-concept differed slightly from the assessment in Study 2, as the German adaption of the math self-concept scale from the Self-Description Questionnaire was used (Arens et al., 2011; Marsh, 1990). The scale consists of eight items (e.g. “I like math.”), which have to be answered on a five-point Likert scale (1 = *wrong* to 5 = *true*). Cronbach’s alpha was .92 in the pretest and .93 in the posttest. I found significant correlations between pre- and posttest ( $r_{tt} = .61$ ).

As in Study 2, math anxiety was measured with the subscale of the Math Anxiety Rating Scale for Fourth to Sixth Grades by Roick et al. (2013). Cronbach’s alpha was .91 in the pretest and .91 in the posttest. Pretest and posttest values were also significantly correlated ( $r_{tt} = .50$ ).

#### 6.3.3.2 Practice Behavior

In line with the original study, I developed six practice frequency scores with the trace data collected in Math Garden: I summed up the number of tasks completed in the domain of (1) addition and (2) subtraction, as well as the number of weeks for which students practiced (3) addition tasks and (4) subtraction tasks. Furthermore, I calculated an (5) overall practice frequency score by summing up the total number of math-related tasks completed in Math Garden, which included tasks on addition, subtraction, multiplication, division, mix (mixture of basic arithmetic operations), counting (of objects), series (logical completion of a given number series), numerals (tasks in which numbers must be combined to obtain a desired target number), and tables (multiplication table tasks from 1 to 10). Finally, I calculated a (6) score for each student using the total number of weeks for which they practiced math-related tasks (see above). The descriptives for these six different scores are reported in Table 6.2.

#### 6.3.3.3 Covariates

At the student level, students’ gender was coded as 0 = *male* and 1 = *female*. Instead of assessing a migration background with a dummy coded item (Study 2: “Are other languages besides German spoken at home?” with 0 = *no* and 1 = *yes* as the answer options), the migration background in the replication was measured continuously, that is, with the question “How often do you speak German at home?”, with 1 = *I always speak German at home*, 2 = *I almost always speak German at home*, 3 = *I sometimes speak German at home and sometimes another*



language, 4 = *I never speak German at home* as answer options. However, for the analyses, I built a dummy coded version in terms of the direct replication approach (0 = *I always speak German at home*, 1 = *I almost always speak German at home*, *I sometimes speak German at home and sometimes another language*, *I never speak German at home*). I used the speed scale of the Heidelberg calculation test (Haffner et al., 2005) to control for students' tablet typing speed.

#### **6.3.4 Data Analysis**

The statistical approach was identical to that of Study 2. Addressing the effects of the provision of Math Garden on math performance, math-self-concept, and math anxiety, I ran multiple regression analyses in *Mplus* (Version 8.5; Muthén and Muthén, 2008–2020) with the group assignment as the predictor (0 = *waitlist control group*, 1 = *experimental group*). The dependent variables were the posttest scores after the pretest scores and the other covariates were controlled for. To investigate the relationship between students' practice behavior and the suggested outcomes, I analyzed only the values of the experimental group. The dependent variables were the posttest scores after the pretest scores and other covariates were controlled for, and the predictors were either the number of tasks completed or the number of weeks spent practicing. However, due to the two distinct performance measures (i.e., the domains of addition and subtraction), I made further differentiations, as was done in Study 2.

Due to the hierarchical structure of the data (students within classes), I used the TYPE=COMPLEX command to adjust standard errors for nonindependence within classes. I used the robust maximum likelihood estimator, and all continuous independent variables were *z*-standardized ( $M = 0$ ,  $SD = 1$ ). Significance testing was undertaken at the 5% level.

I found missing data on all variables. Regarding the self-report and test measures, this was because students in the final analytic sample were absent during at least one of the two testing sessions or left out single items. The percentage of missing data varied from 0% on group assignment to 15.7% on students' math anxiety, with an average missing rate of 8.3%. Missing data regarding the practice behavior variables occurred because students participated in the testing session and received login data but it was not possible to match their responses from the testing session to their practice data ( $n = 161$ ; 19.4%). I used full information maximum likelihood to account for missing data on the independent variables and on the dependent variable (Enders, 2010).

## 6.4 Results

### 6.4.1 Descriptives and Preliminary Analyses

The means and standard deviations for the self-report and performance measures are displayed in Table 6.1. Regarding the relevant study variables, I noticed that, on average, the mean values for the performance measures descriptively increased in both conditions from pretest to posttest, which was also the case for the math anxiety scores. In contrast, math self-concept scores dropped slightly in both conditions.

**Table 6.1**

*Descriptives of Self-Report and Performance Measures*

	Condition					
	All students		Waitlist control		Experimental	
	<i>N</i> = 1,032		<i>n</i> = 203		<i>n</i> = 829	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Math addition performance T <sub>1</sub>	7.50	2.94	7.84	2.81	7.42	2.97
Math addition performance T <sub>2</sub>	8.21	2.70	8.26	2.87	8.20	2.66
Math subtraction performance T <sub>1</sub>	6.36	3.74	6.74	3.78	6.26	3.73
Math subtraction performance T <sub>2</sub>	7.07	3.59	7.23	3.60	7.03	3.59
Math self-concept T <sub>1</sub>	3.46	0.94	3.61	0.99	3.43	0.93
Math self-concept T <sub>2</sub>	3.41	0.98	3.43	1.08	3.40	0.96
Math anxiety T <sub>1</sub>	3.18	1.16	3.14	1.22	3.19	1.14
Math anxiety T <sub>2</sub>	3.22	1.18	3.17	1.13	3.23	1.19
Gender <sup>a</sup>	0.47	0.50	0.41	0.49	0.49	0.50
Migration background T <sub>1</sub> <sup>b</sup>	0.49	0.50	0.48	0.50	0.49	0.50
Tablet typing speed T <sub>1</sub>	9.88	4.17	9.79	4.12	9.91	4.19

*Note.* Reference categories: <sup>a</sup>male. <sup>b</sup>no other languages spoken at home besides German.

Table 6.2 displays the descriptives of practice behavior in Math Garden. As in Study 2, the high variance in practice over the 28-week period is noticeable. Overall, students practiced, on average, for 3.56 weeks (*SD* = 3.96) with Math Garden, with a minimum of zero weeks and a maximum of 22 weeks.

**Table 6.2***Descriptives of Practice Behavior in Math Garden*

	<i>M</i>	<i>SD</i>	Range
Completed addition tasks	24.97	48.33	0–771
Completed subtraction tasks	19.92	43.86	0–486
Overall completed tasks	312.98	625.04	0–7771
Practiced weeks addition tasks	1.36	1.71	0–10
Practiced weeks subtraction tasks	1.12	1.66	0–12
Overall practiced weeks	3.56	3.69	0–22

*Note.*  $N = 829$ .

Tables 6.3 and 6.4 show the correlations between the independent variables and the dependent variables for all students and for the experimental group only. Regarding the possible effects of Math Garden on the outcome variables, I found no significant correlations between students' group assignment and the outcomes in the posttest. Still, for the experimental group, I found a significant but small correlation between the number of subtraction tasks completed and students' subtraction performance in the posttest ( $r = .08$ ; Cohen, 1988).

**Table 6.3***Correlation Matrix of the Study Variables*

	1	2	3	4	5	6	7	8	9	10	11
1. Group assignment <sup>a</sup>											
2. Math addition performance T <sub>1</sub>	-.06										
3. Math subtraction performance T <sub>1</sub>	-.05	<b>.74</b>									
4. Math self-concept T1	<b>-.08</b>	<b>.34</b>	<b>.36</b>								
5. Math anxiety T <sub>1</sub>	.02	<b>-.11</b>	<b>-.13</b>	<b>-.22</b>							
6. Gender <sup>b</sup>	.06	<b>-.13</b>	<b>-.23</b>	<b>-.17</b>	<b>.22</b>						
7. Migration background T <sub>1</sub> <sup>c</sup>	.01	-.02	-.01	.02	<b>.12</b>	<b>.09</b>					
8. Tablet typing speed	.01	<b>.25</b>	<b>.21</b>	<b>.13</b>	.04	<b>.07</b>	-.02				
9. Math addition performance T <sub>2</sub>	-.01	<b>.64</b>	<b>.61</b>	<b>.26</b>	<b>-.15</b>	<b>-.07</b>	.01	<b>.21</b>			
10. Math subtraction performance T <sub>2</sub>	-.02	<b>.59</b>	<b>.70</b>	<b>.31</b>	<b>-.17</b>	<b>-.16</b>	-.02	<b>.15</b>	<b>.71</b>		
11. Math self-concept T <sub>2</sub>	-.01	<b>.25</b>	<b>.28</b>	<b>.61</b>	<b>-.20</b>	<b>-.16</b>	.01	<b>.12</b>	<b>.28</b>	<b>.34</b>	
12. Math anxiety T <sub>2</sub>	.02	<b>-.15</b>	<b>-.16</b>	<b>-.18</b>	<b>.50</b>	<b>.23</b>	<b>.12</b>	-.06	<b>-.10</b>	<b>-.14</b>	<b>-.32</b>

*Note.* Correlations refer to all students,  $N = 1,032$ . Reference categories: <sup>a</sup>waitlist control condition. <sup>b</sup>male. <sup>c</sup>no other languages spoken at home besides German.

**Table 6.4***Correlation Matrix for Practice Behavior and Outcomes*

	1	2	3	4	5	6	7	8	9
1. Completed addition tasks									
2. Completed subtraction tasks	<b>.74</b>								
3. Overall completed tasks	<b>.76</b>	<b>.76</b>							
4. Practiced weeks addition tasks	<b>.69</b>	<b>.65</b>	<b>.71</b>						
5. Practiced weeks subtraction tasks	<b>.58</b>	<b>.75</b>	<b>.72</b>	<b>.78</b>					
6. Overall practiced weeks	<b>.59</b>	<b>.62</b>	<b>.78</b>	<b>.78</b>	<b>.80</b>				
7. Math addition performance T <sub>2</sub>	.06	.03	.03	.01	.04	.05			
8. Math subtraction performance T <sub>2</sub>	.07	<b>.08</b>	.04	.03	.04	.05	<b>.70</b>		
9. Math self-concept T2	.00	.01	.00	.06	.04	.07	<b>.26</b>	<b>.33</b>	
10. Math anxiety T <sub>2</sub>	-.07	-.05	-.06	-.04	-.02	-.06	<b>-.09</b>	<b>-.13</b>	<b>-.34</b>

*Note.* Correlations refer to the experimental condition only,  $n = 829$ .

### 6.4.2 Effects of Providing Math Garden and of Practice Behavior

The first research question focused on the effects of the provision of Math Garden, that is, on the differences in the posttest scores regarding the addition and subtraction performance, math self-concept, and math anxiety of the waitlist control group compared to the experimental group after controlling for baseline values. I found no statistically significant effect for the provision of Math Garden on performance or on the affective-motivational variables (H1.1–H1.4; see Table 6.5) when taking students' gender, a possible migration background, and tablet typing speed as covariates into account.

I also investigated how the number of tasks completed in Math Garden was related to addition and subtraction performance, math self-concept, and math anxiety in the experimental group (see Table 6.6). After including all covariates in the statistical model, I found a significant effect of the number of subtraction tasks completed on students' subtraction performance. As expected (H2.2), the more subtraction tasks students completed, the higher their subtraction performance after the intervention was ( $\beta = .08$ ,  $p = .037$ ). Regarding H2.1, H2.3, and H2.4, I did not find any statistically significant effects.

Addressing the last research question, I examined how the number of weeks spent practicing with Math Garden predicted addition and subtraction performance, math self-concept, and math anxiety in the experimental group (see Table 6.6). I did not find any statistically significant effects regarding the proposed hypotheses (H3.1–H3.4).

**Table 6.5***Results on Effects of Providing Math Garden*

	Addition performance			Subtraction performance			Math self-concept			Math anxiety		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<i>Predictors</i>												
Outcome at T <sub>1</sub>	<b>.63</b>	.03	.000	<b>.70</b>	.03	.000	<b>.59</b>	.03	.000	<b>.47</b>	.03	.000
Math Learning Program <sup>a</sup>	.03	.03	.248	.02	.03	.359	.02	.04	.619	.00	.03	.986
<i>Covariates</i>												
Gender <sup>b</sup>	.01	.03	.664	-.00	.03	.916	-.07	.04	.070	<b>.13</b>	.03	.000
Migration background T <sub>1</sub> <sup>c</sup>	.00	.03	.975	-.03	.02	.119	-.01	.03	.637	<b>.06</b>	.03	.046
Tablet typing speed T <sub>1</sub>	<b>.05</b>	.03	.039	.00	.03	.924	.05	.03	.125	<b>-.08</b>	.03	.009
<i>R</i> <sup>2</sup>	.42			.49			.38			.28		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $N = 1,032$ . Reference categories: <sup>a</sup>wait-list control condition. <sup>b</sup>male. <sup>c</sup>no other languages spoken at home besides German.

**Table 6.6***Results on Effects of Practice Behavior*

	Addition			Subtraction			Math			Math		
	performance			performance			self-concept			anxiety		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
COMPLETED TASKS												
<i>Predictors</i>												
Outcome at T <sub>1</sub>	<b>.61</b>	.03	.001	<b>.70</b>	.03	.000	<b>.57</b>	.04	.000	<b>.47</b>	.03	.000
Completed tasks	.02	.04	.767	<b>.08</b>	.02	.037	.00	.04	.833	−.02	.03	.470
<i>Covariates</i>												
Gender <sup>a</sup>	−.01	.03	.205	−.01	.03	.824	−.07	.04	.099	<b>.11</b>	.04	.003
Migration background T1 <sup>b</sup>	−.01	.03	.593	<b>−.05</b>	.03	.034	−.00	.03	.926	.06	.03	.084
Tablet typing speed T <sub>1</sub>	.04	.03	.128	−.01	.03	.782	.03	.03	.394	−.06	.03	.054
<i>R</i> <sup>2</sup>	.40			.49			.35			.26		
PRACTICED WEEKS												
<i>Predictors</i>												
Outcome at T <sub>1</sub>	<b>.62</b>	.03	.001	<b>.70</b>	.03	.000	<b>.57</b>	.04	.000	<b>.47</b>	.03	.000
Practiced weeks	−.02	.04	.611	.00	.03	.792	.03	.04	.420	−.01	.04	.704
<i>Covariates</i>												
Gender <sup>a</sup>	−.01	.03	.707	−.01	.03	.798	−.07	.04	.111	<b>.11</b>	.04	.003
Migration background T1 <sup>b</sup>	−.02	.03	.631	<b>−.05</b>	.03	.028	.00	.03	.924	.06	.03	.078
Tablet typing speed T <sub>1</sub>	.04	.03	.166	−.01	.03	.754	.03	.03	.370	−.06	.03	.060
<i>R</i> <sup>2</sup>	.40			.48			.35			.26		

*Note.* Significant coefficients are printed in bold ( $p \leq .05$ ). Continuous predictors were standardized ( $M = 0$ ;  $SD = 1$ ).  $n = 829$ . Reference categories: <sup>a</sup>male. <sup>b</sup>no other languages spoken at home besides German.



### 6.5 Discussion

The additional analyses aimed to directly replicate the results of Study 2, as replications in the field of educational psychology are rare but are urgently needed to provide evidence for the robustness of results, especially when it comes to the effects of interventions. My findings showed that the more subtraction tasks students completed in Math Garden, the higher their subtraction performance in the posttest was. This beneficial result was in line with my hypothesis and with the results from the original study. It again underlines the importance of ensuring that students systematically engage with a math learning program if they are to benefit from its implementation (Spitzer, 2022). However, similar to the results of Study 2, the results of the replication analysis did not find any link between the number of addition tasks completed and addition performance; this might support the assumption that the amount of practice might be especially helpful for more difficult domains such as subtraction (Anderson et al., 2022). Future conceptual replications could therefore examine effects on multiplication and division performance to further investigate this assumption.

The results of the additional analyses did not support any other hypothesis. However, the original study showed a beneficial (but rather small) effect of the provision of Math Garden on students' math self-concept in the experimental condition. A possible reason for why the replication study did not replicate this result might be its use of a different math self-concept scale. This instrument was validated for fifth-grade students and showed high internal reliability scores, which is why measurement problems can be ruled out. Thus, the nonsignificant finding might indicate that the effect is not stable across different operationalizations. In addition, the sample characteristics of the replication study might also explain the missing effect, as students in the replication had, on average, higher math self-concept scores in the pretest than students in the original study. Therefore, the beneficial features of adaptivity and feedback, which were assumed to be the underlying theoretical mechanisms of the expected effect, might no longer have come to bear because students already had a positive image of their math-related abilities. Ideas on how future studies could investigate the effects of adaptivity and feedback on motivational outcomes in more in-depth can be found in the General Discussion (Chapter 8).

All other nonsignificant results were in line with the findings of Study 2. The missing effects of the program provision on the performance measures might be due to students in the experimental group, on average, not practicing enough with Math Garden to benefit from the program in contrast to the control group. This pattern of findings supports the call for researchers to include objectively measured practice behavior in their analyses when

investigating the implementation quality of such learning program interventions (Vanbecelaere et al., 2023); such analyses could, in turn, facilitate potential conclusions about missing effects.

Contrary to my expectations, the number of weeks for which students practiced with Math Garden was not associated with performance. This might be due to the same problem as in the original study, namely, that students only practiced for a small number of weeks, which means that probably not enough meaningful variance was created within this variable to explain differences in students' performance.

Regarding the nonsignificant findings regarding the effect of both the provision of and practice behavior with Math Garden on math anxiety in the replication study and in Study 2, there are two possible explanations for the missing effects. First, math anxiety might have been too deeply rooted in the students, so thoughts of failure might have taken up too much of students' working memory (Ashcraft & Kirk, 2001). Therefore, they might not have had the resources to perceive that the tasks in Math Garden were adapted to their ability level; such adaptivity is hypothesized to reduce the likelihood of failure and thus lead to more control over successfully solving a task (Pekrun, 2006). The way in which such assumed effect mechanisms could be further investigated is discussed more precisely in the General Discussion (see Chapter 8). Second, the measure used to assess math anxiety in both Study 2 and its replication asked about test anxiety. However, math anxiety is a multidimensional construct (Barroso et al., 2021; Lukowski et al., 2019) and, as Math Garden is understood less as a test and more as a tool for individual skill development, the program might be more likely to affect other math anxiety components, for instance, math calculation anxiety (Lukowski et al., 2019).

## 6.6 References

- Anderson, S., Sommerhoff, D., Schurig, M., Ufer, S., & Gebhardt, M. (2022). Developing learning progress monitoring tests using difficulty-generating item characteristics. An example for basic arithmetic operations in primary schools. *1866-6671*, 14.  
<https://doi.org/10.25656/01:24875>
- Arens, A. K., Trautwein, U., & Hasselhorn, M. (2011). Erfassung des Selbstkonzepts im mittleren Kindesalter: Validierung einer deutschen Version des SDQ I [Assessment of self-concept in middle childhood: Validation of a German version of the SDQ I]. *Zeitschrift Für Pädagogische Psychologie*, 25(2), 131–144.  
<https://doi.org/10.1024/1010-0652/a000030>

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology*, 130(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Carstensen, B., Hilz, A., Benckwitz, L., Asbury, K., Roloff, J., & Guill, K. Intervention in mathematics to professionally support low-achieving students. (2023). Retrieved from [osf.io/kdbcr](https://osf.io/kdbcr). Advance online publication. <https://doi.org/10.17605/OSF.IO/KDBCR>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York: Guilford Press. <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10389908>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4: Heidelberger Rechentest – Erfassung mathematischer Basiskompetenzen im Grundschulalter [HRT 1-4: Heidelberger calculation test – Assessment of basic mathematical skills at primary school age]*. Göttingen: Hogrefe.
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>
- Lukowski, S. L., DiTrapani, J. B., Jeon, M., Wang, Z., Schenker, V. J., Doran, M. M., Hart, S. A., Mazzocco, M. M. M., Willcutt, E. G., Thompson, L. A., & Petrill, S. A. (2019). Multidimensionality in the measurement of math-specific anxiety and its relationship with mathematical performance. *Learning and Individual Differences*, 70, 228–235. <https://doi.org/10.1016/j.lindif.2016.07.007>
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82(4), 623–636. <https://doi.org/10.1037/0022-0663.82.4.623>
- Muthén, B. O., & Muthén, L. K. (2008-2020). *Mplus computer software (version 8.5)*. Los Angeles, CA: Muthén & Muthén.

- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, 56(2), 90–100. <https://doi.org/10.1080/00461520.2021.1895796>
- Roick, T., Göllitz, D., & Hasselhorn, M. (2013). Affektive Komponenten der Mathematikkompetenz. Die Mathematikangst-Ratingskala für vierte bis sechste Klassen (MARS 4-6) [Affective components of mathematics competence. The Mathematics Anxiety Rating Scale for 4th to 6th grade (MARS 4-6)]. In M. Hasselhorn, A. Heinze, W. Schneider & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen* (pp. 205–219). Göttingen: Hogrefe.
- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Vanbecelaere, S., Demedts, F., Reynvoet, B., & Depaepe, F. (2023). Toward a framework for analyzing adaptive digital games' research effectiveness. *International Journal of Serious Games*, 10(4), 77–91. <https://doi.org/10.17083/ijsg.v10i4.618>

# 7

## STUDY 3: TRACING STUDENTS’ PRACTICE BEHAVIOR IN AN ADAPTIVE MATH LEARNING PROGRAM: DOES IT MEDIATE THE MATH ANXIETY– PERFORMANCE LINK?

*A revised version of this chapter has been published as:*

Hilz, A., Hofman, A., Jansen, B., & Aldrup, K. (2025). Tracing students’ practice behavior in an adaptive math learning program: Does it mediate the math anxiety–performance link? *Learning and Instruction*, 98, 102113.

<https://doi.org/10.1016/j.learninstruc.2025.102113>

### Abstract

**Background:** It has been suggested that math anxiety negatively affects students' performance because math anxious students tend to avoid math-related tasks. As most studies that investigated this assumed relationship focused on self-report measures to assess avoidance behavior, approaches that use objective process measures, which are less likely to be affected by biases, are needed.

**Aims:** We conceptualized avoidance behavior in terms of trace data within an adaptive math learning program. We used a three-wave longitudinal study design to approach the question of how math anxiety is associated with learning behavior that contributes to math performance improvement in school.

**Sample:** The study comprised 890 students at the beginning of secondary school in Germany.

**Methods:** We traced students' practice behavior (i.e., completed tasks within an adaptive learning program; Math Garden ) over 45 weeks.

**Results:** The results of the mediation analyses indicated that math anxious students did not practice less with the math learning program, which we found to be effective in improving math performance. However, this means that avoidance behavior does not explain the weaker performance of math anxious students.

**Conclusions:** Math anxiety is not the central influencing factor for practice behavior in our model, but practitioners need to be aware of the fact that interindividual student differences are still associated with students' practice (e.g., having a migration background). Thus, students need to be encouraged to practice with an adaptive math learning program, as this contributes to their performance improvement

**Keywords:** mathematics, anxiety, engagement, practice behavior, math learning programs

### 7.1 Introduction

Math anxiety is a widespread issue faced by students across the globe; researchers have shown that it affects one in three individuals (Sorvo et al., 2017). This finding underscores the significance of addressing the consequences of math anxiety. From a theoretical perspective, it is assumed that math anxiety is linked to avoidance strategies when it comes to math and math-related tasks (Eysenck & Calvo, 1992), consequently resulting in poorer math skills. Research has already found evidence that math anxiety negatively impacts students' math performance (e.g., Aldrup et al., 2020), but only little is known about the mediating role of avoidance behavior in this relationship. Studies investigating avoidance strategies have, so far, mostly focused on math-related course choices in higher education (Hembree, 1990; LeFevre et al., 1992). However, opting out of math at that point is likely a result of adverse math-related trajectories throughout the school career. Math anxiety already occurs in elementary and secondary education (Ramirez et al., 2013). Therefore, this may be associated with the avoidance of math learning opportunities at an early stage and with difficulties in math skill acquisition. In turn, the math skills acquired during elementary and early secondary education profoundly affect subsequent math learning and performance (Andersson, 2010).

The few studies that investigated the relationship between math anxiety and avoidance behavior in secondary school students mostly focused on self-report measures regarding students' avoidance strategies, but this led to mixed results (Hasty et al., 2021; Quintero et al., 2022). One problem with self-reports is that they are often biased (Donaldson & Grant-Vallone, 2002); for instance, math anxious students might report higher levels of avoidance behavior to attribute their poorer math performance to low engagement rather than to low abilities (e.g., self-serving bias). In contrast, process measures, such as data collected when students practice with learning programs, have the potential to objectively and economically measure students' avoidance behavior. Hence, the current study used an adaptive arithmetic learning program to trace the tasks completed by students as an indicator of their avoidance behavior. Using this new approach, we examined the mediating role of students' practice behavior with the learning program in a sample of 890 fifth-grade students in Germany; we aimed to gain information that would help to explain the link between math anxiety and math performance.

### 7.1.1 Defining Math Anxiety

Math anxiety is characterized by feelings of helplessness, tension, and worry in math and math-related situations (Hembree, 1990). Regarding its conceptualization, an increasing number of studies point out the multidimensionality of math anxiety (e.g., Lee, 2009; Lukowski et al., 2019). Despite the absence of a final consensus regarding the quantity and nature of the dimensions, two complementary assumptions are most commonly encountered.

Wigfield and Meece (1988) assumed that math anxiety has two components, namely an affective emotionality component, which includes perceptions of fear as well as physical reactions, and a cognitive worry component, which involves negative thoughts about one's math competence. The second assumption regarding the conceptualization of math anxiety is that it occurs in different contexts. These situation-related contexts have been identified as math learning anxiety and math test anxiety (Carey et al., 2017; Caviola et al., 2017; Hill et al., 2016). Math learning anxiety is related to unpleasant feelings when it comes to everyday math instruction, for example, when performing calculations. In contrast, math test anxiety occurs especially in test- and performance-related situations.

Despite these various dimensions of math anxiety, studies that investigate the associations between math anxiety and performance outcomes often only use a one-dimensional approach. However, the effects on performance can differ for the math anxiety dimensions, for instance, two studies found that the relationship with performance was strongest for the worry component (Barroso et al., 2021; Lukowski et al., 2019). Hence, more research that considers a multidimensional approach is needed.

### 7.1.2 How is Math Anxiety Linked to Math Performance?

Comprehensive empirical evidence suggests that there is a low to moderate correlation between math anxiety and students' math performance (Barroso et al., 2021; Ma, 1999). Two prominent approaches have been proposed to explain this relationship. The deficit theory claims that low competencies in math lead to academic failure and, thus, to anxiety (Carey et al., 2015). In contrast, Ashcraft and colleagues' disruption account (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007) suggests that math anxiety causes lower math performance because rumination and negative thoughts interfere with the processing of mathematical tasks, and they distract attention and demand working memory resources.

However, to explain the effects of math anxiety on performance, the behavioral aspect of math anxiety can also be considered, which has received less attention so far. This aspect suggests that math anxious students avoid math-related tasks and activities as the confrontation



is perceived as aversive in terms of negative physiological and psychological reactions (Eysenck & Calvo, 1992; Eysenck et al., 2007). Previous research has already shown that students with elevated levels of math anxiety show more avoidance strategies (Hembree, 1990; LeFevre et al., 1992), but, instead of examining the mediating role of avoidance behavior, most studies have only investigated the relationship between math anxiety and avoidance behavior (cf. Quintero et al., 2022). Still, the fewer opportunities that math anxious students have to engage with mathematical situations might explain their weaker math performance. Thus, when attempting to extend previous research by considering avoidance behavior as the mediating mechanism between math anxiety and math performance, the question arises as to which different forms of avoidance behavior individuals can display.

### **7.1.3 Conceptualizing Avoidance Behavior**

Following Wong and Liem (2022), we define behavioral engagement as “the extent to which students intentionally exert effort during learning activities” (Wong & Liem, 2022, p. 123). By showing behavioral engagement in relevant learning tasks, students’ competencies can consequently develop (Bråten et al., 2022; Cheung & Slavin, 2013). In contrast, low behavioral engagement can be understood as avoidance behavior that interferes with competence improvement.

Examining course and career decisions, or the intention to make such decisions, has been a common approach to assessing avoidance behavior since the 1990s (Hembree, 1990; LeFevre et al., 1992; Meece et al., 1990). However, whereas these indicators are applicable in higher education, opting out of math class is rarely an option at lower grade levels in the school context. Thus, different approaches for operationalizing math avoidance behavior are needed to understand the consequences of math anxiety in lower grade levels.

Hasty et al. (2021) investigated the relationship between math anxiety and potential avoidance behavior in secondary school students by asking students about the time spent on math-related activities, but they found no association. Quintero et al. (2022) operationalized avoidance behavior as the lack of secondary school students’ self-reported cognitive-behavioral engagement during math instruction, and their results supported the suggested negative relationship. Furthermore, they extended Hasty et al.’s approach by also investigating effects on students’ performance. As they found that lower levels of cognitive-behavioral engagement indeed explained the negative link between math anxiety and math performance, they provided first evidence for the mediating role of avoidance behavior (Quintero et al., 2022).

However, although using self-report measures to assess avoidance strategies is a common method (Muenks et al., 2017; Skinner et al., 2009), the divergent results of Quintero et al. (2022) and Hasty et al. (2021) could be an indicator that at least one self-report assessment was potentially biased.

### **7.1.4 The Potential of Process Measures**

A common bias of self-reports is a self-serving bias (Donaldson & Grant-Vallone, 2002); students may downplay their actual behavioral engagement in order to avoid attributing failures to their abilities. To overcome this problem, researchers should focus on measures that are more objective (Azevedo, 2015).

One possibility might be to use trained observers during math instruction (Pianta et al., 2012). However, this approach is time-consuming and requires many human resources. Another possibility is the assessment through process measures (Wong & Liem, 2022). Nowadays, programs and applications usually record their users' behavior in terms of log and trace data, which have the potential to economically assess students' behavioral engagement mostly without biases. This also applies to many math learning programs. A wide variety of programs that have been designed for scientific purposes can objectively operationalize behavioral engagement as the tasks completed in the program or the time spent on the program (e.g., see Hassler Hallstedt et al., 2018; Hilz, Guill, Roloff, Sommerhoff, & Aldrup, 2023; Spitzer, 2022).

In this study, we used the adaptive arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014). Math Garden is a drill-and-practice program that allows students to strengthen basic arithmetic operations. It works adaptively, which means that an algorithm regulates the assignment of the tasks in such a way that students have a preset average percentage of correct answers of 75% (for a more detailed description, see Klinkenberg et al., 2011). After each task, direct corrective feedback is given. Math Garden automatically traces the number of tasks completed per domain per week for each practicing student. With regard to the relationship between math anxiety and avoidance behavior, a study conducted by Hilz, Guill, Roloff, Aldrup, and Köller (2023) already showed that math anxious students completed fewer tasks with Math Garden. However, the authors did not investigate the relation with math performance.

### **7.1.5 The Present Study**

This study addressed the existing research gap by aiming to explain how the relationship between math anxiety and math performance is mediated by students' avoidance behavior. We aimed to answer the research question of whether math anxious students avoided practicing

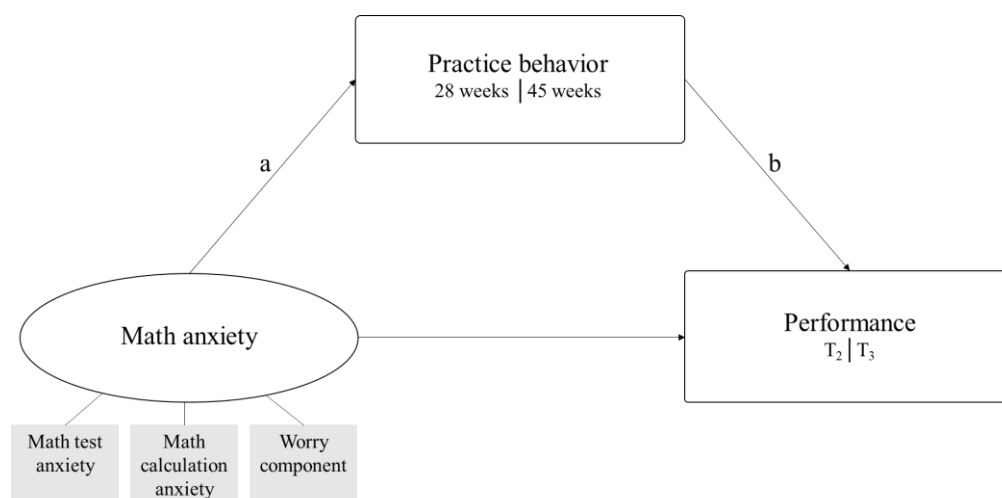
(i.e., practice fewer tasks) with the arithmetic learning program Math Garden and, in turn, whether this hindered them in improving their math performance. Thus, we extended the suggested mediation model of Quintero et al. (2022) but we assessed avoidance behavior with less biased process measures than students' self-reports. Moreover, we used a longitudinal design with three measurement points: the beginning ( $T_1$ ), the middle ( $T_2$ ), and the end ( $T_3$ ) of a school year (interval  $T_1$ – $T_2$ : 28 weeks of practicing, interval  $T_1$ – $T_3$ : 45 weeks of practicing). This allowed us to examine whether math anxiety at the beginning of the school year predicted students' practice behavior in the medium and long term, potentially precluding achievement gains due to continued avoidance of practice.

We used a multidimensional approach to examine students' math anxiety, comprising the subcomponents affective "math test anxiety", affective "math calculation anxiety", and cognitive "worry", as earlier research emphasized the multidimensionality of math anxiety (Carey et al., 2017; Caviola et al., 2017; Hill et al., 2016; Lukowski et al., 2019; Wigfield & Meece, 1988) and already showed that different components of math anxiety vary in the strength of their relationship with math performance (Barroso et al., 2021; Lukowski et al., 2019). For this reason, we aimed to test our research question with both a global math anxiety construct and the subcomponents. The following hypotheses and the statistical approach used to investigate the hypotheses were preregistered<sup>1</sup>.

First, we hypothesized for all suggested anxiety models that students' math anxiety (i.e., global construct and single subcomponents) would be associated with less practice in Math Garden (Hypothesis 1, see Path a in Figure 7.1). Second, we hypothesized that higher levels of practice behavior would be linked to more improvement in student performance (Hypothesis 2, see Path b). To answer our main research question, we hypothesized that math anxiety would impair improvements in math performance because math anxious students would practice less with Math Garden (Hypothesis 3, Path a\*b). We tested all hypotheses for both practice periods (i.e.,  $T_1$ – $T_2$  and  $T_1$ – $T_3$ ).

---

<sup>1</sup><https://doi.org/10.17605/OSF.IO/38GFN>

**Figure 7.1***Mediation Model*

*Note.* T<sub>2</sub>|T<sub>3</sub>=second/third point of measurement.

**7.2 Methods****7.2.1 Design and Procedure**

The study was part of a larger project investigating the effectiveness of different approaches to supporting students in math (Carstensen et al., 2023). Within this project, students were randomly assigned at the school level to either an experimental group or a waitlist control group. For this study, we only considered the experimental group that practiced with Math Garden. An ethics committee reviewed the study; student data were collected in schools from October 2021 to July 2022 in compliance with all data protection requirements. Participation in the study was voluntary and parental permission was obtained before students participated in the study.

In a pretest (T<sub>1</sub>) at the beginning of the school year, the participating classes completed math performance tests and answered questionnaires that assessed students' math anxiety on tablets in order to measure their baseline values. At the end of the testing session, students received a 10-minute introduction to Math Garden and tried it out on tablets using personalized login data provided by their teachers. While practicing, they received a brief oral introduction to the main principles of the program. At the end of the introductory practice session, students were told that they should practice with the program at home and in class on their own devices. In cases in which students who had permission to participate in the study were absent at T<sub>1</sub>, we asked their teachers to introduce these students to the program later. In February 2022 (T<sub>2</sub>), after 28 weeks of practice, and in July 2022 (T<sub>3</sub>), after 45 weeks of practice, students completed

questionnaires and math tasks again. Between the testing sessions, the tasks completed in Math Garden were traced.

### 7.2.2 Sample

The data for this study were obtained from a sample of 890 fifth-grade students in 53 classes at 29 urban comprehensive schools across Germany. Comprehensive schools offer different levels of school degrees but primarily prepare students for vocational training (Maaz et al., 2008). Consequently, the percentage of lower-achieving students is higher at this type of school. After data inspection, we excluded two students due to an implausibly high level of practice (more than 20,000 tasks within 45 weeks), which might have been caused by siblings also practicing with the program, and we excluded another 11 students because they had missing data on all relevant variables, such as the questionnaire, test, and practice data. Furthermore, we excluded four classes that withdrew their participation in the study after the pretest ( $n = 48$ ). The  $t$ -tests we conducted revealed that the excluded students did not differ from the included students regarding their pretest math anxiety scores and their pretest math performance. The final analytic sample consisted of 829 students (48.7% female, 49.5% with a migration background).

### 7.2.3 Measures

#### 7.2.3.1 Math Anxiety

We used five items from the Math Anxiety Rating Scale for fourth to sixth grades (MARS 4–6; Roick et al., 2013) to measure math test anxiety (e.g., “How nervous do you feel when a math test is written?”;  $\alpha = .91$ ). We used another six items from MARS to assess math calculation anxiety (e.g., “How nervous do you feel when you have to do some division tasks without a calculator”;  $\alpha = .82$ ). As these items ask about nervousness, they address the affective math anxiety dimension in different contexts. With regard to the third subcomponent, we used three items based on a scale by Faber (1995; e.g., “How much do you worry about the math grade on your next report?”) to assess the worry dimension ( $\alpha = .76$ ). All math anxiety items were assessed on a five-point Likert scale ranging from 1 = *not at all* to 5 = *very much*.

Using the guidelines of Hu and Bentler (1999), we conducted confirmatory factor analyses; the results showed a satisfactory fit for the suggested three-factor structure (i.e., math test anxiety, math calculation anxiety, and worry component; CFI = .95, TLI = .94, RMSEA = .06, SRMR = .05; see Table S7.1 in the Supplementary Material for one- and two-factor model fits). In addition, we conducted chi-square-difference tests, using the guidelines of Cheung and Rensvold (2002), which also indicated that the three-factor model outperformed the one- and two-factor models (see Table S7.1, Supplementary Material).

### 7.2.3.2 Practice Behavior

Math Garden automatically traces the number of tasks completed per domain per week. We calculated two total practice frequency scores for each student by adding all completed math-related tasks within 28 weeks of practice and within 45 weeks of practice. The scores included tasks in the math-related domains of addition, subtraction, multiplication, division, mix (mixture of basic arithmetic operations), counting (of objects), series (logical completion of a given number series), numerals (tasks in which numbers must be combined to obtain a desired target number), and tables (multiplication tables tasks from 1 to 10). 86 students only practiced non-math-related domains, i.e. Math Garden also includes few language-related domains. These students were nevertheless included in our final sample; we coded their number of completed tasks as zero.

### 7.2.3.3 Math Performance

We used the four scales of the Heidelberg calculation test by Haffner et al. (2005) to measure math performance: addition, subtraction, multiplication, and division. Each scale consists of 40 items that become increasingly difficult; students were instructed to solve as many items as possible in 120 seconds per scale. Correctly solved tasks from all scales were added to obtain a total score. We found good retest-reliability values ( $r_{T1-T2} = .74$ ;  $r_{T1-T3} = .72$ ).

### 7.2.3.4 Covariates

Students' gender was coded as 0 = *male* and 1 = *female*. A migration background was assessed with the following item: "How often do you speak German at home?", with a four-point response scale ranging from 1 = *always* to 4 = *never* (Hussmann et al., 2017). The speed scale of the Heidelberg calculation test (Haffner et al., 2005) was used to control for students' tablet typing speed. Furthermore, we controlled for participation in remedial math lessons for low-achieving students who were part of the larger project (0 = *no participation*, 1 = *participation*). As students might have been absent at T<sub>1</sub>, we also controlled for whether they had received a standardized introduction to Math Garden (0 = *no standardized introduction*, 1 = *standardized introduction*). Lastly, we coded whether students had internet access at home through a smartphone, a tablet, a laptop, or a computer (0 = *no device available*, 1 = *at least one device available*).

## 7.2.4 Data Analysis

To test our hypotheses, we set up structural equation models in *Mplus*, Version 8.5 (Muthén & Muthén, 2008–2020) with math anxiety as predictor, the completed tasks in Math Garden as mediator, and performance as outcome, while controlling for pretest values (i.e., investigating

potential performance improvement). This approach was applied for both practice periods. Regarding math anxiety, we combined the items of the three subcomponents into one parcel for each component to obtain a more parsimonious model (Little et al., 2002). We ran the analyses with the latent global score and also investigated the three subcomponents separately; this resulted in four different models for each practice period. We used the `TYPE = COMPLEX` command to adjust the standard errors for the nonindependence of students, who were nested within classes. To ease interpretation, all continuous variables were  $z$ -standardized ( $M = 0$ ,  $SD = 1$ ).

The practice behavior data from Math Garden were log-transformed due to their right-skewed distribution (Van der Maas & Nyamsuren, 2017). The substantial number of students who did not practice any math-related tasks in Math Garden (i.e., 86 students) resulted in a bimodal distribution for practice behavior. To ensure the robustness of the results, we followed two different approaches: First, we ran all models with  $z$ -standardized practice behavior (i.e., right-skewed) data. Second, we treated the raw completed tasks as count data. Therefore, we modeled the relationship between math anxiety and practice behavior under the assumption of a Poisson distribution, using an overdispersion parameter to account for the high variance on the completed tasks variable (Gardner et al., 1995) in the software R (Version 2023.06.1) with cluster bootstrap to account for the hierarchical data (Deen & Rooij, 2020). These supplementary analysis results are provided in the online material (see Tables S7.3–S7.5).

In empirical studies, the presence of missing data is a potential methodological issue. In our study, data were missing on the self-report and test measures because some students in our final analytic sample were absent during at least one of the three testing sessions. The percentage of missing data varied from 0% on students' gender to 14.5% on students' math performance ( $T_2$ ), with an average missing rate of 7.1%. Regarding the practice behavior data, we also found missing data, which means that we could not match all students to their practice data ( $n = 161$ ; 19.4%). We performed statistical comparisons between the groups of students with and without missing practice behavior data regarding their math anxiety level, and we did not find any differences between the two groups ( $t[176] = -1.34$ ,  $p = .189$ ). To handle all missing data, we used the full information maximum likelihood approach (Enders, 2022).

### 7.3 Results

#### 7.3.1 Descriptives and Preliminary Analyses

The means, standard deviations, and correlations for math anxiety, practice behavior, and performance measures and the covariates are reported in Table 7.1. On average, students solved

297 tasks in Math Garden within 28 weeks and 355 tasks within 45 weeks, but practice behavior varied substantially between students ( $SD_{T1-T2} = 600.11$ ,  $SD_{T1-T3} = 713.41$ ,  $\min = 0$ ,  $\max = 9,053$ ). Regarding the relationships between our study variables, we found that higher global math anxiety scores were associated with fewer completed tasks in Math Garden ( $r_{T1-T2} = -.10$ ;  $r_{T1-T3} = -.10$ ). Similar findings emerged for two of the three specific anxiety subcomponents (math test anxiety:  $r_{T1-T2} = -.09$ ,  $r_{T1-T3} = -.09$ ; worry component:  $r_{T1-T2} = -.11$ ,  $r_{T1-T3} = -.10$ ). Interestingly, migration background was also negatively associated with less practice in Math Garden ( $r_{T1-T2} = -.11$ ,  $r_{T1-T3} = -.11$ ). Furthermore, more completed tasks in Math Garden were related to higher performance at T<sub>2</sub> and T<sub>3</sub> ( $.10 \leq r \leq .14$ ). All correlations can be described as being small to moderate (Cohen, 1988).



**Table 7.1***Descriptive Statistics and Correlation Matrix of the Study Variables*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Global math anxiety	2.88	0.88														
2. Math test anxiety	3.19	1.14	<b>.88</b>													
3. Math calculation anxiety	2.50	0.90	<b>.83</b>	<b>.54</b>												
4. Math worry component	3.08	1.14	<b>.78</b>	<b>.62</b>	<b>.49</b>											
5. Completed tasks (T <sub>1</sub> –T <sub>2</sub> )	297.31	600.11	<b>–.10</b>	<b>–.09</b>	–.06	<b>–.11</b>										
6. Completed tasks (T <sub>1</sub> –T <sub>3</sub> )	355.10	713.41	<b>–.10</b>	<b>–.09</b>	–.07	<b>–.10</b>	<b>.99</b>									
7. Math performance T <sub>1</sub>	71.54	21.62	<b>–.24</b>	<b>–.10</b>	<b>–.33</b>	<b>–.18</b>	.03	.04								
8. Math performance T <sub>2</sub>	78.58	21.87	<b>–.27</b>	<b>–.14</b>	<b>–.35</b>	<b>–.20</b>	<b>.10</b>	<b>.11</b>	<b>.74</b>							
9. Math performance T <sub>3</sub>	77.10	24.55	<b>–.25</b>	<b>–.12</b>	<b>–.31</b>	<b>–.20</b>	<b>.14</b>	<b>.14</b>	<b>.72</b>	<b>.78</b>						
10. Gender: female	48.7%		<b>.21</b>	<b>.19</b>	<b>.17</b>	<b>.18</b>	–.04	–.04	<b>–.17</b>	<b>–.09</b>	–.05					
11. Migration background	1.95	0.96	<b>.19</b>	<b>.16</b>	<b>.11</b>	<b>.24</b>	<b>–.11</b>	<b>–.11</b>	–.04	–.03	–.04	<b>.11</b>				
12. Tablet typing speed T <sub>1</sub>	9.91	4.19	.00	.07	<b>–.08</b>	.01	–.07	–.07	<b>.27</b>	<b>.17</b>	<b>.19</b>	<b>.08</b>	–.03			
13. Remedial math lessons <sup>a</sup>	20.0%		<b>.22</b>	<b>.17</b>	<b>.19</b>	<b>.20</b>	<b>–.16</b>	<b>–.16</b>	<b>–.27</b>	<b>–.26</b>	<b>–.24</b>	<b>.14</b>	<b>.11</b>	–.03		
14. Program introduction <sup>b</sup>	93.0%		.03	.04	–.01	.06	.01	–.01	<b>.21</b>	<b>.08</b>	<b>.11</b>	–.01	.04	.06	<b>.13</b>	
15. Availability of devices <sup>c</sup>	74.1%		–.07	–.03	<b>–.08</b>	–.06	.00	–.01	<b>.15</b>	<b>.16</b>	<b>.16</b>	–.01	<b>–.18</b>	<b>.13</b>	.04	<b>.44</b>

*Note.* *N* = 829. Variables measured on a metric scale are presented as mean values; the reference categories of the categorical variables are presented as percentages. Significant correlations are printed in bold ( $p \leq .05$ ). Reference categories: <sup>a</sup>0 = no participation/1 = participation.

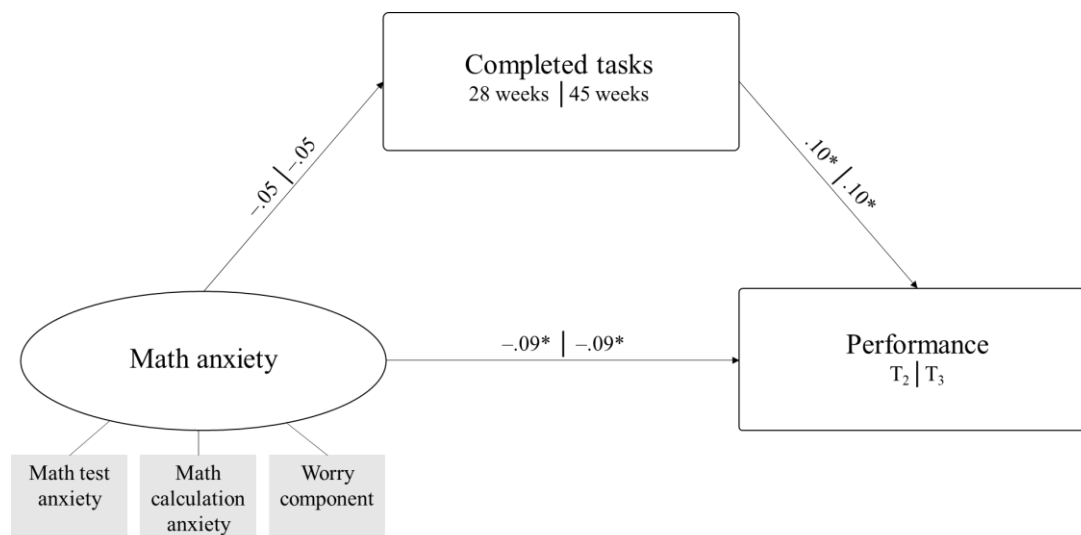
<sup>b</sup>0 = no standardized introduction/1 = standardized introduction. <sup>c</sup>0 = no device available/1 = at least one device available.

### 7.3.2 The Relationship Between Math Anxiety, Avoidance Behavior, and Performance

To answer our research question of whether the avoidance behavior of math anxious students can potentially explain their lower achievement gains, we investigated whether the tasks completed in Math Garden mediated the relationship between math anxious students and their math performance. Considering all of the covariates, we did not find any statistically significant links between global math anxiety or its subcomponents and the number of tasks completed in Math Garden (Hypothesis 1; see Figures 7.2–7.3). However, we found that students with a migration background completed fewer tasks (see Table S7.2 in the Supplementary Material;  $\beta_{T1-T2} = -.14, p = .014$ ;  $\beta_{T1-T3} = -.13, p = .020$ ). Further, we found support for our Hypothesis 2, namely a positive effect of the number of tasks completed in Math Garden on performance in both data collection periods when controlling for students' pretest performance (see Figure 7.2;  $\beta_{T1-T2} = .10, p = .001$ ;  $\beta_{T1-T3} = .10, p = .002$ ). Regarding Hypothesis 3, we found no indirect effect of math anxiety on math performance mediated by the number of completed tasks (see Table 7.2). This was the case both for the analysis of math anxiety as a global score and for the analysis of the subcomponents. Overall, the results seemed to be robust regarding the distribution of the practice behavior variable (see Tables S7.2–S7.5, Supplementary Material).

**Figure 7.2**

*Results of Mediation Analysis Regarding the Global Math Anxiety Score*



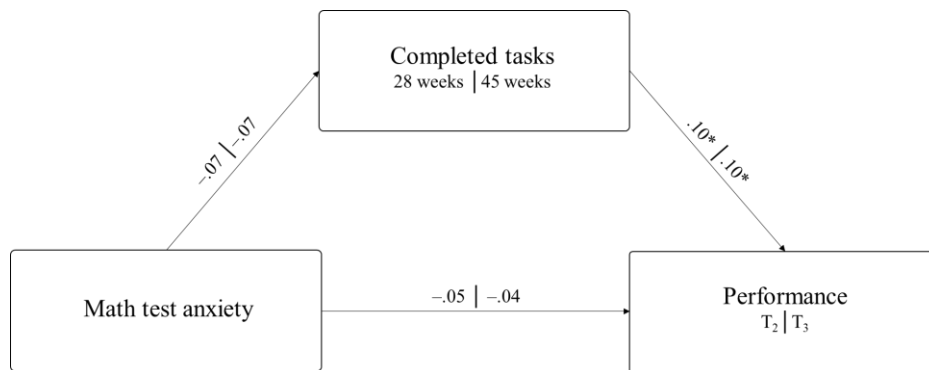
*Note.* Standardized coefficients. All covariates were included: pretest performance, gender, migration background, tablet typing speed, remedial math lessons, program introduction, and availability of devices. T<sub>2</sub>|T<sub>3</sub>=second/third point of measurement.

\* $p \leq .05$ .

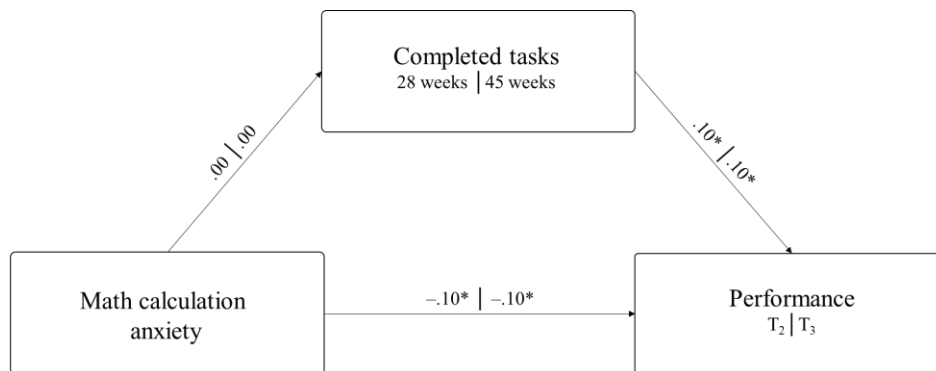
**Figure 7.3**

*Results of Mediation Analysis Regarding Math Test Anxiety, Math Calculation Anxiety, and the Math Worry Component*

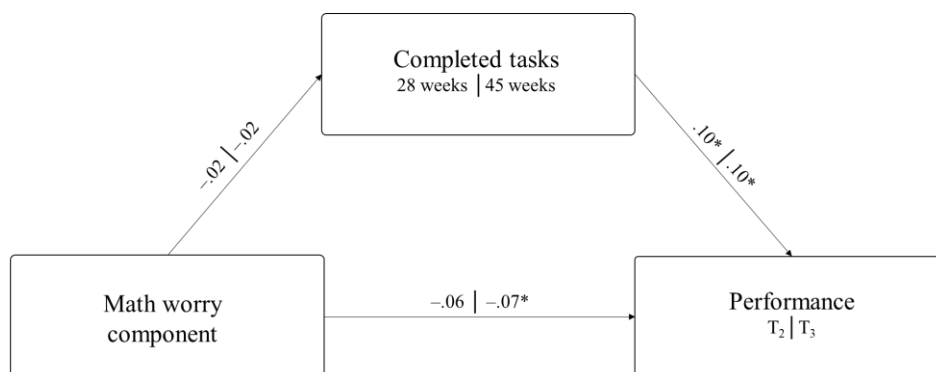
a)



b)



c)



*Note.* Standardized coefficients. All covariates were included: pretest performance, gender, migration background, tablet typing speed, remedial math lessons, program introduction, and availability of devices. T<sub>2</sub>|T<sub>3</sub>=second/third point of measurement.

\* $p \leq .05$ .

**Table 7.2***Total Effects, Indirect Effects, and Direct Effects for all Anxiety Mediation Models*

	Math performance T <sub>2</sub>			Math performance T <sub>3</sub>		
	$\beta$	<i>SE</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>
<b>Global math anxiety</b>						
Total effect	<b>−.09</b>	.04	.024	<b>−.10</b>	.04	.014
Total indirect effect	−.01	.01	.348	−.01	.01	.414
Direct effect	<b>−.09</b>	.04	.027	<b>−.09</b>	.04	.020
<b>Math test anxiety</b>						
Total effect	−.05	.03	.092	−.05	.03	.145
Total indirect effect	−.01	.01	.142	−.01	.01	.230
Direct effect	−.05	.03	.127	−.04	.03	.219
<b>Math calculation anxiety</b>						
Total effect	<b>−.10</b>	.03	.002	<b>−.10</b>	.03	.001
Total indirect effect	−.00	.00	.992	−.00	.00	.986
Direct effect	<b>−.10</b>	.03	.002	<b>−.10</b>	.03	.001
<b>Math worry component</b>						
Total effect	−.06	.04	.077	<b>−.07</b>	.03	.016
Total indirect effect	−.01	.01	.718	−.01	.01	.735
Direct effect	−.06	.03	.077	<b>−.07</b>	.03	.017

*Note.*  $N = 829$ . Coefficients are standardized. Significant coefficients are printed in bold ( $p \leq .05$ ).

## 7.4 Discussion

Avoidance behavior might explain the negative relationship between math anxiety and math performance. By investigating its mediating role in a longitudinal design, this study focused on assessing avoidance behavior with objective process measures within the adaptive arithmetic learning program Math Garden (i.e., by examining the numbers of tasks completed). Furthermore, we operationalized math anxiety as a multidimensional construct. However, the results supported only some of our hypotheses. Students who practiced more with Math Garden showed a higher improvement in their performance (Hypothesis 2). In contrast to our assumptions, math anxious students did not practice less with the program (Hypothesis 1), and practice behavior did not mediate the observed negative math anxiety–performance link (Hypothesis 3).

### 7.4.1 Math Anxiety and Avoidance Behavior

Considering the behavioral aspect of math anxiety, math anxious students are assumed to avoid math-related tasks due to related unpleasant negative arousals (Eysenck & Calvo, 1992; Eysenck et al., 2007). A previous study by Hilz, Guill, Roloff, Aldrup, and Köller (2023) investigating the effects of students' individual characteristics on practice behavior in Math Garden already provided first evidence for this assumption using a process data approach. In the current study, we also found a negative bivariate correlation between math anxiety and the number of tasks completed, but when controlling for other variables, the statistical model revealed that having a migration background was a more powerful predictor of lower levels of practice behavior than math anxiety was. Considering the beneficial effect of practice behavior on math performance, this pattern of results could suggest that math anxious students do have the chance to achieve learning gains from practicing with learning programs such as Math Garden.

However, other circumstances might also explain the difference between the results of our study (i.e., no mediation effect of avoidance behavior) and those of previous research, namely, the period during which the study was conducted. Whereas the study of Hilz, Guill, Roloff, Aldrup, and Köller (2023) was conducted before the COVID-19 pandemic, our study took place after the pandemic. The pandemic led to the promotion of teachers' use of digital tools (Runge et al., 2020) and potentially made their support of students' use of digital learning programs more effective. Such supportive behavior might have the potential to reduce math anxious students' unpleasant feelings in math-related situations (Jansen et al., 2013) and, in turn, might have helped to overcome the inhibited practice behavior hypothesized in our study. Nonetheless, this suggestion needs to be clarified in future studies.

Furthermore, the fact that we did not find a relation between math anxiety and practice behavior might be explained by the beneficial adaptive function in Math Garden. In line with the taxonomy framework of Plass and Pawar (2020) on adaptive learning, Math Garden can diagnose students' skill level and, to accommodate their specific needs, can make adjustments to the task difficulty level in such a way that students can solve at least 75% of the assigned tasks correctly. As a driving factor in anxiety is uncertainty (Buckley & Sullivan, 2023; Pekrun, 2006), math anxious students might perceive adaptive learning programs as more controllable as there is only a low chance of being confronted with unsolvable tasks according to the individual skill level. According to the control-value theory (Pekrun, 2006), this is assumed to lead to reduced anxiety and, thus, more practice. As Hilz, Guill, Roloff, Aldrup, and Köller (2023) also found only a small effect in their study, this could indicate that math anxiety might

only play a small role, or might not even play any role, when it comes to practice behavior with adaptive learning programs. However, such assumptions about the mechanisms of adaptivity and especially about the effects of math anxious students' perception of it need to be tested in future research.

Taken together, the divergent results between our study and that of Hilz, Guill, Roloff, Aldrup, and Köller (2023) suggest that the relationship between math anxiety and avoidance behavior stated in the literature is not trivial and might be influenced by various underlying mechanisms. This is also indicated by the mixed result pattern found in studies using self-reported avoidance behavior. Whereas Quintero et al. (2022) found a negative relationship between math anxious students and their engagement during math instruction, Hasty et al. (2021) did not find any relationship between math anxiety and students' engagement when practicing math after school. The results seem to vary across educational contexts, which could imply that engaging with math outside of school might be favorable for math anxious students as, for instance, social comparisons are omitted (Bekdemir, 2010). In this context, the use of process data also has the potential to comprehensively investigate this implication, as process data are time-stamped. Thus, future research could examine whether math anxious students tend to avoid practicing with a math learning program more in the morning (i.e., during school hours) than in the afternoon (i.e., at home).

### **7.4.2 The Potential of Process Measures**

Different process measures could be included in future studies to further investigate the math anxiety–performance link, as the number of completed tasks might not be the optimal indicator to display students' avoidance strategies. This could also explain the missing effects in our study. Wong and Liem (2022) defined behavioral disengagement as the extent to which students intentionally withdraw effort and give up on learning activities. Thus, quitting could be a useful parameter for measuring avoidance, which is understood on a short timescale, for instance, quitting the learning program and trying again later (Broeke et al., 2022).

Besides behavioral engagement, which our study refers to, theoretical engagement frameworks also consider a cognitive component (Fredricks et al., 2004). According to Wong and Liem (2022), cognitive avoidance is defined as the extent to which students are distracted during learning activities. Thus, future research might also use process measures to test whether math anxious students avoid math-related tasks on a cognitive level. One approach could be to use eye-tracking measures during practice with learning programs (Miller, 2015).

Our results also indicate the importance of taking process-measured engagement into account when examining effects on performance. A problem with rather traditional approaches to investigating the effects of math learning programs is that researchers mostly only focus on comparing the program provision in an experimental group to a waitlist control group (e.g., Bai et al., 2012; Chang et al., 2015; Shin et al., 2012). When the results reveal null effects, this leads many researchers to believe that math learning programs are not suitable for fostering students' math skills (Pellegrini et al., 2021). However, we found that students who completed more tasks with Math Garden showed higher performance improvements, which aligns with previous research on students' engagement within math learning programs (Hassler Hallstedt et al., 2018; Jansen et al., 2013). This beneficial finding therefore illustrates that the missing effects found in previous studies that only investigated the program provision might have occurred because students just practiced too little. Based on this, future research should continue to investigate the learning environment and the person-related circumstances under which students engage with math learning programs, especially in the long term, as it is already evident that engagement often decreases drastically a few weeks after the implementation (Spitzer, 2022).

### 7.4.3 Limitations

Even though our study provides new insights into the mediating role of avoidance behavior in the math anxiety–math performance link, it has limitations that need to be considered when interpreting the results. Firstly, our results are limited to students' practice behavior in an adaptive arithmetic learning program; they cannot be applied to other learning environments. Due to Math Garden's adaptive nature, math anxious students might have overcome their assumed lower engagement, but other nonadaptive learning environments might continue to be avoided due to less perceived control and more persistent failures (Pekrun, 2006). As we did not intend to manipulate the algorithm responsible for providing adaptive tasks in Math Garden, future studies might experimentally vary task difficulties when examining practice effects on math anxious students' engagement. To follow this suggestion, however, it would be necessary to assess students' perception of task difficulty, for example, with an experience sampling approach, in order to understand the underlying mechanisms of math anxiety, adaptivity, and students' engagement.

In addition, Math Garden is limited to fast mental arithmetic tasks, which are less complex. But, as more feelings of math anxiety are experienced when completing tasks at higher complexity levels (Barroso et al., 2021), programs that challenge students with word

problems might continue to be avoided. Thus, our assumptions need further investigation regarding programs with more complex demands.

Regarding our sample and the study design, two further limitations need to be considered. Firstly, our sample consisted, on average, of lower-achieving students. Therefore, the results cannot be generalized to all achievement levels and need further investigation. Secondly, teacher and parent characteristics might have affected our outcomes, especially students' practice behavior with Math Garden. Even though we considered differences between classes in our analytic approach, the teachers might have differed in the support they gave the students while the students practiced with Math Garden (Eickelmann et al., 2019). The same point might apply to parental support at home (Moroni et al., 2015). This could also explain the missing effects of math anxiety on practice behavior, as varying contextual support rather than anxiety would then have been the more relevant variable; future research should consider such variables.

### **7.4.4 Implications**

When implementing supportive digital tools for math instruction, there is a pressing need for practitioners to understand how various student characteristics are associated with practice behavior in order to encourage those students at risk of practicing less. Even though our results do not provide evidence that math anxious students practice less, this does not necessarily mean that math anxious students are not at risk, and we believe that such students might well need further support due to their lower achievement gains. Moreover, our results suggest that teachers need to especially encourage students with a migration background in order for them to have the same chances to benefit from the implementation of digital tools.

The results of the present study also indicate that the effects of digital learning programs can vary according to how much students engage with them. Thus, teachers are not only responsible for the successful implementation of a program that provides adaptive support; they might also play an important role by supporting their students in persistent practicing. In line with the positive effect of teacher training programs on implementing digital tools (Hillmayr et al., 2020), we therefore recommend that such training programs should highlight how teachers can monitor students' practice. For instance, the teachers' backend of the Math Garden program offers them a very detailed overview of which domains their students are practicing in and how many tasks they have completed.



### 7.4.5 Conclusions

Although the negative relationship between math anxiety and math performance is highly evident in the literature, the mediating role of avoidance behavior remains unclear at this point. By investigating students' avoidance behavior in terms of the number of tasks completed (i.e., by using process measures) in an adaptive arithmetic learning program over a period of 45 weeks, this study adds a new perspective to the current state of research. Even though we did not find evidence that avoidance behavior can explain the negative relationship between math anxiety and math performance, we do not believe that this hypothesis should be rejected as strong theoretical assumptions indicate that it is a possibility (Eysenck & Calvo, 1992; Eysenck et al., 2007). Thus, more research is needed to identify the chain of effects behind the use of (adaptive) learning technologies by math anxious students. Moreover, our results suggest that varying student characteristics are associated with students' practice, for instance, students with a migration background need additional encouragement because they are at risk of practicing less. Similarly, our findings indicate the importance of regular student practice with learning programs because, without consistent practice, students may not reap the benefits of the implementation. Consequently, further research should aim to identify the factors that lead to persistent and regular practice with such learning programs.

### 7.5 References

- Aldrup, K., Klusmann, U., & Lüdtke, O. (2020). Reciprocal associations between students' mathematics anxiety and achievement: Can teacher sensitivity make a difference? *Journal of Educational Psychology, 112*(4), 735–750. <https://doi.org/10.1037/edu0000398>
- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology, 102*(1), 115–134. <https://doi.org/10.1037/a0016838>
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology. General, 130*(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248. <https://doi.org/10.3758/BF03194059>

- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology*, 43(6), 993–1003. <https://doi.org/10.1111/j.1467-8535.2011.01269.x>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. *Educational Studies in Mathematics*, 75(3), 311–328. <https://doi.org/10.1007/s10649-010-9260-7>
- Bråten, I., Latini, N. & Haverkamp, Y.E. (2022). Predictors and outcomes of behavioral engagement in the context of text comprehension: when quantity means quality. *Reading and Writing*, 35(3), 687–711. <https://doi.org/10.1007/s11145-021-10205-x>
- Broeke, N., Hofman, A. D., Kruis, J., Mooij, S. M. de, & Van der Maas, H. (2022). *Predicting and Reducing Quitting in Online Learning*. <https://doi.org/10.31219/osf.io/htzvm>
- Buckley, S., & Sullivan, P. (2023). Reframing anxiety and uncertainty in the mathematics classroom. *Mathematics Education Research Journal*, 35(S1), 157–170. <https://doi.org/10.1007/s13394-021-00393-8>
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2015). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, 6, 1987. <https://doi.org/10.3389/fpsyg.2015.01987>
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology*, 8, 11. <https://doi.org/10.3389/fpsyg.2017.00011>
- Carstensen, B., Hilz, A., Benckwitz, L., Asbury, K., Roloff, J., & Guill, K. (2023). Intervention in mathematics to professionally support low-achieving students. Retrieved from [osf.io/kdbcr](https://doi.org/10.17605/OSF.IO/KDBCR). Advance online publication. <https://doi.org/10.17605/OSF.IO/KDBCR>
- Caviola, S., Primi, C., Chiesi, F., & Mammarella, I. C. (2017). Psychometric properties of the Abbreviated Math Anxiety Scale (AMAS) in Italian primary school children. *Learning and Individual Differences*, 55, 174–182. <https://doi.org/10.1016/j.lindif.2017.03.006>

- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International*, 52(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203771587>
- Deen, M., & Rooij, M. de (2020). Clusterbootstrap: An R package for the analysis of hierarchical data using generalized linear models with the cluster bootstrap. *Behavior Research Methods*, 52(2), 572–590. <https://doi.org/10.3758/s13428-019-01252-y>
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2), 245–260. <https://doi.org/10.1023/A:1019637632584>
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Eds.). (2019). *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking [ICILS 2018 #Germany: Computer and information-related competencies of students in the second international comparison and competencies in computational thinking]*. Waxmann.
- Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). The Guilford Press.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition & Emotion*, 6(6), 409–434. <https://doi.org/10.1080/02699939208409696>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Faber, G. (1995). Die Diagnose von Leistungsangst vor schulischen Rechtschreibsituationen: Neue Ergebnisse zu den psychometrischen Eigenschaften und zur Validität einer

- entsprechenden Kurzskala [The diagnosis of anxiety in spelling situations: New results on the psychometric properties and validity of a corresponding short scale]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 44, 110–119.  
<https://doi.org/10.25656/01:880>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.  
<https://doi.org/10.3102/00346543074001059>
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392–404. <https://doi.org/10.1037/0033-2909.118.3.392>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4: Heidelberger Rechentest – Erfassung mathematischer Basiskompetenzen im Grundschulalter [HRT 1-4: Heidelberg Arithmetic test - Assessment of basic mathematical skills at primary school age]*. Hogrefe.
- Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology*, 110(8), 1127–1148. <https://doi.org/10.1037/edu0000264>
- Hasty, L. M., Malanchini, M [Margherita], Shakeshaft, N., Schofield, K., Malanchini, M [Maddalena], & Wang, Z. (2021). When anxiety becomes my propeller: Mental toughness moderates the relation between academic anxiety and academic avoidance. *The British Journal of Educational Psychology*, 91(1), 368–390.  
<https://doi.org/10.1111/bjep.12366>
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33. <https://doi.org/10.2307/749455>
- Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, 48, 45–53. <https://doi.org/10.1016/j.lindif.2016.02.006>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.  
<https://doi.org/10.1016/j.compedu.2020.103897>
- Hilz, A., Guill, K., Roloff, J., Aldrup, K., & Köller, O. (2023). The relationship between individual characteristics and practice behaviour within an adaptive arithmetic

- learning program. *Journal of Computer Assisted Learning*, 39(3), 970–983.  
<https://doi.org/10.1111/jcal.12780>
- Hilz, A., Guill, K., Roloff, J., Sommerhoff, D., & Aldrup, K. (2023). How to continue? New approaches to investigating the effects of adaptive math learning programs on students' performance, self-concept, and anxiety. *Journal of Intelligence*, 11(6), 108.  
<https://doi.org/10.3390/jintelligence11060108>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hussmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C., & Valtin, R. (Eds.). (2017). *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [IGLU 2016: Reading competence of elementary students in Germany in international comparison]*. Waxmann.
- Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, 24, 190–197. <https://doi.org/10.1016/j.lindif.2012.12.014>
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824.  
<https://doi.org/10.1016/j.compedu.2011.02.003>
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- LeFevre, J.-A., Kulak, A. G., & Heymans, S. L. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, 24(3), 276–289.  
<https://doi.org/10.1037/h0078742>
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173.  
[https://doi.org/10.1207/S15328007SEM0902\\_1](https://doi.org/10.1207/S15328007SEM0902_1)

- Lukowski, S. L., DiTrapani, J. B., Jeon, M., Wang, Z., Schenker, V. J., Doran, M. M., Hart, S. A., Mazzocco, M. M. M., Willcutt, E. G., Thompson, L. A., & Petrill, S. A. (2019). Multidimensionality in the measurement of math-specific anxiety and its relationship with mathematical performance. *Learning and Individual Differences, 70*, 228–235. <https://doi.org/10.1016/j.lindif.2016.07.007>
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520–540. <https://doi.org/10.2307/749772>
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. <https://doi.org/10.1111/j.1750-8606.2008.00048.x>
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60–70. <https://doi.org/10.1037/0022-0663.82.1.60>
- Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist, 50*(1), 31–42. <https://doi.org/10.1080/00461520.2015.1004068>
- Moroni, S., Dumont, H., Trautwein, U., Niggli, A., & Baeriswyl, F. (2015). The need to distinguish between quantity and quality in research on parental involvement: The example of parental help with homework. *The Journal of Educational Research, 108*(5), 417–431. <https://doi.org/10.1080/00220671.2014.901283>
- Muenks, K., Wigfield, A., Yang, J. S., & O'Neal, C. R. (2017). How true is grit? Assessing its relations to high school and college students' personality characteristics, self-regulation, engagement, and achievement. *Journal of Educational Psychology, 109*(5), 599–620. <https://doi.org/10.1037/edu0000153>
- Muthén, B. O., & Muthén, L. K. (2008–2020). *Mplus computer software (Version 8.5)*. Muthén & Muthén.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>

- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7.  
<https://doi.org/10.1177/2332858420986211>
- Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom Assessment Scoring System (CLASS)*. Teachstone.
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52(3), 275–300.  
<https://doi.org/10.1080/15391523.2020.1719943>
- Quintero, M., Hasty, L., Li, T., Song, S., & Wang, Z. (2022). A multidimensional examination of math anxiety and engagement on math achievement. *The British Journal of Educational Psychology*, 92(3), 955–973.  
<https://doi.org/10.1111/bjep.12482>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, 14(2), 187–202. <https://doi.org/10.1080/15248372.2012.664593>
- Roick, T., Gölit, D., & Hasselhorn, M. (2013). Affektive Komponenten der Mathematikkompetenz. Die Mathematik-Angst-Ratingskala für vierte bis sechste Klassen [Affective components of math competence. The Mathematics Anxiety Rating Scale for fourth to sixth grades] (MARS 4-6). In M. Hasselhorn, A. Heinze, W. Schneider, & U. Trautwein (Eds.), *Diagnostik mathematischer Kompetenzen* (pp. 205–219). Hogrefe.
- Runge, I., Rubach, C. & Lazarides, R. (2020). Digitale Kompetenzselbsteinschätzung von Lehrkräften – Welche Bedeutung haben Schulausstattung und Fortbildungsteilnahme angesichts aktueller Herausforderungen [Digital competence self-assessment of teachers – What significance do school equipment and training participation have in view of current challenges]? In Huber, S. G. (Eds.), *Jahrbuch Schulleitung 2021*. Carl Link.
- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540–560. <https://doi.org/10.1111/j.1467-8535.2011.01197.x>
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection. *Educational and Psychological Measurement*, 69(3), 493–525. <https://doi.org/10.1177/0013164408323233>

- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *The British Journal of Educational Psychology*, 87(3), 309–327. <https://doi.org/10.1111/bjep.12151>
- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Van der Maas, H. L. J., & Nyamsuren, E. (2017). Cognitive analysis of educational games: The number game. *Topics in Cognitive Science*, 9(2), 395–412. <https://doi.org/10.1111/tops.12231>
- Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80(2), 210–216. <https://doi.org/10.1037/0022-0663.80.2.210>
- Wong, Z. Y., & Liem, G. A. D. (2022). Student engagement: Current state of the construct, conceptual refinement, and future research directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09>



## 7.6 Supplementary Tables

**Table S7.1***Fit Indices and Model Comparisons for Confirmatory Factor Analyses*

	Fit indices							Model comparisons	
	$\chi^2$	<i>df</i>	SCF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	$p_{\text{min-max}}$
One-factor model	875.02	77	1.24	.79	.75	.12	.09		
Two-factor model									
<i>Affective/cognitive</i>	772.12	76	1.23	.81	.78	.12	.09	73.98 <sup>a</sup>	.000
<i>Calculation/test</i>	395.85	75	1.21	.91	.90	.08	.06	203.02 <sup>a</sup>	172.43 <sup>b</sup>
Three-factor model	285.10	74	1.21	.95	.94	.06	.05	388.94 <sup>a</sup>	306.97 <sup>b</sup> 110.75 <sup>c</sup>

*Note.* SCF = scaling correction factor; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual. Reference model: <sup>a</sup>one-factor model. <sup>b</sup>two-factor affective/cognitive model. <sup>c</sup>two-factor calculation/test model.

**Table S7.2***Model Results Regarding Covariates for Both Practice Periods*

	Completed tasks				Math performance			
	$\beta_{\min}$	$\beta_{\max}$	$SE_{\min-\max}$	$p_{\min-\max}$	$\beta_{\min}$	$\beta_{\max}$	$SE_{\min-\max}$	$p_{\min-\max}$
Global math anxiety		-.05	.06	.370-.371		<b>-.09</b>	.04	.020-.027
Math test anxiety		-.07	.05	.144-.145	-.04	-.05	.03	.127-.219
Math calculation anxiety		.00	.04	.986-.992		<b>-.10</b>	.03	.001-.002
Math worry component		-.02	.05-.06	.722-.728	-.06	<b>-.07</b>	.03	.017-.077
Completed tasks						<b>.10</b>	.03	.001-.003
Math performance T <sub>1</sub>	.01	.02	.07	.749-.941	<b>.87</b>	<b>.93</b>	.06-.13	.000
Gender <sup>a</sup>	-.04	-.05	.04	.179-.293	.05	<b>.09</b>	.02-.03	.001-.052
Migration background	<b>-.13</b>	<b>-.14</b>	.05-.06	.014-.020	.01	<b>.05</b>	.02-.03	.049-.834
Tablet typing speed T <sub>1</sub>	.02	.04	.07	.683-.762	-.01	-.05	.03	.076-.847
Remedial math lessons <sup>b</sup>	<b>-.17</b>	<b>-.18</b>	.05-.06	.001-.002	-.04	<b>-.05</b>	.02-.03	.034-.201
Program introduction <sup>c</sup>	-.01	-.04	.06-.07	.565-.866	<b>-.43</b>	<b>-.46</b>	.06-.19	.000-.025
Availability of devices <sup>d</sup>		-.04	.05	.395-.427	.03	.04	.03-.04	.264-.334
$R^2$			.07				.55-.59	

*Note.* Results are presented as ranges. Coefficients are standardized. Significant coefficients are printed in bold ( $p \leq .05$ ). Reference categories: <sup>a</sup>male, <sup>b</sup>no participation, <sup>c</sup>no introduction, <sup>d</sup>no device available.

**Table S7.3***Results Regarding all Math Anxiety Models Without Log-Transformed Practice Data for Both Practice Periods*

	Completed tasks				Math performance			
	$\beta_{\min}$	$\beta_{\max}$	$SE_{\min-\max}$	$p_{\min-\max}$	$\beta_{\min}$	$\beta_{\max}$	$SE_{\min-\max}$	$p_{\min-\max}$
Global math anxiety		-.07	.04	.107-.118		<b>-.09</b>	.04	.018-.025
Math test anxiety		-.04	.03	.147-.153		-.05	.03	.100-.176
Math calculation anxiety		-.03	.02	.153-.156		<b>-.09</b>	.03	.002
Math worry component		-.06	.05	.224-.286		<b>-.07</b>	.03	.019-.075
Completed tasks								
Math performance T <sub>1</sub>		-.01	.05-.06	.861-.983		<b>.12</b>	.02-.03	.000-.096
Gender <sup>a</sup>		.00	.03	.861-.985		<b>.87</b>	.06-.12	.000
Migration background		<b>-.08</b>	.03	.002-.009		.04	.03	.002-.106
Tablet typing speed T <sub>1</sub>		-.07	.05-.06	.138-.206		.04	.02-.03	.082-.913
Remedial math lessons <sup>b</sup>		<b>-.14</b>	.03	.000		-.04	.02-.03	.021-.135
Program introduction <sup>c</sup>		.07	.06-.07	.117-.294		<b>-.44</b>	.06-.19	.000-.022
Availability of devices <sup>d</sup>		-.01	.05	.826-.878		.04	.03-.04	.308-.426
$R^2$			.04				.55-.59	

*Note.* Results are presented as ranges. Coefficients are standardized. Significant coefficients are printed in bold ( $p \leq .05$ ). Reference categories: <sup>a</sup>male, <sup>b</sup>no participation, <sup>c</sup>no introduction, <sup>d</sup>no device available.

**Table S7.4**

*Total Effects, Indirect Effects, and Direct Effects for Both Practice Periods Without Log-Transformed Practice Data*

	Math performance			
	$\beta_{\min}$	$\beta_{\max}$	$SE_{\min-\max}$	$p_{\min-\max}$
<b>Global math anxiety</b>				
Total effect		<b>-.10</b>	.04	.011-.022
Total indirect effect	-.00	-.01	.00-.01	.125-.156
Direct effect		<b>-.09</b>	.04	.018-.025
<b>Math test anxiety</b>				
Total effect		-.05	.03	.090-.137
Total indirect effect		-.00	.00	.180-.190
Direct effect	-.04	-.05	.03	.176-.100
<b>Math calculation anxiety</b>				
Total effect		<b>-.10</b>	.03	.001-.002
Total indirect effect		-.00	.00	.141-.180
Direct effect	<b>-.09</b>	<b>-.10</b>	.03	.002
<b>Math worry component</b>				
Total effect	-.06	<b>-.08</b>	.03-.04	.011-.064
Total indirect effect	.00	-.01	.01	.224-.316
Direct effect	-.06	<b>-.07</b>	.03	.019-.075

*Note.* Results are presented as ranges. Coefficients are standardized. Significant coefficients are printed in bold ( $p \leq .05$ ).

**Table S7.5**

*Relationship Between Math Anxiety and Practice Behavior for Both Practice Periods Modeled with Poission Distribution*

	Completed tasks				
	$b_{\min}$	$b_{\max}$	$SE_{\min-\max}$	95% $CI_{\min}$	95% $CI_{\max}$
Global math anxiety	-35.32	-43.38	24.97–30.91	-133.90	0.41
Math test anxiety	-23.69	-29.08	18.84–23.29	-87.45	7.12
Math calculation anxiety	-26.36	-35.27	18.31–23.00	-100.33	1.34
Math worry component	-43.99	-49.87	39.44–48.07	-210.91	10.85
Math performance T <sub>1</sub>	2.23	9.82	29.16–37.35	-82.90	70.21
Gender <sup>a</sup>	0.53	5.96	37.26–45.51	-75.65	109.65
Migration background	<b>-52.14</b>	<b>-62.46</b>	21.39–27.15	-131.30	-16.53
Tablet typing speed T <sub>1</sub>	-44.63	-51.62	39.03–46.47	-197.86	16.28
Remedial math lessons <sup>b</sup>	<b>-208.33</b>	<b>-242.41</b>	71.65–87.66	-493.60	-101.37
Program introduction <sup>c</sup>	<b>369.70</b>	<b>440.85</b>	127.80–154.77	186.52	873.92
Availability of devices <sup>d</sup>	-31.09	-42.85	73.22–91.90	-250.81	117.43

*Note.* Results are presented as ranges. CI = confidence interval. Significant coefficients are printed in bold ( $p \leq .05$ ; zero not included within confidence interval). Reference categories:

<sup>a</sup>male. <sup>b</sup>no participation. <sup>c</sup>no introduction. <sup>d</sup>no device available.



# 8

## GENERAL DISCUSSION

### 8 General Discussion

The use of math learning programs might be a possible way to address the recent challenges presented by the worrying declines in students' math competence and the growing achievement gap (OECD, 2023; Stanat et al., 2022). Thus, based on an adapted supply-use model (Helmke & Weinert, 1997; Seidel, 2014), the present dissertation aimed to investigate the important role that students' practice behavior might play when such supplementary digital supply structures are offered to students.

For that reason, the computer-based adaptive arithmetic learning program Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014) was implemented over two substantially long periods, in both cases as part of a randomized pretest-posttest control group design in samples of fifth-grade students. The first implementation period included 22 weeks in the 2019/2020 school year; in the second implementation period, students practiced for 45 weeks in the 2021/2022 school year. In both implementation periods, researchers did not supervise students' practice, but teachers were instructed to integrate the program into their math lessons in a way that best suited them. Therefore, a huge strength of this dissertation is the high ecological validity of the program implementation.

Based on the longitudinal data, three empirical studies and two additional replication analyses were conducted. Study 1 and Study 2 refer to the first implementation period, whereas Study 3 is based on the data collected within the second implementation period. To investigate the robustness of the findings, Study 1 and Study 2 were replicated using the data from the second implementation period. In the three studies and the two replication analyses, the focus was on students' practice behavior, which can be understood as behavioral engagement. It was operationalized with objective trace data that were stored during students' practice with Math Garden. This is an underrepresented but sorely needed approach within educational research as it enables a deeper understanding than that provided by the common self-reported engagement measures (Azevedo, 2015).

To examine students' practice behavior, two overarching research questions were addressed within this dissertation. First, drawing on several established psychological theories (i.e., the cognitive load theory, the trait activation theory, the situated expectancy-value theory, and the attentional control theory; Eccles & Wigfield, 2020; Eysenck et al., 2007; Sweller et al., 1998; Tett et al., 2013), Study 1 and its replication as well as parts of Study 3 examined how students' individual characteristics, namely, prior math knowledge, conscientiousness, math self-concept, math anxiety, socioeconomic status (SES), and a migration background,



were linked to the number of tasks completed in Math Garden. The results could help to identify certain groups of students who are at risk of practicing less and who, consequently, need more initial support in order to have the chance of benefitting from math learning program implementations. Indeed, Study 1 revealed that especially math anxious students, on average, practiced less with Math Garden. However, the findings from the replication of Study 1 and the results of Study 3 showed that, although math anxiety was negatively associated with the number of tasks completed in Math Garden, students with a migration background practiced less on average. Regarding prior math knowledge, conscientiousness, math self-concept, and SES, the links found in both Study 1 and its replication were insignificant.

The second research question of this dissertation, which was based on the cognitive theory of multimedia learning, the self-determination theory, and the control-value theory (Deci et al., 1991; Mayer, 2014; Pekrun, 2006), examined the effects of Math Garden on students' learning outcomes. Hereby, it considered students' (addition and subtraction) performance, their math self-concept, and their math anxiety. In addition to testing the effects of the program provision (Study 2 and its replication analysis), students' practice behavior was also taken into account (Study 2 and its replication analysis for all outcomes, Study 3 for performance). This is a huge strength as most studies so far have neglected students' actual practice behavior with math learning program. However, one could assume that practice behavior might be a decisive factor in whether or not an implementation works. Indeed, in Study 2, there were no statistically significant effects of the program provision on students' addition or subtraction performance in the experimental group compared to the waitlist control group. However, the more subtraction tasks students in the experimental group completed, the higher their subtraction performance after the intervention was. This beneficial finding was replicated in the additional analyses regarding Study 2. Furthermore, the results of Study 3 also supported this assumption, that is, the number of math-related tasks students completed was positively related to their performance improvement, whereby performance was measured using a composite score of the domains of addition, subtraction, multiplication, and division. Regarding an assumed increase in students' math self-concept, the results of Study 2 revealed a positive effect of the program provision in the experimental group compared to the waitlist control group. However, in contrast to expectations, practice behavior in the experimental group was not related to this increase. This positive effect of the program provision on students' math self-concept was, however, not replicated in the additional analysis. With regard to math anxiety, neither the program provision nor students' practice behavior had any effect in Study 2 or its replication.

In the following sections, the results of the empirical studies and the additional replication analyses are discussed in relation to the two research questions (Section 8.1 and Section 8.2). After that, the theoretical and practical implications are derived in Section 8.3, before the limitations are discussed in Section 8.4. Finally, an outlook for future research is given in Section 8.5, and the dissertation closes with the general conclusions in Section 8.6.

### **8.1 RQ1: The Associations Between Individual Student Characteristics and Students' Practice Behavior With a Math Learning Program**

On the basis of the adapted supply-use model (Helmke & Weinert, 1997; Seidel, 2014) as well as established psychological theories on students' behavioral engagement (i.e., the cognitive load theory, the trait activation theory, the situated expectancy-value theory, and the attentional control theory; Eccles & Wigfield, 2020; Eysenck et al., 2007; Sweller et al., 1998; Tett et al., 2013), Study 1 and the additional replication analysis regarding Study 1 investigated how students' prior math knowledge, conscientiousness, math self-concept, math anxiety, SES, and a migration background were linked to students' practice behavior in the Math Garden program. Study 3 contributed to this research question by focusing more closely on students' math anxiety and examined how the affective dimensions of math test anxiety and math calculation anxiety, as well as the cognitive worry dimension, were related to students' practice behavior. Within all approaches, students' practice behavior was operationalized in terms of the number of tasks completed within the respective implementation period, which varied between 22 weeks in Study 1 and 45 weeks in Study 3.

Study 1 and its replication consistently showed that most variation in students' practice behavior could be explained across time. This means that students' practice varied mostly on an intraindividual level over the respective implementation periods. However, there was still a substantial amount of variation that could be attributed to differences between students. Such interindividual differences in practice might be explained by the individual student characteristics examined.

#### **8.1.1 Math Anxiety and Students with a Migration Background**

Altogether, the results showed that math anxiety and having a migration background in particular appeared to hinder students from practicing. However, the result pattern varied across the implementation periods. In the first implementation period (Study 1), math anxious students practiced less. Hereby, the items specifically measured math test anxiety. In the second implementation period (the replication of Study 1 and Study 3), a negative bivariate correlation

was found between a global math anxiety score and the two subcomponents of math test anxiety and the math worry component and the number of completed tasks. However, the correlations found were descriptively smaller than those found in the first implementation period. Furthermore, migration background was also negatively correlated with students' practice behavior. It turned out that, according to the statistical models of these investigations, having a migration background was the more powerful negative predictor of practice behavior, whereas math anxiety and its individual subcomponents turned out to be insignificant predictors.

This varying pattern of findings might be due to the interplay of two explanations, namely, (1) sample characteristics with regard to students' math anxiety level, and (2) the assessment of students' migration background. First, it is noticeable that students had on average descriptively less math anxiety in the second implementation period. Of the subcomponents, math test anxiety had the highest average mean value in the second implementation period (the replication of Study 1 and Study 3); however, it was lower than the average mean value for math text anxiety in the first implementation period (Study 1). The average mean value for the worry component was even lower, and the average mean value for math calculation anxiety was, descriptively, the lowest in the second implementation period (Study 3). These overall lower math anxiety levels might have resulted in the small and only partially significant bivariate relationships with students' practice behavior. Taken together, due to their, on average, lower math anxiety level, students from the second implementation period might have experienced fewer unpleasant negative feelings with regard to the tasks in Math Garden, and this might have led to less avoidance behavior (Eysenck et al., 2007).

Second, the assessment of students' migration background differed between the two implementation periods. In the first implementation period, students were asked whether they spoke other languages besides German at home, with answer options *yes* or *no*. In the second implementation period, students were asked how often they spoke German at home and this question had a four-level response format. Hence, due to the broader answer format, more substantial, meaningful variance might have been captured, which, in turn, might have explained more variance in students' practice. This might also be indicated by the more complex correlation pattern. In Study 1, a migration background was not correlated with students' math anxiety. However, in the replication of Study 1 and in Study 3, significant correlations were found between having a migration background and math anxiety ( $r = .19$ ). This significant link between the two predictor variables could lead to the assumption of a suppressor effect, but as both variables also showed very similar moderate correlations with the criteria (i.e., the number of completed tasks), the validity of this assumption is rather unlikely

(Conger, 1974). Overall, the pattern of results is therefore comprehensible and might suggest that when students experience less math anxiety, a (continuously measured) migration background might be the more relevant determinant of students' practice behavior.

When interpreting this pattern of findings, at least three suggestions can be made. First, although only the results of Study 1 provided evidence that math anxious students practiced less, it can hardly be argued that math anxious students are not at risk. This is suggested by the negative relationship between math achievement and math anxiety stated within the literature (Aldrup et al., 2020; Barroso et al., 2021; Carey et al., 2015; Ramirez et al., 2018) and by the lower achievement gains found for math anxious students in Study 3 of the dissertation. Thus, math anxious students might still need further support when practicing with a math learning program, as they might experience less control over their skills and resources (Pekrun, 2006). Specific support strategies will be presented in Section 8.3.

Second, when interpreting the effect of students' migration background on their practice behavior, two underlying mechanisms might explain this negative relationship. On the one hand, previous research has already shown that students with a migration background perceive their parents to be less responsive when it comes to homework and less interested in school activities (Dumont et al., 2014). A lack of parental support at home could therefore be one reason for the lower levels of practice observed in students with a migration background. On the other hand, it might be important to consider that students' migration background was identified based on the language(s) spoken at home. According to Portes and Rumbaut (2001), a language barrier makes it particularly difficult for students with a migration background to get involved in learning processes. However, it might be important to note that tasks in Math Garden are less language-intensive than tasks in many other math learning programs, such as the word problem-solving tasks in the Bettermarks program. Therefore, it is possible that students with a migration background might be even more disadvantaged in more language-intensive programs. In other words, this suggested interplay of less support at home and a language barrier is indeed unfavorable and might intensify social disparities. To counteract such widening disparities, Section 8.3 will discuss in more detail how students with a migration background might be supported when practicing with math learning programs in order to achieve a similarly high level of practice as their peers without a migration background.

Third, it is also important to keep in mind that assessing a migration background according to the language(s) spoken at home is certainly one but is not the only indicator of a migration background (Kast et al., 2021; Konrad-Ristau & Burghardt, 2021). Hence, future studies might consider the duration of a family's residence in the respective country and

whether family members were born abroad. This information could make it possible to further differentiate between first- and second-generation immigrants; previous research has already pointed out that first-generation immigrants face greater disadvantages (OECD, 2019; Wendt et al., 2020). This might also apply to practice with a math learning program.

### **8.1.2 Prior Knowledge, Conscientiousness, Math Self-Concept, and SES**

Although theoretical assumptions and empirical evidence suggest positive relationships between higher prior math knowledge, higher levels of conscientiousness, a higher math self-concept, a higher SES, and students' practice behavior (e.g., see Dong et al., 2020; Tomaszewski et al., 2020; Trautwein & Lüdtke, 2007; Trautwein et al., 2009), no relationship was found in any of the investigations conducted within this dissertation. This robust pattern of results, obtained in two different samples, might lead to the cautious suggestion that there are no associations between the individual student characteristics investigated and students' practice behavior with a math learning program. However, before these results can be generalized, further investigations are definitely needed; such investigations could comprise different sample characteristics (e.g., different age groups, other school tracks) and implement different math learning programs, as certain program characteristics of Math Garden might explain the missing effects (see Section 8.4 for the limitations of the studies conducted for this dissertation). In addition, further alternative explanations cannot be ruled out for the single predictors.

The tasks in Math Garden are limited to fast mental arithmetic tasks, which have less complex demands than, for instance, word problem-solving tasks. Therefore, although the program adapts to different ability levels, students with higher prior math knowledge and a higher math self-concept might not have found the program more challenging or valuable than students with lower levels of prior math knowledge or a lower math self-concept. Thus, contrary to expectations, they may not have practiced more. Conversely, however, this result is also promising with regard to the question of whether there are at-risk groups who practice less and, therefore, may not benefit from the implementation of math learning programs. As especially low-achieving students have low prior knowledge and mostly unfavorable motivational prerequisites (Chiu & Klassen, 2010; Van der Beek et al., 2017), they are most dependent on the use of supportive learning opportunities. As students with these characteristics at least did not avoid practicing with Math Garden, the implementation of the program did not seem to disadvantage them further.

Regarding the findings for conscientiousness, the characteristics of the sample used in this dissertation could be a reason for the lack of effects found, for instance, the younger age group. In the studies by Trautwein and colleagues, which investigated the effects of conscientiousness on students' homework engagement (Trautwein & Lüdtke, 2007; Trautwein et al., 2006), the samples comprised older students. As research has already shown that parental engagement is particularly high among younger children (Hoover-Dempsey & Sandler, 1997; Silinskas & Kikas, 2019), parents of less conscientious children could have tried to compensate for their lower practice behavior by constantly reminding them to use the learning program. Therefore, parental engagement might explain more variance in practice behavior than conscientiousness in younger children. In this light, the way in which teachers integrated Math Garden into math lessons could also have been more decisive than the level of students' conscientiousness. In a well-structured learning environment, the assumed beneficial characteristics of more conscientious students, such as higher self-regulation and self-control (Poropat, 2009), might simply not come to bear. However, this assumption needs to be tested in further studies.

Regarding students' SES, the underlying idea behind the expected relationship with students' practice was that families with a higher SES generally place higher value on education, which might lead to more parental involvement and support during learning processes (Stocke, 2007). Furthermore, families with a higher SES are expected to have more opportunities to provide students with resources for learning, especially when it comes to access to and assistance with technology (Napoli & Obar, 2014; Van Dijk, 2005; Vekiri, 2010). However, in this dissertation, no effect of students' SES was found on their practice behavior. In the first implementation period (Study 1), a reason for the absence of the suggested link could be an invalid assessment. SES was measured via the parents' educational background (i.e., whether they had an Abitur [school leaving certificate enabling university entrance] or not); however, around 60% of students were not able to provide any information about this variable. Thus, this measure might not be an optimal proxy; this is also indicated by the missing relationship with students' migration background frequently reported in the literature (e.g., OECD, 2019; OECD, 2023, Stanat et al., 2022). Consequently, the results regarding students' SES and their practice with Math Garden need to be interpreted with caution. In the second implementation period (the replication of Study 1), SES was again assessed by asking about the parents' educational background in terms of the Abitur but also by measuring how many books students had at home. Again, about 60% of students were not able to provide information about their parents' educational background, and this variable was not correlated with students'

migration background. However, the number of books at home reported by students was correlated with a migration background in the expected direction (i.e., students with fewer books at home also reported having a migration background). Thus, the results of the replication analysis regarding Study 1 provided more valid information on the relationship between students' SES and their practice with Math Garden. Nonetheless, SES again turned out to be an insignificant predictor of the number of tasks completed in Math Garden. A reason for this missing effect could be that the proxy used referred to students' cultural capital (i.e., books). According to Bourdieu's capital theory (1986), SES is defined not only by families' cultural capital but also by their economic capital, such as financial and material assets that can be converted into money or that are already monetary. This would also include, for instance, digital devices. In this regard, research has already provided evidence that people with a low SES are more likely to own a smartphone and to not additionally own a desktop computer or laptop (Napoli & Obar, 2014). However, this is problematic because smartphones are used more frequently for entertainment-related or socially interactive purposes and less frequently for information and learning-oriented purposes, for instance, practicing with a math learning program (Napoli & Obar, 2014). Additionally, as Math Garden is a web-based and not an app-based program, its display on a small smartphone screen could be rather unfavorable in contrast to the display on a tablet or desktop computer. To summarize, to better understand how students' SES is related to their practice with a math learning program, it might be necessary to consider other SES aspects in future studies, particularly the economic capital.

### **8.2. RQ2: Effects of Program Provision and Students' Practice with a Math Learning Program on Students' Performance, Academic Motivation, and Emotions**

Using an adapted supply-use model as well as the cognitive theory of multimedia learning, the self-determination theory, and the control-value theory (Deci et al., 1991; Mayer, 2014; Pekrun, 2006), Study 2 and its replication investigated how the provision of the Math Garden program and also students' practice behavior with the program were related to students' performance, their math self-concept, and their math anxiety. Hereby, practice behavior was operationalized in terms of the number of tasks completed by students, but also via the number of weeks for which students practiced with Math Garden. Regarding students' performance, conclusions were drawn separately for students' addition and subtraction performance. Additionally, Study 3 contributed to this research question by investigating how the number of math-related tasks completed in Math Garden was related to students' overall arithmetic performance (i.e., a composite score for addition, subtraction, division, and multiplication).

No effect of the program provision was found on students' performance in the experimental group in contrast to the waitlist control group in Study 2 or its replication. However, significant links emerged between the number of (subtraction) tasks completed and students' (subtraction) performance (Study 2, the replication of Study 2, and Study 3). This means that the more (subtraction) tasks students completed, the higher their (subtraction) performance improvement was. Regarding students' math self-concept, the program provision led to a significant increase in the experimental group in contrast to the waitlist control group (Study 2). Surprisingly, this effect seemed to be independent of the actual practice behavior of the students. However, the significant effect of the program provision on math self-concept was not replicated in the additional analyses. Regarding students' math anxiety, no effects were found for the program provision or for students' practice behavior.

### 8.2.1 Performance

Study 2, including the replication analysis, failed to yield any empirical evidence regarding the effect of the provision of Math Garden on students' performance in addition or subtraction. Given the ongoing debate about the effectiveness of math learning programs (Byun & Joung, 2018; Pellegrini et al., 2021), the absence of performance effects is not entirely surprising and underscores the pattern of inconsistent research findings in this field (Tokac et al., 2019). However, the limitation of many previous studies was due to the fact that they exclusively examined the effects of the program provision, which only leaves room for speculation about the reasons for the lack of effects.

One plausible explanation for the missing effects could be that programs such as Math Garden might not effectively foster students' math skills. However, considering the fact that the program is designed to practice what has previously been learned and not to introduce new content, any anticipated minor effects on performance might be challenging to detect (Kraft, 2020), particularly if students, on average, do not engage enough with the program. Therefore, an alternative explanation for the missing effects could be that students did not practice sufficiently with Math Garden to derive any benefits. Nevertheless, to draw further conclusions, it is necessary to consider the quality of the program provision and also to account for students' practice behavior. Within this dissertation, the number of tasks completed was primarily understood as a measure of quantity, whereas the number of weeks spent practicing might better reflect how regularly students practiced with the program.

The results of both Study 2 and its replication, as well as the results of Study 3, consistently showed that the more (subtraction) tasks students practiced with Math Garden, the



higher their (subtraction) performance increases were. These findings are promising and lead to two considerations, namely (1) how missing effects might be explained regarding the provision of Math Gardens and (2) that it seems relevant to differentiate between arithmetic domains.

First, the positive effects of the number of tasks completed by students on their performance contribute to the idea that students need to actively engage with math learning programs to benefit from them. However, the large variance in students' practice behavior and the rather low average number of tasks completed was striking in all three investigations (i.e., Study 2, the replication of Study 2, and Study 3). A subgroup of students seemed to be highly engaged; thus, effects of the program provision on this subgroup could perhaps be expected in contrast to the control group. However, the low practice behavior of all other students might have leveled out this suggested positive effect. Thus, the pattern of findings could indicate that students in the experimental group, on average, engaged too little with Math Garden, which might explain the lack of effects regarding the program provision found in Study 2, as well as its replication. This suggestion needs further testing.

Second, Study 3 showed a positive effect for all math-related tasks that students completed in Math Garden on their performance improvement. However, performance improvement was measured in this study with a composite score for the domains of addition, subtraction, multiplication, and division. This could lead to the question of whether students benefitted equally regarding the separate arithmetic domains. Study 2 and its replication, which measured the domains of addition and subtraction separately, consistently showed an expected effect of the number of subtraction tasks completed on subtraction performance but did not find an effect for the number of addition tasks on addition performance. It could be argued that the missing addition effect found in Study 2 might be mainly due to a lack of power because of sample dropout in the posttest. However, the replication analysis also revealed this nonsignificant result pattern, even though it comprised a sample three times as large as that in Study 2. Therefore, a content-related explanation might be needed rather than a statistical explanation for the lack of effects on addition performance. In this vein, research has already provided evidence that subtraction is more difficult and complex than addition (Anderson et al., 2022; Kamii et al., 2001). Possibly, this is why practicing with Math Garden was particularly helpful for more demanding arithmetic domains (e.g., subtraction). Based on these findings, future research might also address separate effects in the domains of multiplication and division. As division tasks are suggested to be more difficult than multiplication tasks (Anderson et al.,

2022), one could expect to find similar result patterns for these two domains as for subtraction and addition.

Surprisingly, no effect on performance was found for the number of weeks for which students practiced with Math Garden. According to the idea of distributed learning behavior (Toppino & Gerbier, 2014), it is reasonable to assume that the number of weeks spent practicing is a relevant indicator. This means that practice is most effective when practice sessions are spread over time with breaks in between, rather than when tasks are repeated in immediate sequence (Carpenter et al., 2012; Toppino & Gerbier, 2014). The missing effects might have occurred because the variable was not meaningful enough; most students in both implementation periods practiced for only a small number of weeks and, thus, only a small—and probably not meaningful—variance was created within the variables for the number of weeks. Furthermore, a student who completed 100 tasks in 4 weeks had the same score as a student who completed 10 tasks in 4 weeks. This could indicate that the variable was possibly not differentiated enough. Thus, the number of tasks completed seemed to be a more precise indicator of students' practice behavior.

Still, the approach of taking the number of weeks spent practicing into account is unique and provides an initial starting point to investigate further data-driven indicators regarding students' practice behavior. To use a more fine-grained measure, future studies could take the number of days spent practicing into account. Moreover, in terms of challenging the hypothesis on distributed and massed practicing (e.g., see Carpenter et al., 2012; Toppino & Gerbier, 2014), future investigations could also combine both measurement approaches (i.e., number of tasks completed and number of weeks/days spent practicing) and use profile analyses to distinguish between different practicing styles. Four different styles could perhaps be expected, namely, (1) high in tasks and high in weeks/days, (2) high in tasks but low in weeks/days, (3) low in tasks but high in weeks/days, (4) low in tasks and low in weeks/days. Hereby, Style 1 and Style 3 might indicate distributed practice on different intensity levels, Style 2 might represent rather massed practice, and Style 4 might display an unfavorable avoidance type. Given the high correlation found between the number of tasks completed and the number of weeks spent practicing in both implementation periods of this dissertation, it could be expected that most students might belong to either Style 1 or Style 4. However, there could also be other styles of practice behavior as the relationship is not perfect (i.e.,  $r < 1$ ).

To sum up, the effects of practice behavior (i.e., the number of tasks completed) are promising and indicate that students need to actively engage with math learning programs to yield any performance increases. This might apply in particular to more difficult domains, such

as subtraction. However, regarding the insignificant results of the mere program provision, it is reasonable to assume that students, on average, practiced too little to show any substantial performance improvement in contrast to the control group. No effect was found for the number of weeks spent practicing, but this variable might not have been differentiated enough. Therefore, the idea of developing other practice behavior indicators suggests valuable avenues for future research.

### **8.2.2. Math Self-Concept**

The findings from Study 2 indicate that, as expected, the provision of Math Garden was positively related to students' math self-concept in the experimental group compared to the waitlist control group. Given that math self-concept is a crucial predictor of later academic performance (Eccles & Wigfield, 2020; Guay et al., 2003; Marsh & Martin, 2011), this result is highly promising. What is notable about the result is the buffering mechanism observed. While there was a descriptive decline in math self-concept in the waitlist control group over the intervention period—consistent with prior research suggesting an overall decrease in self-concept over a school year (Eshel & Klein, 1981; Fraine et al., 2007)—the provision of Math Garden appeared to counteract this decline and led to a significant increase in math self-concept in the experimental group. This result aligns with earlier research, for instance, Jansen et al. (2013) demonstrated a small but positive effect of providing Math Garden on perceived math competence in a Dutch sample; perceived math competence is a construct closely related to math self-concept (Lee, 2009).

However, the additional analyses regarding Study 2 were not able to replicate this beneficial finding of Math Garden's provision. Still, the descriptive results pattern was similar to the one observed in Study 2 in terms of a potential buffering effect. This means that, although math self-concept descriptively dropped within both groups, the decline was descriptively less strong within the experimental group. Nonetheless, the effect turned out to be insignificant. A reason for the unexpected finding in the replication might be the use of a different math self-concept scale to that used in Study 2. This instrument was well suited for the specific age group and reliability analyses provided good results for the internal consistency, which is why measurement problems can be ruled out. Thus, the pattern of findings might indicate that the effect is not stable across different operationalizations. Furthermore, the sample characteristics might also explain the missing effect, as students in the replication analysis obtained, on average, higher math self-concept scores in the pretest than students in Study 2 did. The characteristics of adaptivity and feedback, which were hypothesized as theoretical mechanisms

for an expected positive effect, might therefore no longer have come into play, as the students, on average, already had a more positive image of their mathematical abilities.

In short, the pattern of findings in this dissertation might indicate that the effect of Math Garden's provision on students' math self-concept is small ( $.16 \leq d \leq .26$ ) and statistically not robust. Regarding this result, one might argue that the effect could be negligible. However, Math Garden's focus is not exclusively to foster students' motivation but, rather, to foster students' performance; the hypotheses regarding motivation effects were made solely on the basis of the features of adaptivity and feedback included in the program. This is why large effects regarding students' math self-concept were also not expected (Kraft, 2020). Keeping this in mind, the size of the effect demonstrated in Study 2 ( $d = .26$ ) is similar to that of an intervention that solely aimed to increase students' math motivation (i.e., value beliefs; Gaspard et al., 2015). Thus, the effects of Math Garden on students' math self-concept might be promising, especially when taking specific sample characteristics into account. The samples from both implementation periods comprised students from non-academic-track schools, who are, on average, lower-achieving students. As math performance and math self-concept are closely linked constructs (Marsh & Martin, 2011), lower-achieving students often also have unfavorable motivational preconditions. Although prior research has already shown that math learning programs can be particularly effective for low-achieving students in terms of performance (e.g., Hassler Hallstedt et al., 2018; Ran et al., 2021), the effects on affective-motivational outcomes have not yet been investigated for such students. Thus, the results of this dissertation might provide first evidence that interventions such as Math Garden have the potential to foster lower achievers' math self-concept (see also Section 8.4 for limitations). To support this assumption, however, future research is needed.

Contrary to expectations, however, the results of Study 2 and the replication of Study 2 did not provide empirical evidence regarding an effect of practice behavior on math self-concept in the experimental group. This result is in line with that of Jansen et al. (2013), who also found no effect of the tasks completed in Math Garden on perceived math competence. Thus, it appears that the positive association with students' math self-concept (i.e. the positive effect of the program provision found in Study 2) seemed to be independent of students' actual practice behavior. Against the theoretical suggestions, this would mean that there was no relationship between the amount of positive feedback students received during practice and the effect of the program provision on math self-concept. This discrepancy might be explained by the results of a study conducted by Bernacki et al. (2015), who investigated the effect of feedback on the motivational construct of self-efficacy. They found that feedback on task

correctness was related to students' self-efficacy, particularly at the beginning of a training session. However, over time, previous self-efficacy became more predictive, while the number of correct answers and the amount of positive feedback became less relevant. This mechanism might also apply to students' math self-concept in the studies of the present dissertation. Thus, students' math self-concept might have increased shortly after the provision but then remained stable at this higher level. To further explore this assumption, future research could assess math self-concept multiple times during students' practice with the program.

### 8.2.3 Math Anxiety

Neither the results of Study 2 nor its replication provided empirical evidence for the expected effects of providing Math Garden or practicing with Math Garden on students' math anxiety. These outcomes align with the findings of Jansen et al. (2013), who similarly did not find any effects of Math Garden on math anxiety. Drawing on the results of Study 1 and past studies that showed that math anxious students tend to avoid math and math-related tasks due to unpleasant thoughts and feelings (Ashcraft & Kirk, 2001; Hembree, 1990; LeFevre et al., 1992; Quintero et al., 2022), it is reasonable to assume that these students consequently might not have derived benefits from the adaptive algorithm because they might have practiced too infrequently to notice that the program adjusts task based on the ability level. This means that they may not have perceived control over successfully solving the tasks in Math Garden, which has been proposed to be the underlying theoretical mechanism for potentially reducing math anxiety (Pekrun, 2006).

Furthermore, researchers have emphasized that math anxiety often begins at a young age and becomes deeply embedded in students over time (Ashcraft & Krause, 2007; Bekdemir, 2010; Ramirez et al., 2013). Thus, even though students' experience with Math Garden could have been positive due to the adaptive algorithm and the positive feedback, this might not outweigh negative math-related experiences from the past (Jansen et al., 2013).

Additionally, the assessment of math anxiety predominantly focused on math test anxiety in all investigations. However, students could have perceived Math Garden less as a test tool and more as an everyday practice tool, which is why their unpleasant feelings regarding math-related test situations might not have been minimized. Thus, Math Garden might be more effective in reducing other subcomponents, such as math calculation anxiety (Lukowski et al., 2019). In this light, a study by Lukowski and colleagues (2019) showed a strong negative relationship between math calculation anxiety and performance when controlling for other anxiety subcomponents. However, due to Math Garden's adaptive algorithm, students' chances

of solving tasks correctly are high (i.e., at least 75%). Hereby, a substantial amount of positive program feedback should make this beneficial process salient to students and, consequently, imply good performance. Due to the strong link between math calculation anxiety and math performance, one could suggest that students suffering from math calculation anxiety might be very sensitive to this positive performance feedback, which, in turn, could reduce their math calculation anxiety. However, with regard to the results that math anxious students potentially practice less with math learning programs, it might be crucial to ensure that students with high math calculation anxiety actually perceive the beneficial mechanisms by interacting sufficiently with the program. To gain a more holistic understanding regarding the potential effects of math learning programs on students' math anxiety, future investigations should consider other aspects of math anxiety, such as math calculation anxiety, as an outcome.

### **8.3 Theoretical and Practical Implications**

#### **8.3.1 Theoretical Implications**

Theories aim to reduce complex relationships to the interaction of a limited number of highly relevant variables. Thus, they facilitate the differentiated understanding of connections, enable predictions to be made, and enhance the successful planning of effective interventions (Borsboom et al. 2021). In other words, strong theories are essential for high-quality scientific research (Jonas et al., 2022).

When investigating the effects of (math) learning programs, most studies so far have based their assumptions on why learning programs might be beneficial for students' learning on models such as the cognitive theory of multimedia learning (Mayer, 2014). Indeed, Mayer (2014) provided strong rationale for why a program has the potential to foster students' learning, that is, through interactive elements such as adaptivity and feedback. Still, the assumed relationship refers exclusively to mechanisms within the program, that is, it takes a rather program-centered perspective and does not facilitate any assumptions about how person-related factors, such as students' interaction with a learning program, may contribute to potential effects. Addressing this deficit, Reinhold et al. (2024) very recently introduced a cognitive process framework for the mechanisms of learning with digital tools (CoDiL). In this framework, the idea is that, through digitally enriched instruction, such as the provision of digital tools that include beneficial features like adaptive feedback prompts, and students' learning activity with such digital tools, knowledge could evolve (Reinhold et al., 2024). Hereby, the CoDiL framework puts a special focus on cognitive processes during students'

learning activity and their knowledge acquisition. However, while this framework is progressive in terms of integrating students' learning behavior into an assessment, it focuses solely on knowledge acquisition, that is, performance, and neglects to consider further contextual factors such as students' individual preconditions. Up until now, there is no common comprehensive theoretical model within learning program research that links contextual factors, such as students' characteristics, and program-related factors with students' learning processes and their learning outcomes. Making such links could be important for researchers as it could enable them to derive testable hypotheses; it could also be important from the perspective of forming a common understanding within the learning program research community about which direction future investigations should take.

Even though the three empirical studies and the additional replication analyses conducted in this dissertation were not able to confirm all assumptions regarding the theoretical model, the results show that adapting a supply-use model to the context of providing a math learning program can offer new perspectives. Based on this model, it was hypothesized that contextual factors, such as students' individual characteristics, are related to their practice behavior with a math learning program and, in turn, that students' practice behavior is associated with various learning outcomes, such as students' performance, their academic motivation, and their emotions. The results empirically supported the theoretical assumption that specific student characteristics (math anxiety and a migration background) are related to their practice behavior and that students' practice behavior is related to their performance. Accordingly, the adapted model might be a valuable starting point for future investigations. For instance, the model could further assume that parent and teacher characteristics might also play an important role in the successful implementation of a math learning program. Such contextual factors were not addressed within this dissertation (see limitations in Section 8.4 and future research in Section 8.5).

Besides the described advantages, the adapted supply-use model nonetheless also has limits. For instance, it cannot explain the lack of an effect on students' math anxiety or why Study 2 found an effect of Math Garden's provision on students' math self-concept that appeared to be independent of their practice behavior. For this reason, the rather broad model would benefit from the integration of more in-depth perspectives on the underlying processes. Against this background, Vanbecelaere et al. (2023) recently proposed a theoretical assessment framework for investigating the effectiveness of digital learning programs that includes students' practice behavior and potential cognitive (e.g., performance) and noncognitive outcomes (e.g., academic motivation and emotions). The authors presented two nuanced

measurement strategies, which formed the center of this framework. First, regarding potential outcomes, researchers should differentiate “when” to assess them, such as during the intervention, immediately after the intervention, or significantly after the intervention (Vanbecelaere et al., 2023). Regarding the missing and mixed result pattern of students’ math anxiety and math self-concept found in this dissertation, a measurement during the intervention (i.e., using experience-sampling methods; see also Section 8.5 for future research suggestions) might shed light on how students perceive a program’s adaptivity and feedback (which were assumed to be underlying effect mechanisms) during practice. Second, Vanbecelaere et al. (2023) emphasized the relevance of using various data sources, such as process measures, physiological data, observational data, questionnaire data, test data, and interview data. This applies to both the outcomes and students’ practice behavior. As already pointed out, this is an important approach as it can minimize potential biases (Azevedo, 2015).

To sum up, the adapted supply-use model (Helmke & Weinert, 1997; Seidel, 2014) and the assessment framework proposed by Vanbecelaere et al. (2023) have different focuses. While the supply-use model puts a key focus on students’ use of a supply structure (i.e., math learning programs) and on context-related factors, Vanbecelaere et al.’s framework concentrates more on the differentiated assessment of students’ outcomes. The two theories could therefore complement each other in essential aspects. In combination, they might serve as a starting point for establishing a common model, with a sound theoretical grounding, for examining the contextual factors related to students’ use of (math) learning programs and a wide range of cognitive and noncognitive learning outcomes.

### **8.3.2 Practical Implications**

When discussing practical implications of adaptive math learning programs such as Math Garden, it is important to emphasize that these are supplementary tools to support students in the math learning process, and they are not intended to replace traditional math instruction. As an additional tool, they offer the advantage of being flexible, that is, teachers can either integrate them into lessons to enable individualized learning or assign practice as homework. However, even though all students receive the same learning program, the results of this dissertation suggest that not all students might equally benefit from it. Thus, practitioners might play an important role in making the implementation of math learning programs successful.

This dissertation provides empirical evidence that students with specific characteristics, such as math anxious students and students with a migration background, need additional support in a math learning program implementation in order to achieve a similar level of



practice as their peers with less math anxiety or without a migration background. Without such support, these students may practice less, thereby missing the potential benefits of the program. Math anxious students might especially focus on the unpleasant negative emotions and feelings triggered by the math content, resulting in less perceived control over the resources required to successfully solve a task (Ashcraft & Moore, 2009; Pekrun, 2006). Thus, in line with the control-value theory (Pekrun, 2006), potential support strategies could be based on increasing math anxious students' perceived control. For instance, the initial stage of the program implementation could involve offering an extended guided introductory period. This could help to make math anxious students feel more at ease with the program. Within adaptively working programs, it might also be useful to lower item difficulties to give math anxious students even higher chances of success in the first place (Rocklin & Thompson, 1985). For instance, within Math Garden, teachers can manually set the default setting of 75% solution probability to 90% solution probability. Prior research has suggested that students tend to experience less math anxiety when they perceive their teachers as being highly aware of and responsive to their emotional needs (Aldrup et al., 2020). Consequently, cultivating teachers' sensitivity to the needs of math anxious students could help these students to overcome their inhibited practice behavior. Hereby, teachers can demonstrate sensitivity in various ways. This includes actively and consciously observing indicators of students requiring support or assistance, such as changes in behavior, facial expressions, or gestures (Pianta et al., 2012). Consistent responses to students' needs, such as allowing sufficient time for task processing, are also crucial. Additionally, teachers can contribute to reducing math anxiety by enhancing students' autonomy, offering positive feedback, and fostering a positive error culture (Brandenberger & Moser, 2018). In the context of practicing with math learning programs, this might involve allowing math anxious students to initially choose the domains they wish to practice.

Regarding students with a migration background, studies have shown that these students often perceive migration as a possibility for social advancement and thus come motivated with high ambitions (Kao & Tienda 1995; Stanat et al., 2010). However, they are mostly unable to realize these ambitions, which is usually attributed to a lack of (family) resources and a language barrier (Stanat et al., 2010; Portes & Rumbaut, 2001). Thus, this is where practitioners might take action to overcome the inhibited practice of students with a migration background. Regarding material resources, a digital divide for students with a migration background is evident (Van de Werfhorst et al., 2022). This means that these students are more often smartphone users and do not own a tablet or desktop computer (Ertl et al., 2020). However, smartphones tend to be used more for socially interactive and entertainment-related purposes

(Napoli & Obar, 2014), which may distract students from learning. In contrast, tablets and desktop computers are more likely to be used for learning purposes (Napoli & Obar, 2014). Tablets and desktop computers also offer larger screens, on which web-based learning programs such as Math Garden can be more accurately displayed. Thus, it might be necessary to ensure that students with a migration background have at least a tablet available for practicing with math learning programs at school and at home. In addition to this digital divide, a language barrier might hinder students with a migration background from practicing, which is why teachers need to make sure that these students can clearly follow the instructions when a math learning program is implemented. To address language barriers over the long term, students with a migration background could collaborate with a native-speaking peer while engaging in practice sessions during school hours. This collaborative approach could provide a low-threshold way to ask language-related questions. It is essential to note that tasks in the Math Garden program are less language-intensive, but they still seemed to disadvantage students with a migration background. Thus, practitioners utilizing programs with a greater emphasis on language, for instance, programs including word problem-solving tasks such as Bettermarks, should keep this in mind. However, a language barrier could also hinder the parents of students with a migration background from supporting them with practice at home (Dumont et al., 2014). Thus, it might also be important to provide information about the provision of the program at a parents' evening and to offer translated instructions on how the program works.

The findings of this dissertation regarding the effects of practice behavior on students' performance, including a domain specificity regarding subtraction, have two further important implications for teachers. First, simply providing students with a math learning program might not mean that they will definitely benefit from it. The teacher's role could therefore be to support the students in continuously and systematically engaging with the math learning program provided (Spitzer & Musslick, 2021). In this light, teachers could establish short but regular practice sessions within their math lessons, for instance, 10 minutes at the end of the lesson, to familiarize students with everyday practice. Furthermore, they should prioritize acknowledging students' efforts during practice sessions rather than focusing solely on their performance results (Blackwell et al., 2007).

Second, regarding the result that students' subtraction performance improved depending on the number of subtraction tasks they completed, it could be recommended that Math Garden is used specifically to foster students' subtraction skills. Against this background, it is the teacher's responsibility not only to ensure regular practice but also to monitor the domains in which students practice. For instance, students might not benefit if they practice too few

subtraction tasks or if they only practice easier addition tasks. This underscores the relevance of introducing teachers to the possibility of monitoring their students' practice during the program implementation (Hillmayr et al., 2020). In Math Garden, teachers can, for example, monitor the number of tasks that students complete within each domain in a teachers' back end. In this back end, they also have the possibility to prioritize domains, such as subtraction, so that the domain is highlighted for students. Additionally, they can temporarily block certain domains, which might be useful if students, for instance, practice only addition and no other domain. Future research is warranted to derive practical implications regarding other arithmetic domains, such as multiplication and division.

Lastly, the potential of Math Garden to foster students' math self-concept might also be promising for practitioners. According to the expectancy-value theory (Eccles & Wigfield, 2020), a higher math self-concept could lead to higher student engagement in math and math-related tasks. This suggested beneficial mechanism, in turn, might have lasting effects on students' future math performance (Eccles & Wigfield, 2020; Guay et al., 2003; Marsh & Martin, 2011). However, for the Math Garden learning program, these assumptions need to be tested further.

### 8.4 Limitations

Some important limitations need to be considered when interpreting the results of the present dissertation. Given that the majority of high-achieving students in Germany typically attend academic-track schools (Maaz et al., 2008), it is important to note that the samples from both implementation periods only comprised students from non-academic-track schools, that is, on average, lower-achieving students. Consequently, the findings might not be broadly applicable to all types of schools in Germany, nor to all achievement levels, and they need further investigation.

Furthermore, even though the high ecological validity of the implementation of the learning program Math Garden in this dissertation makes it a desirable approach, it comes with the disadvantage that environment-related variables cannot be adequately controlled for. It was hoped that randomization at the class level in Implementation Period 1 and randomization at the school level in Implementation Period 2 would partially counteract this disadvantage, as relevant unknown third variables would be held constant. Nonetheless, the environment in which students usually practiced with Math Garden (i.e., in school or at home) was not known. However, specific environment-related factors could have affected their way of engaging with the program. It can be assumed that students' use of the program during school hours involves

teacher-related factors, whereas their use at home involves parent-related factors (Wong & Liem, 2022). Such mechanisms were, however, not considered within this dissertation although they may be highly relevant. A successful program implementation in class might depend on the digital skills of the teacher and their attitude towards digital media (Eickelmann et al., 2019) but also on their access to technical devices and additional support, which would enable them to confidently use the program within math instruction (Ozel et al., 2008). Therefore, future research should take such teacher-related characteristics into account. A first approach to how to investigate the implementation quality in connection with teacher-related aspects and based on process measures is discussed in Section 8.5. With regard to parent characteristics, studies on the quality of parental homework support indicated that high levels of responsiveness and structure could have positive effects on students' learning outcomes but that high levels of control appeared to be unfavorable (Benckwitz et al., 2023; Dumont et al., 2014; Senkbeil, 2023). Similar mechanisms might apply when students practice with math learning programs at home. For instance, students might engage most effectively when their parents are available to respond to questions, such as questions on the technical use of the program or questions on solving a specific math task. Furthermore, it might be beneficial if parents remind their children to practice and support them in structuring the time for practicing. However, parents should not control their child's practice, that is, they should not interfere unless the child asks for help. Thus, these forms of parental support could be related to the use of a math learning program at home and might be considered in future research.

Regarding the specific learning program implemented in this dissertation, it should be noted that Math Garden mainly contains tasks on arithmetic operations that are generally less complex. In addition, the program works adaptively on students' individual performance level and also offers corrective feedback and incentives. Therefore, the effects found might be primarily due to these specific program characteristics and features. In this light, future research might investigate how varying task complexity, using adaptive processes that function differently, varying feedback complexity, and using different incentive systems could contribute to similar effects.

### **8.5 Future Research**

The aim of this dissertation was to investigate whether math learning programs such as Math Garden can provide students with a beneficial supply structure for learning and, in particular, how student characteristics and their behavior with the program might contribute to these benefits. These questions are still not definitively answered and the way in which various other

factors may be associated with a successful learning program implementation remains unclear. Based on the ideas and findings of this dissertation, future research might therefore focus on investigating (1) the underlying processes during students' practice with such learning programs and (2) the role of the teacher with regard to the quality of the program implementation.

The results of this dissertation cannot answer the question of why programs like Math Garden might work. The assumed underlying theoretical mechanisms, such as students' perception of adaptivity, feedback, and rewards during practice, were not examined but rather taken as beneficial preconditions. However, one can speculate that, for instance, no effects on students' math anxiety were found because students did not perceive that tasks were more controllable due to the adaptive algorithm. Therefore, future research could investigate students' perceptions of program mechanisms such as adaptivity and feedback during their practice with the program, especially regarding their effects on students' motivation and emotions. Within the field of computer-based educational testing, similar approaches have already been considered, such as on-task investigations that consider the effect of the complexity of feedback error messages on students' situative affect and motivation (e.g., Kuklick et al., 2023; Kuklick & Lindner, 2021). The results of such investigations showed that less complex feedback that was provided after incorrect responses might be detrimental to students' motivation (Kuklick et al., 2023). However, these investigations mostly referred to a single assessment situation that was usually carried out under laboratory-like conditions. In contrast, programs such as Math Garden might be understood more as practice tools and less as test instruments; therefore, it is important to investigate their effects in the long term under highly ecological conditions.

A suitable method for examining such program mechanisms could be the use of experience sampling (i.e., repeated measures, e.g., Niepel et al., 2022; Talić et al., 2022). With this method, students are presented with very short one-item questions several times on a daily basis. Hereby, they might provide information on how difficult they found the previously solved math tasks and on how they perceived their chances of successfully solving the following math tasks. Additionally, students' situative math self-concept or their situative math anxiety could be assessed. Over time, task difficulty could then be varied according to a within-person design. Thus, on the basis of the assessment of these intraindividual trajectories, researchers might be able to investigate how students' affective-motivational states develop during practice with math learning programs by considering the actual and perceived task difficulty level. While additionally considering students' level of engagement in terms of process measures and taking

a strong person-centered perspective, this might shed light on the puzzling but important question of how the reciprocal interplay of these variables can contribute to students' short- and long-term performance increases as a result of their engagement with digital math learning programs.

As mentioned above, this dissertation is limited regarding information on how teachers integrated Math Garden into their instruction or how they structured student practice. Thus, in future studies, researchers could integrate teacher questionnaires that assess, for instance, how often teachers use the tool in class or assign it as homework. However, such self-reports could be biased and might hinder an understanding of the actual quality of the program implementation (Donaldson & Grant-Vallone, 2002). To facilitate assumptions about implementation quality that are based on more objective measures, stored process measures could be further analyzed using methods from the field of learning analytics (e.g., Gašević et al., 2015). Regarding such methods, the idea is that, by clustering process-measured data, meaningful data patterns might emerge (e.g., Khalil & Ebner, 2017; Kizilcec et al., 2017). For instance, based on the cluster ID (i.e., indicating the certain class to which a cluster belongs) and a time stamp, it is possible to determine whether a teacher used a learning program more within instruction (i.e., students practiced more during morning hours) or assigned it more as homework (i.e., students practiced more during afternoon hours). Furthermore, by examining the individual domains that students practiced within their cluster according to a specific time point, it would be possible to derive information about the extent to which teachers structured students' use of the program; if most students in a cluster practiced the same domain at the same time, this might indicate a highly structured use. If students practiced different domains, in contrast, this would indicate a less structured use. Thus, such "features", which are suggested to group the massive process-data structures into meaningful data patterns, could serve as indicators of the implementation quality of math learning programs. Against this background, a high implementation quality might be defined by distributed and consistent program use during school hours and at home, as well as by a structured task assignment, which, in turn, could have a positive effect on students' performance improvement (Toppino & Gerbier, 2014). However, with regard to the basic dimensions of teaching quality (Klieme et al., 2001; Praetorius et al., 2018), it is reasonable to assume that certain dimensions, such as classroom management, might be related to implementation quality. Thus, future research could address how the relationship between teaching quality dimensions and students' math performance is mediated by the objectively measured quality of the program implementation.

## 8.6 Conclusions

In the hope that math learning programs can counteract the worrying decline in students' math competence and the widening achievement gap (OECD, 2023; Stanat et al., 2022), research on math learning programs has taken off in recent years (Byun & Joung, 2018; Ma et al., 2014; Sailer & Homner, 2020; Tokac et al., 2019). Nonetheless, the findings so far are mixed and the programs do not seem to meet the expectation that they can successfully support student learning (Pellegrini et al., 2021). However, it should be noted that the studies conducted up until now usually only investigated the mere provision of programs (e.g., see Bai et al., 2012; Chang et al., 2015; Hung et al., 2014; Shin et al., 2012), which means that the possible variation in students' engagement was largely neglected (cf. Haleva et al., 2021; Hassler Hallstedt et al., 2018; Spitzer, 2022). However, this might be a decisive factor in whether a program implementation is successful or not (Cheung & Slavin, 2013; Ozel et al., 2008).

Employing the supply-use model (e.g., Helmke & Weinert, 1997; Seidel, 2014), the present dissertation went beyond this limited approach to pose two relevant research questions with regard to students' actual engagement (i.e., practice behavior). First, it examined how individual student characteristics, namely, prior math knowledge, conscientiousness, math self-concept, math anxiety, SES, and a migration background, were related to students' practice behavior, and, second, it examined how the program provision, as well as students' practice behavior, contributed to effects regarding students' learning outcomes (i.e., performance, math self-concept, and math anxiety). To address both questions, the adaptive arithmetic learning program Math Garden was offered to students in two implementation periods, varying between 22 and 45 weeks. In the resulting three studies and two additional replication analyses, students' practice behavior was operationalized with objective process data.

First, the findings indicated that math anxious students and students with a migration background displayed lower levels of practice. A migration background might be the more relevant determinant of students' practice when students experience less math anxiety; however, this assumption needs further testing. For practitioners, these results indicate that students with a migration background and math anxious students need more initial support and encouragement in order to reach a similar level of practice as their peers with less math anxiety or without a migration background, i.e. to at least get the chance to benefit from a math learning program implementation. In contrast, prior math knowledge, conscientiousness, math self-concept, and SES were not related to students' practice. These results are positive as they imply that such characteristics do not appear to disadvantage students in terms of practice behavior.

However, further research is needed to draw more general conclusions for other learning programs, age groups, and school tracks.

Second, the results provided evidence that higher levels of practice (i.e., higher numbers of completed tasks) were linked to students' performance increases. Furthermore, the results suggested that this might especially apply to more difficult domains (e.g., subtraction) in contrast to easier domains (e.g., addition). However, it was noticeable that students varied highly regarding their practice and, on average, did not practice much. This might be the reason for the missing effect regarding the mere program provision in the experimental group in contrast to the control group. For researchers, the pattern of findings indicates the relevance of considering students' practice behavior when drawing conclusions about the missing effects of a program provision on students' performance. For practitioners, this underscores the importance of ensuring that students systematically and consistently practice with a math learning program if performance improvements are to be achieved.

With regard to students' academic motivation and emotions, the findings indicate the potential of (adaptive) math learning programs to foster students' math self-concept; however, the effects were small and varied in their statistical significance. Still, keeping in mind the important role that math self-concept plays in students' future performance (Marsh & Martin, 2011), this result might be promising. Interestingly, the effect seemed to be independent of how much students practiced with the program. Regarding students' math anxiety, no effects were found. These findings suggest that students' perception of adaptivity and feedback should be investigated more closely in future research.

To conclude, this dissertation, with its high ecological validity, provides relevant insights for researchers and practitioners, whereby two key messages can be derived. First, certain students may be at risk of avoiding practice with math learning programs. Second, students need to actively engage with a math learning program to unlock its potential to improve their performance. Consequently, future research should aim to identify the contextual factors that lead to persistent and regular practice for all students. Against this background, the ideas and results of this dissertation can provide a nuanced road map for future research on math learning programs that, for instance, could take teachers and parents into account.



## 8.7 References

- Aldrup, K., Klusmann, U., & Lüdtke, O. (2020). Reciprocal associations between students' mathematics anxiety and achievement: Can teacher sensitivity make a difference? *Journal of Educational Psychology, 112*(4), 735–750.  
<https://doi.org/10.1037/edu0000398>
- Anderson, S., Sommerhoff, D., Schurig, M., Ufer, S., & Gebhardt, M. (2022). Developing learning progress monitoring tests using difficulty-generating item characteristics. An example for basic arithmetic operations in primary schools. *1866-6671, 14*.  
<https://doi.org/10.25656/01:24875>
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology. General, 130*(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248. <https://doi.org/10.3758/BF03194059>
- Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27*(3), 197–205.  
<https://doi.org/10.1177/0734282908330580>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43*(6), 993–1003.  
<https://doi.org/10.1111/j.1467-8535.2011.01269.x>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. *Educational Studies in Mathematics, 75*(3), 311–328. <https://doi.org/10.1007/s10649-010-9260-7>
- Benckwitz, L., Kohl, K., Roloff, J., Lüdtke, O., & Guill, K. (2023). Reciprocal relationships between parental and scholastic homework assistance and students' academic

- functioning at elementary school. *Frontiers in Psychology*, 14, 1106362.  
<https://doi.org/10.3389/fpsyg.2023.1106362>
- Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: variability and relations to behavior, performance, and learning. *Metacognition and Learning*, 10(1), 99–117. <https://doi.org/10.1007/s11409-014-9127-x>
- Blackwell, L., Trzesniewski, K., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246–263. <https://doi.org/10.1111/j.1467-8624.2007.00995.x>
- Borsboom, D., Van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: a practical framework for building theories in psychology. *Perspectives on Psychological Science*, 16(4), 756–766.  
<https://doi.org/10.1177/1745691620969647>
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.
- Brandenberger, C. C., & Moser, N. (2018). Förderung der Lernfreude und Reduzierung der Angst im Mathematikunterricht in der Sekundarstufe 1 [Promoting the joy of learning and reducing anxiety in mathematics at secondary level]. In G. Hagenauer & T. Hascher (Eds.), *Emotionen und Emotionsregulation in Schule und Hochschule* (pp. 323–337). Waxmann.
- Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education*, 128, 63–74.  
<https://doi.org/10.1016/j.compedu.2018.09.011>
- Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics*, 118(3-4), 113–126.  
<https://doi.org/10.1111/ssm.12271>
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2015). The chicken or the egg? The Direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, 6, 1987. <https://doi.org/10.3389/fpsyg.2015.01987>
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and

- implications for instruction. *Educational Psychology Review*, 24(3), 369–378.  
<https://doi.org/10.1007/s10648-012-9205-z>
- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International*, 52(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.  
<https://doi.org/10.1016/j.edurev.2013.01.001>
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20(1), 2–17.  
<https://doi.org/10.1016/j.learninstruc.2008.11.002>
- Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 34(1), 35–46. <https://doi.org/10.1177/001316447403400105>
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and Education: The Self-Determination Perspective. *Educational Psychologist*, 26(3-4), 325–346. <https://doi.org/10.1080/00461520.1991.9653137>
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2), 245–260. <https://doi.org/10.1023/A:1019637632584>
- Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, 11, 591203. <https://doi.org/10.3389/fpsyg.2020.591203>
- Dumont, H., Trautwein, U., Nagy, G., & Nagengast, B. (2014). Quality of parental homework involvement: Predictors and reciprocal relations with academic functioning in the reading domain. *Journal of Educational Psychology*, 106(1), 144–161.  
<https://doi.org/10.1037/a0034100>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859.  
<https://doi.org/10.1016/j.cedpsych.2020.101859>

- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Eds.). (2019). *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking [ICILS 2018 #Germany: Computer and information-related competencies of students in the second international comparison and competencies in computational thinking]*. Waxmann.
- Ertl, B., Csanadi, A., & Tarnai, C. (2020). Getting closer to the digital divide: An analysis of impacts on digital competencies based on the German PIAAC sample. *International Journal of Educational Development*, 78, 102259. <https://doi.org/10.1016/j.ijedudev.2020.102259>
- Eshel, Y., & Klein, Z. (1981). Development of academic self-concept of lower-class and middle-class primary school children. *Journal of Educational Psychology*, 73(2), 287–293. <https://doi.org/10.1037/0022-0663.73.2.287>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Fraine, B. de, Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, 32(1), 132–150. <https://doi.org/10.1016/j.cedpsych.2006.10.005>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, 51(9), 1226–1240. <https://doi.org/10.1037/dev000002>
- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95(1), 124–136. <https://doi.org/10.1037/0022-0663.95.1.124>
- Haleva, L., HersHKovitz, A., & Tabach, M. (2021). Students' activity in an online learning environment for mathematics: The role of thinking levels. *Journal of Educational Computing Research*, 59(4), 686–712. <https://doi.org/10.1177/0735633120972057>

- Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology, 110*(8), 1127–1148. <https://doi.org/10.1037/edu0000264>
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen [Conditional factors of academic performance]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie, Bd. 3: Psychologie des Unterrichts und der Schule* (pp. 71–176). Hogrefe.
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33. <https://doi.org/10.2307/749455>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153*, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of Educational Research, 67*(1), 3–42. <https://doi.org/10.3102/00346543067001003>
- Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education, 1*(2-3), 151–166. <https://doi.org/10.1007/s40692-014-0008-8>
- Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24*, 190–197. <https://doi.org/10.1016/j.lindif.2012.12.014>
- Jonas, E., Zerle, G., Mühlberger, C., & Hansen, J. (2022). Nichts ist praktischer als eine gute Theorie—Wenn Theorien praktisch werden [Nothing is more practical than a good theory - when theories become practical]. *Organisationsberatung, Supervision, Coaching, 29*(3), 289–311. <https://doi.org/10.1007/s11613-022-00781-4>
- Kamii, C., Lewis, B. A., & Kirkland, L. D. (2001). Fluency in subtraction compared with addition. *The Journal of Mathematical Behavior, 20*(1), 33–42. [https://doi.org/10.1016/S0732-3123\(01\)00060-8](https://doi.org/10.1016/S0732-3123(01)00060-8)
- Kao, G., & Tienda, M. (1995). Optimism and achievement. The educational performance of immigrant youth. *Social Science Quarterly, 76*(1), 1–19.

- Kast, J., Lehofer, M., Tanzer, N., Hagn, J., & Schwab, S. (2021). Parents' attitudes towards students with a background of migration—Does the background of migration matter? *Studies in Educational Evaluation*, 70, 101014.  
<https://doi.org/10.1016/j.stueduc.2021.101014>
- Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of Computing in Higher Education*, 29(1), 114–132. <https://doi.org/10.1007/s12528-016-9126-9>
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104, 18–33.  
<https://doi.org/10.1016/j.compedu.2016.10.001>
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung [Mathematics teaching in lower secondary schools: Task culture and lesson design]. *TIMSS - Impulse Für Schule Und Unterricht*, 43–57.
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824.  
<https://doi.org/10.1016/j.compedu.2011.02.003>
- Konrad-Ristau, K., & Burghardt, L. (2021). Differences in children's social development: How migration background impacts the effect of early institutional childcare upon children's prosocial behavior and peer problems. *Frontiers in Psychology*, 12, 614844.  
<https://doi.org/10.3389/fpsyg.2021.614844>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Kuklick, L., Greiff, S., & Lindner, M. A. (2023). Computer-based performance feedback: Effects of error message complexity on cognitive, metacognitive, and motivational outcomes. *Computers & Education*, 200, 104785.  
<https://doi.org/10.1016/j.compedu.2023.104785>
- Kuklick, L., & Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: Effects on performance, motivation, and achievement emotions. *Contemporary Educational Psychology*, 67, 102001.  
<https://doi.org/10.1016/j.cedpsych.2021.102001>

- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- LeFevre, J.-A., Kulak, A. G., & Heymans, S. L. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science, 24*(3), 276–289. <https://doi.org/10.1037/h0078742>
- Lukowski, S. L., DiTrapani, J. B., Jeon, M., Wang, Z., Schenker, V. J., Doran, M. M., Hart, S. A., Mazzocco, M. M. M., Willcutt, E. G., Thompson, L. A., & Petrill, S. A. (2019). Multidimensionality in the measurement of math-specific anxiety and its relationship with mathematical performance. *Learning and Individual Differences, 70*, 228–235. <https://doi.org/10.1016/j.lindif.2016.07.007>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. <https://doi.org/10.1037/a0037123>
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. <https://doi.org/10.1111/j.1750-8606.2008.00048.x>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology, 81*(Pt 1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- Napoli, P. M., & Obar, J. A. (2014). The emerging mobile internet underclass: A critique of mobile internet access. *The Information Society, 30*(5), 323–334. <https://doi.org/10.1080/01972243.2014.944726>
- Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. *Journal of Educational Psychology, 114*(6), 1380–1393. <https://doi.org/10.1037/edu0000716>
- OECD (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing.

- OECD. (2023). *PISA 2022 results: The state of learning and equity in education (Vol. I)*. OECD Publishing.
- Ozel, S., Yetkiner, Z. E., & Capraro, R. M. (2008). Technology in K-12 mathematics classrooms. *school science and mathematics*, 108(2), 80–85.  
<https://doi.org/10.1111/J.1949-8594.2008.TB17807.X>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7, 233285842098621.  
<https://doi.org/10.1177/2332858420986211>
- Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom assessment scoring system (CLASS)*. Teachstone.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322–338.  
<https://doi.org/10.1037/a0014996>
- Portes, A., & Rumbaut, R. G. (2001). *Legacies. The story of the immigrant second generation*. University of California Press.
- 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, 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Quintero, M., Hasty, L., Li, T., Song, S., & Wang, Z. (2022). A multidimensional examination of math anxiety and engagement on math achievement. *The British Journal of Educational Psychology*, 92(3), 955–973.  
<https://doi.org/10.1111/bjep.12482>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, 14(2), 187–202. <https://doi.org/10.1080/15248372.2012.664593>
- Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, 53(3), 145–164. <https://doi.org/10.1080/00461520.2018.1447384>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12



- classrooms. *Journal of Educational Computing Research*, 59(1), 119–153.  
<https://doi.org/10.1177/0735633120952063>
- Reinhold, F., Leuders, T., Loibl, K., Nückles, M., Beege, M., & Boelmann, J. M. (2024). Learning mechanisms explaining learning with digital tools in educational settings: A cognitive process pramework. *Educational Psychology Review*, 36(1), 14.  
<https://doi.org/10.1007/s10648-024-09845-6>
- Rocklin, T., & Thompson, J. M. (1985). Interactive effects of test anxiety, test difficulty, and feedback. *Journal of Educational Psychology*, 77(3), 368–372.  
<https://doi.org/10.1037/0022-0663.77.3.368>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. Advance online publication.  
<https://doi.org/10.25656/01:14686>
- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540–560. <https://doi.org/10.1111/j.1467-8535.2011.01197.x>
- Senkbeil, M. (2023). How well does the digital home learning environment predict ICT literacy and ICT self-efficacy? Comparing the predictive power of adolescent and parent reports. *Computers & Education*, 207, 104937.  
<https://doi.org/10.1016/j.compedu.2023.104937>
- Silinskas, G., & Kikas, E. (2019). Math homework: Parental help and children's academic outcomes. *Contemporary Educational Psychology*, 59, 101784.  
<https://doi.org/10.1016/j.cedpsych.2019.101784>
- Spitzer, M. W. H., & Musslick, S. (2021). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic. *PloS One*, 16(8), e0255629.  
<https://doi.org/10.1371/journal.pone.0255629>
- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am*

- Ende der 4. Jahrgangsstufe im dritten Ländervergleich [IQB education trend 2021: Competence in the subjects german and mathematics at the end of year 4 in the third federal state comparison]*. Waxmann Verlag GmbH.  
<https://doi.org/10.31244/9783830996064>
- Stanat, P., Segeritz, M., & Christensen, G. (2010). Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund [School-related motivation and aspiration of students with a migration background]. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung: Festschrift für Jürgen Baumert* (pp. 31–58). Waxmann.
- Stocke, V. (2007). Explaining educational decision and effects of families' social class position: An empirical test of the breen goldthorpe model of educational attainment. *European Sociological Review*, 23(4), 505–519. <https://doi.org/10.1093/esr/jcm014>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.  
<https://doi.org/10.1023/A:1022193728205>
- Talić, I., Scherer, R., Marsh, H. W., Greiff, S., Möller, J., & Niepel, C. (2022). Uncovering everyday dynamics in students' perceptions of instructional quality with experience sampling. *Learning and Instruction*, 81, 101594.  
<https://doi.org/10.1016/j.learninstruc.2022.101594>
- Tett, R. P., Simonet, D. V., Walser, B., & Brown, C. (2013). Trait activation theory. *Handbook of personality at work*, 71–100.
- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Tomaszewski, W., Xiang, N., & Western, M. (2020). Student engagement as a mediator of the effects of socio-economic status on academic performance among secondary school students in Australia. *British Educational Research Journal*, 46(3), 610–630.  
<https://doi.org/10.1002/berj.3599>
- Toppino, T. C., & Gerbier, E. (2014). About practice. In *Psychology of Learning and Motivation* (Vol. 60, pp. 113–189). Elsevier. <https://doi.org/10.1016/B978-0-12-800090-8.00004-4>

- Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology*, 99(2), 432–444. <https://doi.org/10.1037/0022-0663.99.2.432>
- Trautwein, U., Lüdtke, O., Kastens, C., & Köller, O. (2006). Effort on homework in grades 5–9: Development, motivational antecedents, and the association with effort on classwork. *Child Development*, 77(4), 1094–1111. <https://doi.org/10.1111/j.1467-8624.2006.00921.x>
- Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, 97(6), 1115–1128. <https://doi.org/10.1037/a0017048>
- Van de Werfhorst, H. G., Kessenich, E., & Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools. *Computers and Education Open*, 3, 100100. <https://doi.org/10.1016/j.caeo.2022.100100>
- Van der Beek, J. P. J., Van der Ven, S. H. G., Kroesbergen, E. H., & Leseman, P. P. M. (2017). Self-concept mediates the relation between achievement and emotions in mathematics. *The British Journal of Educational Psychology*, 87(3), 478–495. <https://doi.org/10.1111/bjep.12160>
- Van Dijk, J. (2005). *The deepening divide: Inequality in the information society*. SAGE Publications, Inc. <https://doi.org/10.4135/9781452229812>
- Vanbecelaere, S., Demedts, F., Reynvoet, B., & Depaepe, F. (2023). Toward a framework for analyzing adaptive digital games' research effectiveness. *International Journal of Serious Games*, 10(4), 77–91. <https://doi.org/10.17083/ijsg.v10i4.618>
- Vekiri, I. (2010). Socioeconomic differences in elementary students' ICT beliefs and out-of-school experiences. *Computers & Education*, 54(4), 941–950. <https://doi.org/10.1016/j.compedu.2009.09.029>
- Wendt, H., Schwippert, K., Stubbe, T. C., & Jusufi, D. (2020). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund [Mathematical and scientific competencies of students with and without a migration background]. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky, & H. Wendt (Eds.), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 291–314). Waxmann.

Wong, Z. Y., & Liem, G. A. D. (2022). Student engagement: Current state of the construct, conceptual refinement, and future research directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09628-3>

## APPENDIX

**Deutschsprachige Zusammenfassung**

Der Erwerb mathematischer Kompetenzen ist von zentraler Bedeutung für die erfolgreiche Teilhabe an der modernen Gesellschaft (OECD, 2023). Individuen, die über hohe mathematische Kompetenzen verfügen, neigen eher dazu, ihre schulische Ausbildung erfolgreich abzuschließen und anschließend ihre berufliche Karriere erfolgreich zu verfolgen (Bailey et al., 2014; Barbieri et al., 2021; Duncan et al., 2007; OECD, 2019), während Personen mit unzureichend entwickelter mathematischer Kompetenz eher zu Schulabbruch, Arbeitslosigkeit und somit einer erhöhten Wahrscheinlichkeit der Entwicklung von psychischen Erkrankungen neigen (Bynner & Parsons, 2006). Länder wie Deutschland sehen sich jedoch mit einem besorgniserregend starken Rückgang der mathematischen Kompetenzen von Schüler:innen sowie einer zunehmenden Leistungsheterogenität konfrontiert (OECD, 2023; Stanat et al., 2022). Diese beiden ungünstigen Entwicklungen sind alarmierend und es besteht folglich ein erheblicher Bedarf, Lehrkräfte hinsichtlich der Umsetzung eines hochwertigen Mathematikunterrichts auf unterschiedlichen Leistungsniveaus zu unterstützen (Brasiel et al., 2016). Um diesen Herausforderungen zu begegnen, scheint der Einsatz computergestützter Mathematiklernprogramme vielversprechend, da viele Programme die Auswahl der Aufgaben entsprechend dem Leistungsniveau der Lernenden ermöglichen.

Insgesamt weisen metaanalytische Ergebnisse auf das Potenzial von Mathematiklernprogrammen hin, da deren Implementation überwiegend positiv mit der Leistung der Schüler:innen assoziiert ist (vgl. Brasiel et al., 2016; Higgins et al., 2019; Hillmayr et al., 2020; Ma et al., 2014; Ran et al., 2021; Sailer & Homner, 2020). Allerdings sind die Zusammenhänge insgesamt eher klein und die Heterogenität der Befunde in den empirischen Originalstudien ist auffällig (vgl. Byun & Joung, 2018; Tokac et al., 2019). Aufgrund der vielfach nicht signifikanten Effekte wird vermehrt die Wirksamkeit solcher Lernprogramminterventionen infrage gestellt (Pellegrini et al., 2021).

Eine Limitation bei herkömmlichen experimentellen Ansätzen zur Untersuchung der Effekte von Mathematiklernprogrammen besteht jedoch darin, dass Forschende sich meist darauf beschränken, die Programmbereitstellung in einer Experimentalgruppe lediglich mit einer Wartekontrollgruppe zu vergleichen (vgl. Bai et al., 2012; Chang et al., 2015; Shin et al., 2012). Dieser sogenannte Goldstandard-Ansatz gewährleistet zwar wichtige Voraussetzungen für Rückschlüsse auf Kausalität (Eid et al., 2017), führt jedoch dazu, dass bei nicht signifikanten Effekten häufig der Schluss gezogen wird, Mathematiklernprogramme seien möglicherweise nicht geeignet, um Schüler:innen bei der Entwicklung ihrer mathematischen Fähigkeiten zu

unterstützen (Pellegrini et al., 2021). Allerdings mangelt es dieser Vorgehensweise an der erforderlichen Tiefe, um die zugrunde liegenden Wirksamkeitsmechanismen zu verstehen und er bietet darüber hinaus keine überprüfbaren Erklärungen für nicht signifikante Effekte. Daher ergibt sich die Notwendigkeit, traditionelle Fragestellungen und Methoden zu erweitern, um das mögliche Potenzial von Mathematiklernprogrammen weiter zu erforschen.

Ein vielversprechender Ansatz besteht darin, die Interaktion der Schüler:innen mit dem bereitgestellten Programm eingehender zu betrachten. Dabei könnte das variierende Nutzungsverhalten der Schüler:innen eine Schlüsselrolle spielen, da empirische Hinweise bereits darauf hindeuten, dass die Intensität der Nutzung durch die Schüler:innen positiv mit den Effekten des Programms zusammenhängen könnte (Cheung & Slavin, 2013; Ozel et al., 2008). Es reicht somit möglicherweise nicht aus, den Schülern lediglich ein Mathematiklernprogramm anzubieten; vielmehr scheint es von Bedeutung, dass die Schüler:innen über einen längeren Zeitraum regelmäßig mit solchen Lernprogrammen üben, um Lernerfolge zu erzielen. Obwohl diese grundsätzliche Annahme zu erfolgreichem Lernen auch in der Kognitionspsychologie weithin akzeptiert und verbreitet ist (vgl. Gerbier & Toppino, 2015; Toppino & Gerbier, 2014), haben bisherige empirische Studien zu den Effekten von Mathematiklernprogrammen das tatsächliche Nutzungsverhalten der Schüler:innen überwiegend nicht mit einbezogen (vgl. Bai et al., 2012; Brezovszky et al., 2019; Chang et al., 2015; Huang et al., 2014; Hung et al., 2014; Shin et al., 2012). Hinsichtlich der Erfassung des Nutzungsverhaltens bieten (Mathematik-) Lernprogramme aber ein großes Potenzial, da Prozessdaten (z. B. die Anzahl der geübten Aufgaben oder aktive Lernzeit) ökonomisch während des Übens erfasst und gespeichert werden, wodurch behaviorales Engagement objektiv abgebildet werden kann (Azevedo, 2015).

Aus diesen Überlegungen ergeben sich zwei zentrale Ansatzpunkte, die neue vielversprechende Möglichkeiten eröffnen, um das Potenzial von Mathematiklernprogrammen eingehender zu untersuchen. Erstens stellt sich die Frage nach den Prädiktoren des Nutzungsverhaltens der Schüler:innen, da allein die Bereitstellung eines Programms nicht zwangsläufig bedeutet, dass alle Schüler:innen gleichermaßen üben. Daher könnte eine Untersuchung, ob individuelle Schüler:innenmerkmale wie kognitive, persönlichkeitsbezogene, affektiv-motivationale oder familiäre Voraussetzungen mit ihrem Nutzungsverhalten in Verbindung stehen, wichtige Erkenntnisse liefern. Psychologische Theorien, wie die Attentional Control Theory, die Cognitive Load Theory, das situative Erwartungs-Wert-Modell bzw. die Trait Activation Theory (Eccles & Wigfield, 2020; Eysenck et al., 2007; Sweller et al., 1998; Tett et al., 2013) liefern eine theoretische Fundierung für diese

Annahme, da sie Aussagen über Zusammenhänge zwischen individuellen Merkmalen von Person und Verhalten treffen. Ebenso konnten empirische Studien aus der Hausaufgaben- und Unterrichtsforschung beispielsweise bereits zeigen, dass gewissenhafte Schüler:innen bzw. Schüler:innen mit einem höheren mathematischen Selbstkonzept mehr Zeit und Anstrengung hinsichtlich der Aufgabenbearbeitung im Fach Mathematik investierten (vgl. Köller et al., 2000; Trautwein & Lüdtke; 2007; Trautwein et al., 2009). Bisher gibt es jedoch im Kontext des digitalen Lernens keine Untersuchungen zum Zusammenhang von individuellen Schüler:innenmerkmalen und objektiv gemessenem Nutzungsverhalten (d.h. Prozessdaten) in einem Mathematiklernprogramm.

Zweitens könnte variierendes Nutzungsverhalten erklären, warum einige Schüler:innen von der Bereitstellung eines Mathematiklernprogramms hinsichtlich ihrer Leistung profitieren, während andere dies nicht tun. Durch die Betonung der aktiven Auseinandersetzung mit Lerninhalten bei der Nutzung von digitalen Tools stützt die Cognitive Theory of Multimedia Learning (Mayer, 2014) den erwarteten Zusammenhang zwischen dem Nutzungsverhalten der Schüler:innen und deren Mathematikleistung. Um eine ausreichende Auseinandersetzung mit dem Lerninhalt sicherzustellen, wird zwar in den meisten Lernprogramminterventionen üblicherweise darauf geachtet, dass die Schüler:innen eine festgelegte Anzahl von Aufgaben über einen bestimmten Zeitraum hinweg absolvieren (vgl. Brezovszky et al., 2019; Huang et al., 2014; Rodrigo, 2011; Shin et al., 2012; Van den Heuvel-Panhuizen et al., 2013); das Nutzungsverhalten der Schüler:innen könnte jedoch trotzdem variieren. Außerdem kann dieser Ansatz nur begrenzte Informationen darüber liefern, wie die Schüler:innen mit dem Programm interagieren, wenn sie unbeaufsichtigt sind – nicht nur in der Schule, sondern beispielsweise auch zu Hause. Bisher haben nur wenige Studien Schüler:innen ein Mathematiklernprogramm in einer ökologisch validen Umgebung bereitgestellt und das Nutzungsverhalten anhand von objektiven Prozessmaßen berücksichtigt. Die vorhandenen Studien lieferten jedoch erste vielversprechende empirische Hinweise darauf, dass die Mathematikleistung der Schüler:innen zunimmt, je mehr mit dem jeweiligen Programm geübt wurde (Hassler Hallstedt et al., 2018; Jansen et al., 2013; Spitzer, 2022).

Die Annahme, dass Schüler:innen durch die Nutzung von Mathematiklernprogrammen profitieren können, bezieht sich allerdings nicht nur auf die schulische Leistung, sondern auch auf Konstrukte, die eng mit Leistung verknüpft sind, wie akademische Motivation und Leistungsemotionen (z.B. Eccles et al., 1983; Eccles & Wigfield, 2020; Pekrun, 1988; Pekrun, 2006). Basierend auf der Selbstbestimmungstheorie (Deci & Ryan, 1993, 2000; Deci et al., 1991) und der Kontroll-Wert-Theorie (Pekrun, 2006) wäre es plausibel anzunehmen, dass ein



Programm durch bestimmte Merkmale, wie Adaptivität und Feedback, die akademische Motivation der Schüler:innen steigern und negative Leistungseemotionen verringern könnte. Dennoch existiert bislang nur sehr begrenztes Wissen darüber, wie Mathematiklernprogramme, insbesondere im Hinblick auf das Nutzungsverhalten der Schüler:innen, auf deren akademische Motivation bzw. deren Leistungseemotionen wirken (vgl. Dondio et al., 2023; Fadda et al., 2022; Jansen et al., 2013).

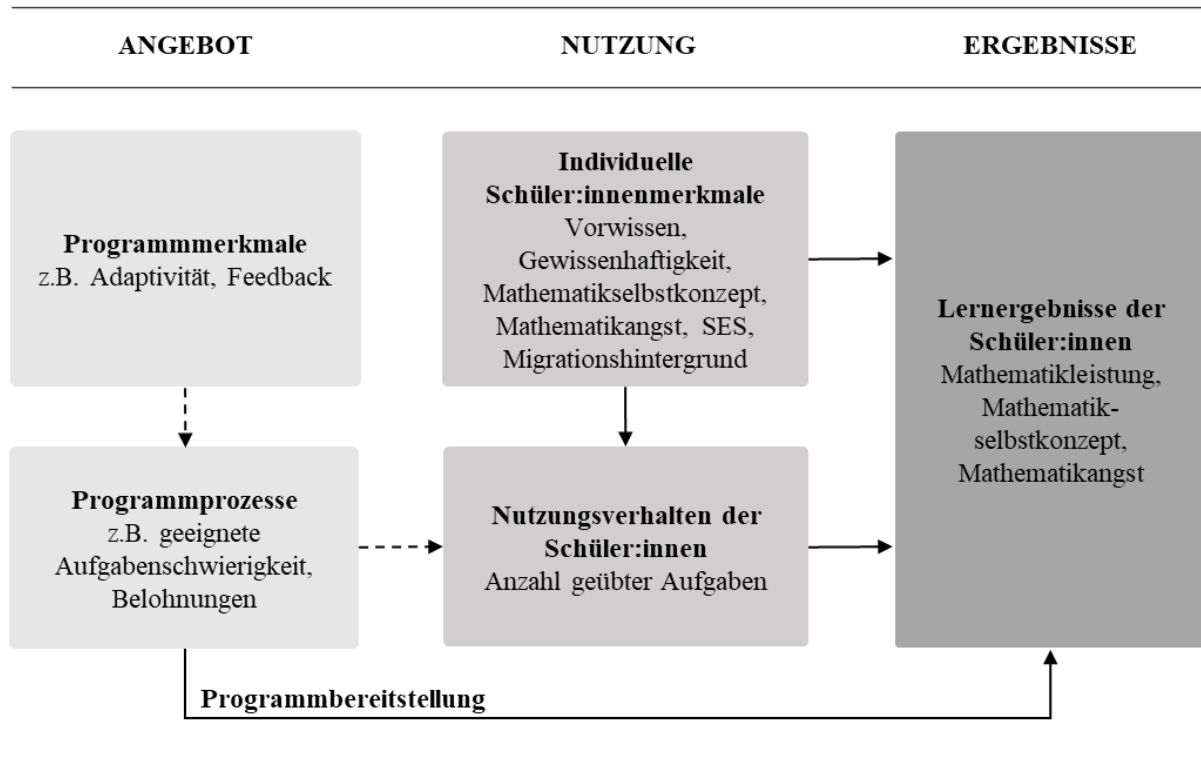
### **Fragestellungen der Dissertation**

Im Einklang mit den aufgezeigten Forschungsdesideraten lieferte ein an den Kontext von Mathematiklernprogrammen angepasstes Angebots-Nutzungs-Modell den heuristischen Rahmen für diese Dissertation (Helmke & Weinert, 1997; Seidel, 2014; siehe Abbildung 1). Das Modell unterstützt (1) die Annahme, dass es Zusammenhänge zwischen den individuellen Merkmalen der Schüler:innen und ihrem Nutzungsverhalten mit einem Mathematiklernprogramm gibt. Hinsichtlich der individuellen Merkmale konzentriert sich diese Dissertation aufgrund deren Einbettung in das Angebots-Nutzungs-Modell sowie deren evidenter Relevanz in der Bildungsforschung (Barroso et al., 2021; Chiu et al., 2012; Dong et al., 2020; Kriegbaum et al., 2018; Poropat, 2009; Tomaszewski et al., 2020) auf das mathematische Vorwissen der Schüler:innen, relevante Persönlichkeitsmerkmale (Gewissenhaftigkeit), die akademische Motivation (Mathematikselbstkonzept) und Leistungseemotionen (Mathematikangst) sowie den familiären Hintergrund, d.h. einen Migrationshintergrund bzw. den sozioökonomischen Status (SES).

Außerdem unterstützt das Angebots-Nutzungs-Modell (2) die Annahme, dass das Nutzungsverhalten der Schüler:innen mit ihrer Mathematikleistung, aber auch ihrer akademischen Motivation und ihren Leistungseemotionen in Verbindung steht. Neben Effekten auf die Mathematikleistung sollen in diese Dissertation daher auch Effekte auf das Mathematikselbstkonzept sowie auf die Mathematikangst der Schüler:innen berücksichtigt werden, da beide Konstrukte substantiell mit zukünftiger Leistungen verbunden sind (Marsh & Martin, 2011; Ramirez et al., 2018).

**Abbildung 1**

*Heuristisches Rahmenmodell zur Darstellung der zentralen Variablen: Individuelle Merkmale der Schüler:innen, deren Nutzungsverhalten mit einem Mathematiklernprogramm und mögliche Lernergebnisse*



*Anmerkungen.* Der Rahmen dieser Dissertation basiert auf einem an den Kontext von Mathematiklernprogrammen angepassten Angebots-Nutzungs-Modell (Helmke & Weinert, 1997; Seidel, 2014). Gestrichelte Linien stellen theoretische Annahmen dar; durchgängige Linien beziehen sich auf die in dieser Dissertation untersuchten Zusammenhänge.

Zusammenfassend bestand das Hauptziel dieser Dissertation darin, die Rolle des Nutzungsverhaltens der Schüler:innen hinsichtlich eines Mathematiklernprogramms zu untersuchen. Um dieses Ziel zu erreichen, wurden in dieser Dissertation zwei Forschungsfragen postuliert:

**Forschungsfrage 1:** Gibt es einen Zusammenhang zwischen individuellen Schüler:innenmerkmalen und dem Nutzungsverhalten der Schüler:innen in einem Mathematiklernprogramm?

**Forschungsfrage 2:** Wie hängen die Bereitstellung eines Mathematiklernprogramms und das Nutzungsverhalten der Schüler:innen mit diesem Programm mit deren Mathematikleistung, ihrem Mathematikselbstkonzept und ihrer Mathematikangst zusammen?

Um die beiden theoretischen Annahmen zu überprüfen, wurden drei empirische Studien und zwei zusätzliche Replikationsanalysen durchgeführt. Studie 1 und deren Replikation sowie Teile von Studie 3 tragen zur Beantwortung von Forschungsfrage 1 bei, während Studie 2 und deren Replikation sowie Teile von Studie 3 zur Beantwortung von Forschungsfrage 2 dienen. Alle Untersuchungen basieren auf dem Arithmetiklernprogramm Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014; Deutsch: Prowise Learn), welches für diese Dissertation zweimal in großen deutschen Stichproben implementiert wurde, in beiden Fällen als Bestandteil eines randomisierten Pretest-Posttest-Kontrollgruppendesigns. Die erste Implementation umfasste eine Übungsperiode von 22 Wochen im Schuljahr 2019/2020 (Studie 1, Studie 2); in der zweiten Implementation übten die Schüler:innen 45 Wochen lang mit Math Garden im Schuljahr 2021/2022 (Replikation Studie 1, Replikation Studie 2, Studie 3).

Math Garden selbst zeichnet sich durch seinen adaptiven Algorithmus aus und kann als intelligentes Drill und Practice Programm verstanden werden (Nattland & Kerres, 2009). Es integriert Instructional Design Elemente, etwa ein Belohnungssystem, und hat dementsprechend einen game-based Charakter (Plass et al., 2015; Plass et al., 2020). Außerdem bietet Math Garden den großen Vorteil, dass während des Übens relevante Prozessdaten erfasst und gespeichert werden, welche als Indikatoren für das Nutzungsverhalten der Schüler:innen dienen können. Im konkreten Fall handelt es sich in dieser Dissertation um die Anzahl geübter Aufgaben.

### **Studie 1 und zusätzliche Replikationsanalysen: Individuelle Schüler:innenmerkmale als Prädiktoren für deren Nutzungsverhalten**

Hinsichtlich der Beantwortung der ersten Forschungsfrage untersuchte Studie 1, wie individuelle Schüler:innenmerkmale mit deren Nutzungsverhalten in Math Garden über einen Zeitraum von 22 Wochen (Implementationsperiode 1) zusammenhingen. Basierend auf den Ergebnissen vorheriger Forschungen in verschiedenen Lernkontexten (z. B. Hausaufgabenverhalten) wurden die Hypothesen aufgestellt, dass Schüler:innen mit mehr mathematischem Vorwissen, höherem Maß an Gewissenhaftigkeit, einem höheren Mathematikselbstkonzept und einem höheren SES mehr mit Math Garden üben würden, während Schüler:innen mit mehr Mathematikangst bzw. Schüler:innen mit einem Migrationshintergrund weniger mit dem Programm üben würden (Chiu et al., 2012; Dong et al., 2020; Dumfart & Neubauer, 2016; LeFevre et al., 1992; Trautwein & Lüdtke, 2007). Die individuellen Merkmale der Schüler:innen wurden in einem Pretest erfasst und ihr Nutzungsverhalten wurde als die Anzahl der wöchentlich geübten Aufgaben operationalisiert.

Basierend auf diesen Längsschnittdaten wurde eine Mehrebenenanalyse durchgeführt, da die Messzeitpunkte der geübten Aufgaben innerhalb der Schüler:innen ( $N=203$ ) geclustert waren. Mit denselben theoretischen Annahmen und methodischen Ansätzen wurde Studie 1 in der zweiten Implementationsperiode repliziert, um die Robustheit der Ergebnisse zu prüfen ( $N=890$ ).

Die Ergebnisse aus Studie 1 zeigten, dass das Nutzungsverhalten der Schüler:innen über die 22 Wochen abnahm und vor allem in den Ferien besonders niedrig war. Hinsichtlich der Hauptfragestellung der Studie, die die Unterschiede zwischen den Schüler:innen betraf, stellte sich heraus, dass besonders mathematikängstliche Schüler:innen im Durchschnitt weniger Aufgaben mit Math Garden übten. Für mathematisches Vorwissen, Gewissenhaftigkeit, Mathematikselbstkonzept, SES und einen Migrationshintergrund war der Zusammenhang mit der Anzahl geübter Aufgaben statistisch nicht signifikant.

Die Ergebnisse der Replikationsanalysen stimmten weitgehend mit denen aus Studie 1 überein. Das Nutzungsverhalten der Schüler:innen nahm ebenfalls im Laufe der Zeit ab und es wurde in Ferienzeiten signifikant weniger geübt. Deskriptiv zeigten sich für Mathematikangst und einen Migrationshintergrund negative Zusammenhänge mit der Nutzungshäufigkeit. Im statistischen Modell bestätigte sich dieser Zusammenhang jedoch nur für den Migrationshintergrund. Das bedeutet, dass Schüler:innen mit einem Migrationshintergrund im Durchschnitt weniger mit Math Garden übten. Dementsprechend deuten die Ergebnisse von Studie 1 und der Replikationsanalyse darauf hin, dass besonders mathematikängstliche Schüler:innen und Schüler:innen mit einem Migrationshintergrund mehr initiale Unterstützung beim Üben benötigen, um ein ähnliches Nutzungsniveau wie ihre Mitschüler:innen mit geringerer Mathematikangst bzw. ohne Migrationshintergrund zu erreichen. Mögliche Erklärungen für das variierende Befundmuster werden in der Gesamtdiskussion aufgegriffen.

### **Studie 2 und zusätzliche Replikationsanalysen: Effekte der Programmbereitstellung von und des Nutzungsverhaltens mit Math Garden auf die Lernergebnisse der Schüler:innen**

Um die zweite Forschungsfrage der Dissertation zu beantworten, konzentrierte sich Studie 2 auf die Effekte von Math Garden hinsichtlich der Lernergebnisse der Schüler:innen. Mit einem randomisierten Pretest-Posttest-Kontrollgruppendesign ( $N=376$ ) wurden nicht nur die Auswirkungen der Programmbereitstellung untersucht, sondern auch, wie sich das Nutzungsverhalten über den Bereitstellungszeitraum von 22 Wochen (Implementationsperiode 1) innerhalb der Experimentalgruppe auf die Lernergebnisse der Schüler:innen auswirkte. Dabei sollte bisherige Forschung durch einen differenzierteren Ansatz erweitert werden (Ran

et al., 2021). In Bezug auf die Lernergebnisse der Schüler:innen bedeutete dies, dass getrennte Schlussfolgerungen über deren Addition- und Subtraktionsleistung gezogen wurden, im Gegensatz zu Schlussfolgerungen hinsichtlich eines Gesamtscores. Außerdem wurden neben der Leistung auch Effekte auf das Mathematikselbstkonzept und die Mathematikangst berücksichtigt. Das Nutzungsverhalten wurde nicht nur über die Anzahl geübter Aufgaben operationalisiert, sondern auch hinsichtlich der Anzahl der Wochen, über die die Schüler:innen mit Math Garden geübt haben. Im Einklang mit vorheriger Forschung und theoretischer Annahmen über die vorteilhaften Merkmale von Adaptivität und Feedback (Dondio et al., 2023; Fadda et al., 2022; Jansen et al., 2013; Mayer, 2014; Marsh & Martin, 2011; Pekrun, 2006) wurden die Hypothesen aufgestellt, dass die Bereitstellung von Math Garden ebenso wie die Anzahl der geübten Aufgaben und die Anzahl der geübten Wochen, die Additions- und Subtraktionsleistung der Schüler:innen verbessern würde, das Mathematikselbstkonzept steigern und die Mathematikangst verringern würde. Die Hypothesen wurden mit Regressionsanalysen statistisch geprüft. Wie Studie 1 wurde auch Studie 2 mit Daten aus der zweiten Implementationsperiode repliziert, um die Robustheit der Ergebnisse unter Verwendung derselben theoretischen Annahmen und methodischen Ansätze zu prüfen.

Die Ergebnisse aus Studie 2 unterstützten die angenommenen Hypothesen nur teilweise. Hinsichtlich der Additions- und Subtraktionsleistung der Schüler:innen gab es in der Experimentalgruppe im Gegensatz zur Wartekontrollgruppe keine statistisch signifikanten Effekte bezüglich der Programmbereitstellung. Allerdings war die Anzahl geübter Subtraktionsaufgaben ein positiver Prädiktor hinsichtlich der Subtraktionsleistung der Schüler:innen in der Experimentalgruppe. Der erwartete Zusammenhang zwischen der Anzahl geübter Additionsaufgaben und der Additionsleistung zeigte sich jedoch nicht. Weiterhin konnten keine Zusammenhänge zwischen der Anzahl geübter Wochen und der Additions- bzw. Subtraktionsleistung festgestellt werden. Für das Mathematikselbstkonzept zeigte sich ein signifikanter Effekt der Programmbereitstellung in der Experimentalgruppe im Vergleich zur Wartekontrollgruppe, wobei sich für die Indikatoren des Nutzungsverhaltens (Anzahl geübter Aufgaben bzw. Wochen) in der Experimentalgruppe allerdings keine Effekte ergaben. In Bezug auf die Mathematikangst hatten weder die Programmbereitstellung noch das Nutzungsverhalten einen Effekt.

Die Ergebnisse der Replikationsanalyse stimmten teilweise mit denen aus Studie 2 überein. Sie zeigten ebenfalls, dass die von den Schüler:innen geübte Anzahl an Subtraktionsaufgaben signifikant mit deren Subtraktionsleistung zusammenhingen. Im Gegensatz zu Studie 2 wurde jedoch kein Effekt der Bereitstellung von Math Garden auf das

Mathematikselbstkonzept gefunden. Mögliche Erklärungen hinsichtlich des unterschiedlichen Befundmusters werden in der Gesamtdiskussion aufgegriffen. Für alle anderen angenommenen Zusammenhänge konnten, ebenso wie in Studie 2, keine statistisch signifikanten Effekte gezeigt werden.

Studie 2 und die zusätzlichen Replikationsanalysen liefern somit Hinweise auf die Relevanz der Berücksichtigung von tatsächlichem Nutzungsverhalten, um besser zu verstehen, wie Schüler:innen von der Implementation von Mathematiklernprogrammen hinsichtlich ihrer Leistungen profitieren können. Die unterschiedlichen Effekte hinsichtlich der einzelnen Arithmetikdomänen könnten ein Indikator dafür sein, dass Schüler:innen eher vom Üben schwierigerer Aufgaben (d.h. Subtraktion; Anderson et al., 2020) profitieren. Außerdem könnten die Ergebnisse darauf hindeuten, dass Mathematiklernprogramme das Potenzial haben, das Mathematikselbstkonzept der Schüler:innen zu steigern. Aufgrund des variierenden Befundmusters bedarf es jedoch weiterer Forschung, um diese Frage abschließend zu beantworten.

### **Studie 3: Die mediierende Rolle des Nutzungsverhaltens hinsichtlich des Zusammenhangs zwischen Mathematikangst und Mathematikleistung**

Studie 3 kombinierte die Überlegungen aus Studie 1 und Studie 2 und adressierte somit beide Forschungsfragen dieser Dissertation. Basierend auf verhaltens- und emotionstheoretischen Annahmen, dass insbesondere mathematikängstliche Schüler:innen mathematikbezogene Aufgaben vermeiden (Eysenck & Calvo, 1992; Eysenck et al., 2007), zielte diese Studie darauf ab zu untersuchen, ob mathematikängstliche Schüler:innen Math Garden weniger nutzen und ob dies wiederum die Verbesserung ihrer Mathematikleistung hemmen würde. Das Nutzungsverhalten wurde, wie auch in Studie 1 und Studie 2, objektiv mit Prozessdaten in Form der Anzahl geübter Aufgaben operationalisiert. Dieser Ansatz erweitert bisherige Forschung, da Vermeidungsverhalten mathematikängstlicher Schüler:innen bisher nur mit möglicherweise verzerrten Selbstberichtsfragebögen erhoben wurde (siehe Hasty et al., 2021; Quintero et al., 2022). Es wurden die Hypothesen aufgestellt, dass mathematikängstliche Schüler:innen weniger Aufgaben in Math Garden üben, aber auch, dass mehr geübte Aufgaben mit einer besseren Leistung einhergehen würden. Daher könnte ein niedrigeres Nutzungsverhalten die angenommene negative Beziehung zwischen Mathematikangst und Mathematikleistung erklären. Die Studie umfasste drei Messzeitpunkte innerhalb der zweiten Implementationsperiode: zu Beginn, in der Mitte und am Ende eines Schuljahres ( $T_1$ – $T_2$ : 28 Wochen Programmnutzung,  $T_1$ – $T_3$ : 45 Wochen Programmnutzung). Im ersten Messzeitpunkt

beantworteten die Schüler:innen Mathematikangstfragebögen (Mathematiktestangst, Mathematikrechenangst, kognitive Sorgenkomponente) und zu allen drei Messzeitpunkten wurde ein Mathematiktest absolviert, der Rechenoperationen in den vier Grundrechenarten testete. Es wurden Mediationsanalysen durchgeführt, um die angenommenen Hypothesen zu testen.

Die Ergebnisse unterstützten die Hypothesen jedoch nur teilweise. Das deskriptive Befundmuster zeigte zwar, dass ein Mathematikangstgesamtscore, ebenso wie die Subkomponenten Mathematiktestangst und die kognitive Sorgenkomponente, negativ mit der Anzahl geübter Aufgaben in Math Garden korreliert war; allerdings ergaben sich auch ein negativer Zusammenhang hinsichtlich der Kontrollvariable Migrationshintergrund. Im statistischen Modell zeigte sich jedoch nur der Migrationshintergrund als relevanter negativer Prädiktor. Dieses Ergebnis deutete darauf hin, dass Schüler:innen mit einem Migrationshintergrund weniger übten. Im Einklang mit der aufgestellten Hypothese verbesserten sich aber die Mathematikleistungen der Schüler:innen, je mehr Aufgaben sie übten. Es stellte sich jedoch heraus, dass die Anzahl geübter Aufgaben in Math Garden die negative Beziehung zwischen Mathematikangst und Mathematikleistung nicht mediierte.

Zusammenfassend lässt sich somit sagen, dass die Ergebnisse von Studie 3 auf die Relevanz einer regelmäßigen Nutzung von Mathematiklernprogrammen hinweisen, denn konsequentes Üben scheint, wie auch bereits in Studie 2 angedeutet, eine wichtige Rolle für die mathematische Leistungsverbesserung zu spielen. Die Rolle der Mathematikangst bleibt zu diesem Zeitpunkt jedoch unklar und muss weiter untersucht werden.

### **Gesamtdiskussion**

Um die Rolle des Nutzungsverhaltens von Schüler:innen bei der Implementation von Mathematiklernprogrammen zu untersuchen, wurden in dieser Dissertation zwei übergreifende Forschungsfragen gestellt. Erstens wurden, basierend auf mehreren etablierten psychologischen Theorien (Attentional Control Theory, Cognitive Load Theory, situatives Erwartungs-Wert-Modell, Trait Activation Theory; Eccles & Wigfield, 2020; Eysenck et al., 2007; Sweller et al., 1998; Tett et al., 2013), in Studie 1 und deren Replikation sowie Teilen von Studie 3 untersucht, wie individuelle Merkmale der Schüler:innen, d.h. mathematisches Vorwissen, Gewissenhaftigkeit, Mathematikselbstkonzept, Mathematikangst, SES und ein Migrationshintergrund, mit der Anzahl der in Math Garden geübten Aufgaben zusammenhängen. Die Ergebnisse können dazu beitragen, bestimmte Risikogruppen von Schüler:innen zu identifizieren, die möglicherweise weniger üben und daher mehr

Unterstützung benötigen, um die Möglichkeit zu erhalten, von Mathematiklernprogrammimplementationen profitieren zu können.

Tatsächlich zeigte Studie 1, dass insbesondere mathematikängstliche Schüler:innen im Durchschnitt weniger Aufgaben in Math Garden übten. Die Replikationsanalysen von Studie 1 und die Ergebnisse von Studie 3 ergaben jedoch, dass Mathematikangst zwar negativ mit der Anzahl der geübten Aufgaben in Math Garden zusammenhing, insbesondere Schüler:innen mit einem Migrationshintergrund im Durchschnitt aber weniger übten. Eine mögliche Erklärung für dieses Befundmuster könnte das Zusammenspiel von zwei Faktoren sein, nämlich Stichprobenmerkmalen bezogen auf die Ausprägung von Mathematikangst und die Art der Erfassung des Migrationshintergrunds der Schüler:innen. Hinsichtlich des ersten Erklärungsfaktors berichteten die Schüler:innen in der zweiten Implementationsperiode insgesamt ein geringeres Level an Mathematikangst. Dieses niedrigere Mathematikangstniveau könnte dementsprechend mit weniger negativen Gefühlen beim Üben verbunden gewesen sein, was wiederum zu weniger Vermeidungsverhalten geführt haben könnte (Pekrun, 2006). Im Hinblick auf den zweiten Erklärungsfaktor unterschied sich die Erfassung des Migrationshintergrunds zwischen den Implementationsperioden. Aufgrund des breiteren Antwortformats<sup>1</sup> in der zweiten Implementationsperiode könnte eine substanziell aussagekräftigere Varianz geschaffen worden sein, die wiederum eher Unterschiede im Hinblick auf das Nutzungsverhalten der Schüler:innen hat erklären können. Zusammenfassend scheint es, dass bei einem geringeren Mathematikangstlevel der Migrationshintergrund der Schüler:innen möglicherweise eine relevantere Rolle hinsichtlich des Nutzungsverhaltens spielt.

In Bezug auf mathematisches Vorwissen, Gewissenhaftigkeit, Mathematikselbstkonzept und SES war der Zusammenhang mit dem Nutzungsverhalten in Studie 1 und der Replikation nicht signifikant. Dieses robuste Muster von Ergebnissen in zwei unterschiedlichen Stichproben könnte darauf hindeuten, dass es möglicherweise keine Zusammenhänge zwischen den untersuchten individuellen Schüler:innenmerkmalen und dem Üben mit Math Garden gab. Dieses Ergebnis wäre durchaus als positiv zu bewerten, da es bedeutet, dass Schüler:innen mit diesen Merkmalen (d.h. geringeres Mathematikvorwissen, geringere Ausprägung an Gewissenhaftigkeit, geringeres Mathematikselbstkonzept und geringerer SES) das Üben mit Math Garden zumindest nicht vermieden haben und die Implementation des Programms sie

---

<sup>1</sup>Implementationsperiode 1: „Werden bei dir Zuhause noch andere Sprachen außer Deutsch gesprochen“ mit dichotomem Antwortformat; Implementationsperiode 2: „Wie oft sprichst du Zuhause Deutsch?“ mit vierstufiger Likert-Skala als Antwortformat



dementsprechend nicht hinsichtlich ihres Nutzungsverhaltens zu benachteiligen schien. Um die Ergebnisse allerdings zu generalisieren, sind weitere Untersuchungen erforderlich, die verschiedene Stichprobenmerkmale (z.B. Altersgruppe oder Schultyp) und andere Mathematiklernprogramme berücksichtigen, da bestimmte Programmmerkmale von Math Garden, wie Adaptivität und Feedback, die fehlenden Effekte erklären könnten.

Die zweite Forschungsfrage in dieser Dissertation, die sich auf die Cognitive Theory of Multimedia Learning, die Selbstbestimmungstheorie und die Kontroll-Wert-Theorie stützte (Deci et al., 1991; Mayer, 2014; Pekrun, 2006), untersuchte die Effekte von Math Garden auf die Lernergebnisse der Schüler:innen. Dabei wurden die (Additions- und Subtraktions-) Leistung, das Mathematikselbstkonzept und die Mathematikangst der Schüler:innen berücksichtigt. Neben der Untersuchung der Effekte der reinen Programmbereitstellung (Studie 2 und Replikationsanalysen) wurde auch das Nutzungsverhalten der Schüler:innen berücksichtigt (Studie 2 und Replikationsanalyse: Anzahl geübter Aufgaben und Anzahl der Wochen, in denen mit Math Garden geübt wurde hinsichtlich aller Lernergebnisse; Studie 3: Anzahl geübter Aufgaben hinsichtlich der Mathematikleistung). Entgegen den Erwartungen zeigten sich in Studie 2 keine statistisch signifikanten Effekte der Programmbereitstellung auf die Addition- oder Subtraktionsleistung der Schüler:innen in der Experimentalgruppe im Vergleich zur Wartekontrollgruppe. Allerdings verbesserte sich wie angenommen die Subtraktionsleistung umso mehr, je mehr die Schüler:innen in der Experimentalgruppe Subtraktionsaufgaben übten. Dieses Ergebnis konnte in den zusätzlichen Analysen zu Studie 2 repliziert werden. Ergebnisse aus Studie 3 unterstützen diesen angenommenen Zusammenhang ebenfalls, d. h. die Anzahl der bearbeiteten Aufgaben stand in positivem Zusammenhang mit der Leistung der Schüler:innen, die anhand eines Gesamtscores der Grundrechenarten Addition, Subtraktion, Multiplikation und Division operationalisiert wurde.

Die Zusammenhänge zwischen dem Nutzungsverhalten der Schüler:innen und deren Leistungsverbesserung deuteten wie erwartet weitestgehend darauf hin, dass Schüler:innen ein Mathematiklernprogramm aktiv nutzen müssen, um davon profitieren zu können. Allerdings fiel insgesamt die große Varianz im Nutzungsverhalten der Schüler:innen und die eher geringe durchschnittliche Anzahl geübter Aufgaben in allen drei Untersuchungen auf (d. h. Studie 2, Replikation von Studie 2, Studie 3). Diese deskriptiven Beobachtungen könnten darauf hindeuten, dass die Schüler:innen der Experimentalgruppe im Durchschnitt zu wenig mit Math Garden geübt haben, weshalb in Studie 2 und deren Replikation keine Effekte bezüglich der Programmbereitstellung auftraten. Diese Vermutung bedarf jedoch weiterer Überprüfung. Der domänenspezifische Effekt der Anzahl geübter Subtraktionsaufgaben könnte weiterhin darauf

hindeuten, dass Math Garden eher dafür geeignet ist schwierige Arithmetikdomänen zu fördern (Anderson et al., 2022). Fehlende Effekte der Anzahl geübter Wochen mit Math Garden könnten darauf zurückzuführen sein, dass alle Schüler:innen insgesamt nur wenige Wochen mit dem Programm übten. Dies könnte möglicherweise zu einer unzureichenden Erzeugung von aussagekräftiger Varianz geführt haben.

Für das Mathematikselbstkonzept der Schüler:innen ergab sich in Studie 2 tatsächlich ein positiver Effekt der Programmbereitstellung in der Experimentalgruppe im Vergleich zur Wartekontrollgruppe. Allerdings zeigte sich für Schüler:innen der Experimentalgruppe kein erwarteter Zusammenhang zwischen der Nutzungshäufigkeit und dem Mathematikselbstkonzept. Entgegen der theoretischen Annahmen könnte dies bedeuten, dass der Effekt der Programmbereitstellung unabhängig von der Menge an positivem Feedback zu sein schien, welches die Schüler:innen während des Übens erhielten (Marsh & Martin, 2011). Diese Annahme müsste eingehender geprüft werden. Der positive Effekt der Bereitstellung von Math Garden auf das Mathematikselbstkonzept der Schüler:innen konnte jedoch in der zusätzlichen Analyse nicht repliziert werden. Ein Grund für dieses konträre Ergebnis könnte der Einsatz einer anderen Skala im Gegensatz zu Studie 2 gewesen sein. Somit scheint der Effekt nicht stabil über verschiedene Operationalisierungen hinweg zu sein. Ferner könnten Stichprobenmerkmale ebenfalls das nicht signifikante Ergebnis erklären, da die Schüler:innen in der Replikationsanalyse im Pretest im Durchschnitt deskriptiv höhere Mathematikselbstkonzeptwerte aufwiesen als die Schüler:innen in Studie 2. Die Mechanismen von Adaptivität und Feedback, die als theoretische Mechanismen für einen erwarteten positiven Effekt postuliert wurden, könnten daher weniger relevant gewesen sein, da die Schüler:innen im Durchschnitt bereits ein positiveres Bild ihrer mathematischen Fähigkeiten hatten.

Hinsichtlich der Mathematikangst der Schüler:innen hatten weder die Programmbereitstellung noch das Nutzungsverhalten einen Effekt (Studie 2 und Replikation). Ein Grund für die fehlenden Effekte wäre möglicherweise, dass mathematikängstliche Schüler:innen insgesamt zu wenig mit Math Garden übten, um die durch den adaptiven Algorithmus erhöhte Kontrolle über das erfolgreiche Lösen einer Aufgabe wahrzunehmen (Pekrun, 2006). Weiterhin wird betont, dass Mathematikangst oft schon in jungen Jahren beginnt und sich im Laufe der Zeit tief in den Schüler:innen verankert (Ashcraft & Krause, 2007; Bekdemir, 2010; Ramirez et al., 2013). Selbst wenn die Erfahrungen der Schüler:innen mit Math Garden aufgrund des adaptiven Algorithmus und des Feedbacks positiv waren, überwog dies möglicherweise nicht über bereits erlebte negative Erfahrungen aus der Vergangenheit (Jansen et al. 2013). Außerdem bezogen sich die Fragen zu Mathematikangst

explizit auf die Komponente Mathematiktestangst. Math Garden könnte aber weniger als Test, sondern mehr als Übungstool wahrgenommen worden sein, weshalb anzunehmen wäre, dass das Programm effektiver auf die Reduktion anderer Mathematikangstkomponenten, wie etwa Rechenangst, wirken könnte (Lukowski et al., 2019)

Bei der Interpretation aller Ergebnisse müssen Limitationen der vorliegenden Arbeit berücksichtigt werden. So ist nicht klar, in welchem Kontext die Schüler:innen überwiegend mit Math Garden übten, d. h. in der Schule oder zu Hause. Bestimmte kontextbezogene Faktoren könnten jedoch die Art und Weise, wie Schüler:innen mit dem Programm übten, beeinflusst haben. So könnten bei der Nutzung in der Schule lehrkräftebezogene Faktoren eine Rolle gespielt haben, während bei der Nutzung zu Hause elternbezogene Faktoren möglicherweise relevant waren (Wong & Liem, 2022). Weiterhin muss darauf hingewiesen werden, dass die zugrundeliegenden Stichproben der Untersuchungen ausschließlich Schüler:innen aus Gemeinschaftsschulen, d. h. Schüler:innen mit einem durchschnittlich niedrigeren Leistungsniveau, umfasste. Die Ergebnisse sind folglich nicht auf alle Leistungsniveaus übertragbar. Hinsichtlich des Programms Math Garden ist anzumerken, dass es hauptsächlich Aufgaben zu arithmetischen Operationen enthält, die im Allgemeinen weniger komplex sind. Darüber hinaus arbeitet das Programm adaptiv und bietet korrekatives Feedback. Daher könnten die erzielten Effekte in erster Linie auf diese spezifischen Programmcharakteristika und -merkmale zurückzuführen sein und deren Übertragbarkeit auf andere Programme müsste geprüft werden.

Trotz der genannten Limitationen können die Ergebnisse zu praktischen Implikationen hinsichtlich der Implementation von Mathematiklernprogrammen beitragen. Einerseits zeigen die Ergebnisse, dass Schüler:innen mit bestimmten Merkmalen, wie mathematikängstliche Schüler:innen sowie Schüler:innen mit einem Migrationshintergrund, nach der Bereitstellung eines Mathematiklernprogramms mehr zusätzliche Unterstützung benötigen. Ohne eine solche Unterstützung üben diese Schüler:innen möglicherweise weniger und wären somit hinsichtlich des potenziellen Programmnutzens benachteiligt. Im Hinblick auf die Unterstützung von mathematikängstlichen Schüler:innen könnte darauf fokussiert werden, die wahrgenommene Kontrolle dieser Schüler:innen zu erhöhen (Pekrun, 2006), etwa durch eine manuelle Herabsetzung der Itemschwierigkeit. Bezüglich des Migrationshintergrunds könnte beim Abbau einer möglicherweise vorhandenen Sprachbarriere, ebenso wie bei der Bereitstellung eines geeigneten digitalen Endgeräts angesetzt werden (Portes & Rumbaut, 2001; Van de Werfhorst et al., 2022). Dies scheint vor allem essenziell, um soziale Disparitäten nicht weiter zu verstärken. Andererseits deuten die Ergebnisse darauf hin, dass es möglicherweise nicht

ausreicht, den Schüler:innen lediglich ein Mathematiklernprogramm zur Verfügung zu stellen, damit sie hinsichtlich ihrer Leistung davon profitieren können. Die Aufgabe der Lehrkraft wäre es daher, die Schüler:innen dabei zu unterstützen, kontinuierlich und systematisch mit dem bereitgestellten Mathematiklernprogramm zu üben (Spitzer & Musslick, 2021).

Zukünftige Forschung könnte darauf abzielen, tiefergehende Erkenntnisse in die Prozesse von (Mathematik-) Lernprogrammen zu gewinnen und die zugrunde liegenden Programmmechanismen weiter zu erforschen. In diesem Kontext könnte mittels eines Experience-Sampling-Designs untersucht werden, inwieweit Schüler:innen die Adaptivität, d.h. die Aufgabenschwierigkeit, wahrnehmen und wie sich dies auf ihre Leistung, Motivation und Emotionen während des Übens auswirkt. Darüber hinaus wäre anzunehmen, dass die Lehrkraft eine entscheidende Rolle bei der erfolgreichen Implementation von (Mathematik-) Lernprogrammen spielt. Methodische Ansätze aus dem Bereich des Learning Analytics (vgl. Khalil & Ebner, 2017; Kizilcec et al., 2017) könnten genutzt werden, um Prozessdaten weiter zu analysieren und Aussagen zur Implementationsqualität von (Mathematik-) Lernprogrammen auf Klassenebene zu treffen.

## Literaturangaben

- Anderson, S., Sommerhoff, D., Schurig, M., Ufer, S., & Gebhardt, M. (2022). Developing learning progress monitoring tests using difficulty-generating item characteristics. An example for basic arithmetic operations in primary schools. *1866-6671*, *14*.  
<https://doi.org/10.25656/01:24875>
- Ashcraft, M. H., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243–248. <https://doi.org/10.3758/BF03194059>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology*, *43*(6), 993–1003.  
<https://doi.org/10.1111/j.1467-8535.2011.01269.x>
- Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775–785.  
<https://doi.org/10.1111/desc.12155>

- Barbieri, C. A., Young, L. K., Newton, K. J., & Booth, J. L. (2021). Predicting middle school profiles of algebra performance using fraction knowledge. *Child Development, 92*(5), 1984–2005. <https://doi.org/10.1111/cdev.13568>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. *Educational Studies in Mathematics, 75*(3), 311–328. <https://doi.org/10.1007/s10649-010-9260-7>
- Brasiel, S., Jeong, S., Ames, C., Lawanto, K., Yuan, M., & Martin, T. (2016). Effects of educational technology on mathematics achievement for K-12 students in Utah. *Journal of Online Learning Research, 2*(3), 205–226.
- Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education, 128*, 63–74. <https://doi.org/10.1016/j.compedu.2018.09.011>
- Bynner, J., & Parsons, S. (2006). *Does numeracy matter more?* National research and development centre for adult literacy and numeracy.
- Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics, 118*(3-4), 113–126. <https://doi.org/10.1111/ssm.12271>
- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International, 52*(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Chiu, M. M., Pong, S., Mori, I., & Chow, B. W.-Y. (2012). Immigrant students' emotional and cognitive engagement at school: A multilevel analysis of students in 41 countries. *Journal of Youth and Adolescence, 41*(11), 1409–1425. <https://doi.org/10.1007/s10964-012-9763-x>

- Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39, 223–238.  
<https://doi.org/10.25656/01:11173>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.  
[https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, 26(3-4), 325–346. <https://doi.org/10.1080/00461520.1991.9653137>
- Dondio, P., Gusev, V., & Rocha, M. (2023). Do games reduce maths anxiety? A meta-analysis. *Computers & Education*, 194, 104650.  
<https://doi.org/10.1016/j.compedu.2022.104650>
- Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, 11, 591203. <https://doi.org/10.3389/fpsyg.2020.591203>
- Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences*, 37(1), 8–15. <https://doi.org/10.1027/1614-0001/a000182>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motivation* (pp. 75–146). Freeman.
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859.  
<https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Beltz.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6), 409–434.  
<https://doi.org/10.1080/02699939208409696>

- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353.  
<https://doi.org/10.1037/1528-3542.7.2.336>
- Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2022). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning*, 38(1), 304–325. <https://doi.org/10.1111/jcal.12618>
- Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, 4(3), 49–59.  
<https://doi.org/10.1016/j.tine.2015.01.001>
- Haleva, L., Hershkovitz, A., & Tabach, M. (2021). Students’ activity in an online learning environment for mathematics: The role of thinking levels. *Journal of Educational Computing Research*, 59(4), 686–712. <https://doi.org/10.1177/0735633120972057>
- Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology*, 110(8), 1127–1148. <https://doi.org/10.1037/edu0000264>
- Hasty, L. M., Malanchini, M [Margherita], Shakeshaft, N., Schofield, K., Malanchini, M [Maddalena], & Wang, Z. (2021). When anxiety becomes my propeller: Mental toughness moderates the relation between academic anxiety and academic avoidance. *The British Journal of Educational Psychology*, 91(1), 368–390.  
<https://doi.org/10.1111/bjep.12366>
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie, Bd. 3: Psychologie des Unterrichts und der Schule* (pp. 71–176). Hogrefe.
- Higgins, K., Huscroft-D’Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A Meta-Analysis. *Journal of Educational Computing Research*, 57(2), 283–319.  
<https://doi.org/10.1177/0735633117748416>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.  
<https://doi.org/10.1016/j.compedu.2020.103897>
- Huang, Y.-M., Huang, S.-H., & Wu, T.-T. (2014). Embedding diagnostic mechanisms in a digital game for learning mathematics. *Educational Technology Research and Development*, 62(2), 187–207. <https://doi.org/10.1007/s11423-013-9315-4>

- Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1(2-3), 151–166. <https://doi.org/10.1007/s40692-014-0008-8>
- Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, 24, 190–197. <https://doi.org/10.1016/j.lindif.2012.12.014>
- Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of Computing in Higher Education*, 29(1), 114–132. <https://doi.org/10.1007/s12528-016-9126-9>
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education*, 104, 18–33. <https://doi.org/10.1016/j.compedu.2016.10.001>
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>
- Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift Für Pädagogische Psychologie*, 14(1), 26–37. <https://doi.org/10.1024//1010-0652.14.1.26>
- Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, 25, 120–148. <https://doi.org/10.1016/j.edurev.2018.10.001>
- LeFevre, J.-A., Kulak, A. G., & Heymans, S. L. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science*, 24(3), 276–289. <https://doi.org/10.1037/h0078742>
- Lukowski, S. L., DiTrapani, J. B., Jeon, M., Wang, Z., Schenker, V. J., Doran, M. M., Hart, S. A., Mazzocco, M. M. M., Willcutt, E. G., Thompson, L. A., & Petrill, S. A. (2019). Multidimensionality in the measurement of math-specific anxiety and its relationship with mathematical performance. *Learning and Individual Differences*, 70, 228–235. <https://doi.org/10.1016/j.lindif.2016.07.007>



- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- Nattland, A., & Kerres, M. (2009). Computerbasierte Methoden im Unterricht. In K.-H. Arnold, U. Sandfuchs, & J. Wiechmann (Eds.), *UTB Schulpädagogik, Pädagogik: Vol. 8423. Handbuch Unterricht* (2., aktualisierte Aufl., pp. 317–324). Klinkhardt.
- OECD. (2019). *PISA 2018 results: What students know and can do* (Vol. I). OECD Publishing.
- OECD. (2023). *PISA 2022 results: The state of learning and equity in education* (Vol. I). OECD Publishing.
- Ozel, S., Yetkiner, Z. E., & Capraro, R. M. (2008). Technology in K-12 mathematics classrooms. *School Science and Mathematics*, 108(2), 80–85. <https://doi.org/10.1111/J.1949-8594.2008.TB17807.X>
- Pekrun, R. (1988). Anxiety and motivation in achievement settings: Towards a systems-theoretical approach. *International Journal of Educational Research*, 12, 307–323. [https://doi.org/10.1016/0883-0355\(88\)90008-0](https://doi.org/10.1016/0883-0355(88)90008-0)
- Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7, 233285842098621. <https://doi.org/10.1177/2332858420986211>
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. <https://doi.org/10.1080/00461520.2015.1122533>
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52(3), 275–300. <https://doi.org/10.1080/15391523.2020.1719943>

- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322–338.  
<https://doi.org/10.1037/a0014996>
- Portes, A., & Rumbaut, R. G. (2001). *Legacies. The story of the immigrant second generation*. University of California Press.
- Quintero, M., Hasty, L., Li, T., Song, S., & Wang, Z. (2022). A multidimensional examination of math anxiety and engagement on math achievement. *The British Journal of Educational Psychology*, 92(3), 955–973.  
<https://doi.org/10.1111/bjep.12482>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, 14(2), 187–202. <https://doi.org/10.1080/15248372.2012.664593>
- Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, 53(3), 145–164. <https://doi.org/10.1080/00461520.2018.1447384>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12 classrooms. *Journal of Educational Computing Research*, 59(1), 119–153.  
<https://doi.org/10.1177/0735633120952063>
- Robert P. Tett, Daniel V. Simonet, Benjamin Walser, & Cameron Brown. (2013). Trait activation theory. In *Handbook of Personality at Work*. Routledge.  
<https://doi.org/10.4324/9780203526910.ch5>
- Rodrigo, M. M. T. (2011). Dynamics of student cognitive-affective transitions during a mathematics game. *Simulation & Gaming*, 42(1), 85–99.  
<https://doi.org/10.1177/1046878110361513>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. Advance online publication.  
<https://doi.org/10.25656/01:14686>
- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540–560. <https://doi.org/10.1111/j.1467-8535.2011.01197.x>

- Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, 37(1), 39–53. <https://doi.org/10.1007/s10212-021-00546-0>
- Spitzer, M. W. H., & Musslick, S. (2021). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic. *PloS One*, 16(8), e0255629. <https://doi.org/10.1371/journal.pone.0255629>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann Verlag GmbH. <https://doi.org/10.31244/9783830996064>
- Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument*. Universiteit van Amsterdam.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Tomaszewski, W., Xiang, N., & Western, M. (2020). Student engagement as a mediator of the effects of socio-economic status on academic performance among secondary school students in Australia. *British Educational Research Journal*, 46(3), 610–630. <https://doi.org/10.1002/berj.3599>
- Toppino, T. C., & Gerbier, E. (2014). About practice. In *Psychology of Learning and Motivation* (Vol. 60, pp. 113–189). Elsevier. <https://doi.org/10.1016/B978-0-12-800090-8.00004-4>
- Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology*, 99(2), 432–444. <https://doi.org/10.1037/0022-0663.99.2.432>
- Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent

- predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, 97(6), 1115–1128. <https://doi.org/10.1037/a0017048>
- Van de Werfhorst, H. G., Kessenich, E., & Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools. *Computers and Education Open*, 3, 100100. <https://doi.org/10.1016/j.caeo.2022.100100>
- Van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game. *Educational Studies in Mathematics*, 84(3), 281–307. <https://doi.org/10.1007/s10649-013-9483-5>
- Wong, Z. Y., & Liem, G. A. D. (2022). Student Engagement: Current State of the Construct, Conceptual Refinement, and Future Research Directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09628-3>