Pre-Registration of the Study

Motivation for Change! How Teacher Motivation Relates to the Success of Teacher Professional Development

Abstract

Teacher professional development (PD) is a central prerequisite of high-quality teaching. Evidence shows that PD objectives (e.g., promoting teachers' knowledge and skills, development of new ideas for teaching) are reflected in very different implementations (i.e., what components of a PD program and how PD programs are implemented in teaching practice). Currently, teachers' motivation to implement objectives of PD programs is considered as an important factor to explain differences in the success of implementation processes. The present study aims to examine how differences in teachers' motivation to implement PD objectives can be explained and how teachers' motivation relates to teaching practice and student learning. For this purpose, longitudinal data (three measurement points over five months) from an evaluation of a PD program on the use of technology in heterogeneous history classes will be used (N = 137 history teachers and their N = 1,670 students). By using factor analyses, we firstly aim to examine a three-factor structure (expectancy of success, perceived value, and cost) of teachers' motivation to implement objectives of PD programs as it was found by Osman and Warner (2020). Secondly, by using structural equation modelling, we explore to what extent different (a) PD-related and (b) teacher- and context-specific factors predict differences in teachers' motivation to implement objectives of PD programs. Lastly, we examine how the different factors of teachers' motivation (e.g., perceived value) to implement objectives of PD programs relate to perceived teaching quality and student-level outcomes such as learners' achievement or engagement.

Keywords: teacher professional development, motivation, PD implementation, longitudinal, student achievement

Introduction

Teacher professional development (PD) is a key aspect to prepare teachers for contemporary professional challenges (Darling-Hammond et al., 2009, 2017; Hammerness et al., 2005; Hendriks et al., 2010; OECD, 2019; Sprott, 2019) like technology enhanced teaching (e.g., Hillmayr et al., 2020) or adaptive teaching (e.g., Aas, 2020; Deunk et al., 2018; Forlin et al., 2008). The basic assumption is that teachers increase their knowledge and skills as well as change their attitudes and beliefs in high-quality PD and subsequently change their teaching practice, which leads to improved learning of their students (Darling-Hammond et al., 2017; Desimone, 2009). However, studies analyzed this hypothesized causal sequence of effects of PD activities revealed mixed findings. In general, literature reviews and meta-analyses showed that the effects of PD on teacher outcomes (e.g., self-efficacy, knowledge; Gesel et al., 2020) and student achievement (Basma & Savage, 2018; Blank & de las Alas, 2009) are positive (small to moderate), but that the effects vary considerably. Based on evidence, Hill (2009) argues that in many cases an adequate transfer of PD programs into classrooms lacks. Against this backdrop, findings from recent studies on PD indicate that teacher motivation is an important aspect to explain differences in changes of teaching practices after PD activities (Osman & Warner, 2020). That is, this motivation is discussed as a decisive link between PD activities and the actual implementation of PD programs (Osman & Warner, 2020). Currently, knowledge about what motivates teachers to implement objectives of PD programs and how these motivational states relate to teaching practices as well as student variables is scarce. This planned study examines how theoretically based personal (e.g., self-concept, perception of PD) and contextual factors (e.g., workload) relate to teachers' motivation to implement objectives of PD programs in their classroom. Furthermore, the association between teachers' motivation to implement objectives of PD programs and student-level variables (e.g., student achievement, engagement) will be analyzed. Doing this, longitudinal data (three measurement points over five months) of an evaluation of a PD program on the use of technology in heterogeneous history classes will be used (N = 137 history teachers and their N = 1,670 students).

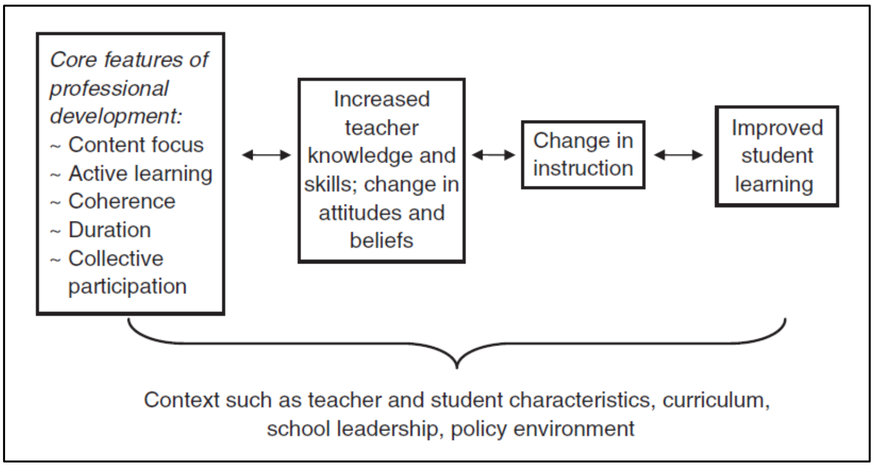
Theoretical Background

Significance and Effectiveness of Teacher Professional Development

Nations all around the world seek to improve their education systems (Hendriks et al., 2010; OECD, 2019). Teachers are crucial for a successful education system (Blömeke & Olsen, 2019; Desimone et al., 2006; Hattie, 2012; Seidel & Shavelson, 2007). Against this backdrop, teacher PD is important to prepare teachers for contemporary challenges (Darling-Hammond et al., 2009, 2017; Hammerness et al., 2005; Hendriks et al., 2010; OECD, 2019; Sprott, 2019). For instance, teachers need to be prepared for technology enhanced teaching in heterogeneous classrooms. The mechanisms of an effective PD are often conceptualized as follows: High-quality PD offers are used by teachers, which promotes changes in their instructional practices, mediated by enhanced professional knowledge of teachers (Figure 1; Clarke & Hollingsworth, 2002; Desimone, 2009). The underlying assumption is that PD participation improves teachers' professional knowledge and motivates them to translate new ideas into their classroom practice (Kennedy, 2016; K. S. Yoon et al., 2007). In turn, changed instructional practices lead to improved student learning outcomes.

Figure 1

Hypothesized Causal Sequence of Effects of Professional Development Activities



Note. Model adopted from Desimone (2009).

On average, literature reviews and meta-analyses show small to moderate positive effects of PD activities on teacher variables (e.g., knowledge: Gesel et al., 2020) and student achievement (Basma & Savage, 2018; Blank & de las Alas, 2009; Didion et al., 2020; Fletcher-Wood & Zuccollo, 2020; K. S. Yoon et al., 2007). However, mixed empirical findings often emerge for this idealized linear causal sequence of effects of PD activities which is well illustrated by the large variance in the effects of PD. For instance, a meta-analysis revealed that the effect size (ES) of PD on teacher knowledge varied from ES = -0.02 to ES = 2.28 and for teacher self-efficacy from ES = -0.08 to ES = 0.78 (Gesel et al., 2020). Similar findings exist for the effectiveness of PD on student achievement. For students' mathematics achievement PD had a positive effect on average but the effect sizes ranged from ES = -.19 to ES = 1.63 (Blank & de las Alas, 2009). For students' reading achievement, PD also had a positive effect on average (Basma & Savage, 2018; Didion et al., 2020). Again, the effect sizes ranged from ES = -.88 to ES = 2.25 (Basma & Savage, 2018) respectively from ES = -0.15 to ES = 0.94 (Didion et al., 2020). To explain such variances in effects, the associations of the mechanisms of effects shown in Figure 1 must be critically examined. One crucial aspect is whether teachers implement objectives of PD programs (e.g., change teaching practices; Sanchez et al., 2018). Hill (2009) argues—based on evidence—that in many cases an adequate transfer of PD programs into classrooms lacks. Against this backdrop, previous studies focused, for instance, on PD quality characteristics such as content focus (i.e., PD activities that focus on subject matter content), coherence (i.e., PD activities that show a fit between teacher learning and, for instance, teachers' beliefs or school policies), or collective participation (i.e., PD activities that allow interaction and discourse of teachers) as potential moderators (e.g., Hill, 2011). Indeed, it has been shown that only a fraction of business as usual PD (i.e., PD that have not been explicitly designed according to high-quality criteria) lead to changes in teaching practice (Darling-Hammond et al., 2017; Hill, 2009, 2011; Opfer & Pedder, 2011; Pedder et al., 2008). In contrast, PD that, for instance, facilitates collective participation or was coherent was found to be able to effect changes in teaching practices (Penuel et al., 2007; Watson & Manning, 2008). However, fitting to recent calls for more comprehensive approaches when analyzing the PD effectiveness (Liu & Phelps, 2020), in other studies further variables like the degree of differentiated PD (i.e., PD that target teachers' needs; Desimone & Garet, 2015; Desimone & Stuckey, 2014), individual teacher characteristics (e.g., treatment fidelity; Desimone & Stuckey, 2014), and contextual characteristics (e.g., school policies; Kennedy, 2016; Opfer & Pedder, 2011; Watson & Manning, 2008) are discussed as factors that explain whether and how teachers implement objectives of PD programs. Findings from recent studies on PD indicate that the previously neglected aspect of teacher motivation is important to explain differences in changes of teaching practices after PD activities (Osman & Warner, 2020). That is, this motivation is discussed as a decisive link between PD activities and the actual implementation of PD programs (Osman & Warner, 2020). However, evidence on this motivation and knowledge about predictors and associated variables of this motivation of teachers to implementation PD programs is scarce.

Teachers' Motivation to Implement Professional Development Programs

The implementation of PD programs in everyday professional practice depends, among other things, on teachers' motivation to do so (Andersson & Palm, 2018; Kennedy, 2016; Opfer & Pedder, 2011; Osman & Warner, 2020). That is, teachers' perception of PD is important for both the implementation of PD programs and student outcomes (Kragler et al., 2008; Rutherford et al., 2017). In general, following one of the most popular theories of motivation (i.e., expectation-value theory: Wigfield & Eccles, 2000), two main factors are considered to be crucial to explain different levels of motivation: (a) individual beliefs about one's competence and ability and the respective expectancy for success when doing a task and (b) the perceived value of a task in itself. The framework of the expectancy-value theory (Wigfield & Eccles, 2000) has mostly and extensively been applied to explain and find ways to foster children's and adolescents' motivation (e.g., Hulleman et al., 2016), but it has also been used in the context of teachers (Watt & Richardson, 2007, 2008, 2015). When aiming to explain mixed findings of the effectiveness of PD, one needs to consider that teachers perceive PD opportunities different (Avalos, 2011; Kennedy, 2016; Lipowsky & Rzejak, 2015; Opfer & Pedder, 2011). Individual perceptions refer to both of the aforementioned motivational factors: (a) teachers' individual competence beliefs and expectations for success, and (b) how much value teachers attribute to the respective tasks (Kennedy, 2016; Opfer & Pedder, 2011; Osman & Warner, 2020). Whereas there is still a debate whether perceived cost constitutes a subordinate (negative) part of value or a factor itself (Barron & Hulleman, 2015; Wigfield et al., 2017), recent extensions of the expectancy-value theory consider (c) perceived cost (e.g., educational, emotional, timely costs) as a third motivational factor (Flake et al., 2015; Jiang et al., 2018) that has also been found relevant in the context of teachers' motivation to implement objectives of PD programs (Osman & Warner, 2020; Rutherford et al., 2017).

Expectancy of Success With Implementation

The first motivational subdimension is closely related to concepts like self-efficacy, but task-specific (Eccles & Wigfield, 2002), and reflects beliefs about one's own competence and abilities and respective expectancies of success (i.e., "Can I do it?"; Wigfield & Eccles, 2000). Imagine a teacher who is still a novice when it comes to teaching with technology. If the teacher perceives a PD program about technology-enhanced teaching in a way that it seems manageable, they will be more likely to implement the newly learned techniques into their teaching. However, if the PD program leaves the teacher with the impression that they are still lacking the ability to successfully apply PD content, the newly learned techniques will less likely find its way into classrooms. Abrami et al. (2004) identified expectancy of success as the most important factor associated with implementation of an educational intervention. Similarly, Wozney et al. (2006) and Foley (2011) measured teachers' expectancy beliefs after PD participation and found a positive relation to implementation of the teaching strategies that were taught.

Perceived Implementation Value

The second motivational subdimension reflects the perceived importance of a task and the purpose and value of a respective action (i.e., "Why do it?"; Wigfield & Eccles, 2000). Teachers likely attribute different value to the same PD content and, accordingly, apply PD content differently in their own teaching (Kennedy, 2016; Kyndt et al., 2016; Opfer & Pedder, 2011; Osman & Warner, 2020; S. Y. Yoon et al., 2013). For instance, some teachers might consider a certain teaching technique that they learn about not as useful to support learning whereas others perceive it as a perfect fit for their own students. Commonly there are three different facets of perceived value pertaining to intrinsic value, attainment value and utility value (Eccles & Wigfield, 2020; Wigfield & Eccles, 2000). Intrinsic value refers to the interest in and enjoyment of doing a task. Attainment value reflects the importance of a task based on its alignment with individuals' values and identities (Wigfield & Eccles, 2000). PD programs that are perceived as in line with the own teaching philosophy (Donnell & Gettinger, 2015) and identity as an educator in general (Emo, 2015) are associated with higher rates of acceptance of educational innovations and respectively related to a positive attitude towards implementation. Perceived utility has been found to lead to higher levels of motivation to implement objectives of PD programs, for instance, when teachers have positive emotions about the PD content being able to improve teaching practices to help students learn (Gaines et al., 2019). With regards to PD about technology-enhanced teaching, higher levels of motivation to implement the PD content were found with regards to utility value-related affordances of teaching with technology such as gaining students' attention (i.e., making use of the technology's novelty effect) and collaborative opportunities (Kale & Akcaoglu, 2018).

Perceived Implementation Cost

Perceived cost reflects what is required, needs to be invested and/or given up doing something(i.e., related to effort, time and resources; Flake et al., 2015; Jiang et al., 2018). Osman & Warner (2020) showed that perceived cost is worthwhile to consider as an independent factor when it comes to teachers' motivation to implement objectives of PD programs. Whereas Wozney et al. (2006) found that cost did not influence implementation of ICT following a PD program as much as expectancy of success and perceived value, a qualitative study by Cameron et al. (2013) provides evidence that teachers always weigh educational as well as emotional costs against the benefits of implementing PD content. With regards to resources, support at school has been identified as a critical factor associated with difficulties to implement new practices in daily professional life (Opfer & Pedder, 2011).

Influencing Factors Explaining Differences in Motivation to Implement Objectives of PD Programs

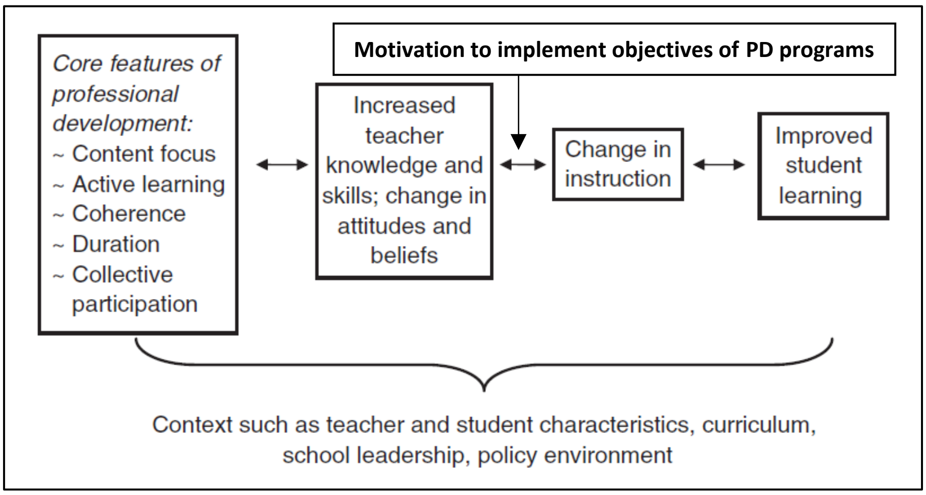
Eccles and Wigfield (2020) recently proposed a situated expectancy-value theory highlighting the situation-specific nature of how expectancies and values develop. This is in line with findings that emphasize situational influences on PD effectiveness (Kennedy, 2016; Kwakman, 2003). Taken together, there are several factors that potentially influence teachers' motivation to implement objectives of PD programs (i.e., specifically their expectancy of success, the perceived value as well as cost of applying the PD content to their own teaching). Based on the already existing findings (from qualitative as well as quantitative studies) about teachers' motivation for PD activities and subsequent implementation of the PD programs, potential influencing factors can be broadly grouped into (a) teacher-specific factors and (b) context-specific factors.

(a) Teacher-specific factors mostly refer to variables like teachers' pedagogical knowledge, teaching experience, and teachers' attitudes towards PD as well as towards changes in their instruction in general. The level of teaching experience is a decisive factor related to competence beliefs. For instance, (Abrami et al., 2004) found expectancies for success as the most important factor associated with implementation of an educational intervention – particularly for less experienced teachers. With regards to technology-enhanced teaching, teachers' motivational beliefs like perceived utility value of educational technology (e.g., Backfisch et al., 2020, 2021) or self-efficacy (e.g., Scherer et al., 2019) are important for effective technology integration in classrooms. Variables related to competence (belief) are particularly interesting in this context, as they are typically affected or even explicitly targeted by PD programs. Moreover, a study by Emo (2015) suggests that teachers' motivation for implementing PD programs is related to an innate wish for change to improve student learning. Thus, it can be assumed that the value of implementing changes in one's own instructional practices is intrinsic to some teachers, similar to traits like innovative behavior (Kleysen & Street, 2001) or teaching enthusiasm (Kunter et al., 2011; Lazarides et al., 2021) which have received attention particularly in the context of technology-enhanced teaching innovations. With regards to teachers' attitudes towards PD, Rzejak et al. (2014) found different motivational forces for PD participation (i.e., social interaction, external expectation, career orientation, self-development orientation) which are likely related to subsequent PD program implementation. That is, teachers' perception of PD seems to be important for both the effectiveness of a PD and the likelihood of implementing a PD program after participating a PD (Rutherford et al., 2017). A particularly important aspect in this context is the perceived quality of a PD program. If a PD program is perceived to be of high quality, then it can be assumed that PD programs are more likely to be implemented (i.e., leading to changes in teaching practice; Desimone, 2009).

Context factors mostly refer to the current teaching situation and school policies of participating teachers. The level of support from school administration and colleagues has been identified a key determinant of teachers' motivation in general (Pelletier et al., 2002). Moreover, overall workload is per definition directly related to perceived implementation cost (Eccles & Wigfield, 2020). Specifically for implementing technology-enhanced teaching practices, factors such as lacking resource availability are suggested to negatively impact the motivation to implement (Wozney et al., 2006).

Figure 2

Teachers' Motivation to Implement Objectives of PD Programs Located in the Model of Hypothesized Causal Sequence of Effects of Professional Development Activities



Note. Model adopted from Desimone (2009). The component of teachers' motivation to implement objectives of PD programs in their teaching practice was added.

Osman and Warner (2020) point out that future research should "consider whether variance in motivated behavior following PD is best explained by context- or person specific characteristics" (p. 10). Figure 2 illustrates where this motivation to implement objectives of PD programs becomes theoretically relevant.

Research Questions and Assumptions

The overall goal of this study is to better understand teachers' motivation to implement objectives of PD programs in their daily professional lives. A better understanding of motivational facets may yield powerful implications for educational stakeholders like PD providers, for instance how to establish conditions that enhance teachers' motivation to implement objectives of PD programs. In this study we will work on this overarching goal by exemplary using data of a PD program on preparing teachers for technology enhanced teaching in heterogeneous history classrooms. To enable teachers to implement technologies in their classrooms in a way that is effective for students' learning, the PD program aims to promote specific professional knowledge like technological-pedagogical knowledge (TPK; Harris et al., 2009; Koehler & Mishra, 2009; Lachner et al., 2019). TPK refers to teachers' cross-domain knowledge of how technologies can support students' learning during teaching (Harris et al., 2009; Lachner et al., 2019). Focusing on heterogeneous classes, teachers should know, for example, which digital tools to use and how to use them to make lessons as adaptive as possible. In this PD context, we are interested in teachers' motivation to implement the PD program. In particular, we will focus on the following three research questions:

1. How are the different motivational factors (expectancy of success, perceived attainment and utility value, and cost) reflected in teachers' motivation to implement objectives of PD programs? [*exploratory*]

   As Osman and Warner (2020) did not find the theoretically well-established four factor structure including the factor *perceived utility*, we will investigate if the three-factor structure of teachers' motivation to implement objectives of PD programs as it was found by Osman and Warner (2020) or the theoretically well-established four-factor structure including the factor *perceived utility* fits better to our data. As we adapted few item formulations (e.g., "What I have learned in this PD has a direct benefit for my work as a history teacher" instead of a very general "The things presented in this training are useful to me") we assume that a four-factor structure fits better to our data and outperforms a three-factor structure.

2. How are different teacher-specific and context-specific factors associated with teachers' motivation to implement objectives of PD programs (e.g., expectancy of success, perceived attainment and utility value, and cost)? [*exploratory*]

   We explore several factors based on and extending currently existing findings. We examine (a) PD-targeted variables (i.e., pedagogical content knowledge [PCK], technological-pedagogical knowledge [TPK], self-efficacy for teaching with technologies), as well as (b) teacher- and context-specific factors (i.e., innovative behavior, and general motivation for PD participation, pressure from colleagues/school management/curriculum, and workload).

3. How is teachers' motivation to implement objectives of PD programs associated with (a) students' achievement and motivation in the subject and (b) their perception of teachers' enthusiasm and teaching practices? [*explanatory*]

   In line with the commonly assumed framework of high-quality PDs leading to changes in teachers' knowledge and practices and respectively improved learning for students (Desimone, 2009), we expect a positive association of teachers' motivation to implement objectives of PD programs with student-level outcomes. We moreover explore in-depth relations with the distinct motivational factors.

Methodology

An ethics committee approved the study and the collection of the data and confirmed that the procedures were in line with ethical standards of research with human subjects (date of approval: 08/07/2019, file number: Drs.EK/2019/017-01).

Sample

The recruited sample consisted of history teachers from the German state of North Rhine-Westphalia (cohort 1) and Bavaria (cohort 2). In terms of recruitment, there were three simultaneous strategies: First, regional educational authorities of all governmental districts in the state were contacted and informed about the project. Upon agreement, all of them subsequently forwarded the information materials to all eligible school principals who were asked to encourage suitable teachers to sign-up. Furthermore, former participants of the PD program from a pilot run in the previous school year (Sept 2019 to June 2020) were used for snowball sampling to suggest signing-up to further interested teachers. Lastly, participants were recruited via social media (i.e., Twitter and Facebook) and calls for application on relevant websites. The incentive provided for teachers was the participation in a 5-month-long PD program for free (details on the PD program see below). All participants that signed up were admitted participating given that they fulfilled the inclusion criteria of the study (i.e., History teachers at secondary schools) and provided written informed consent. There were no further planned participant characteristics, however central background variables were assessed before the start of the PD program. The PD program aimed at history teachers at secondary schools (academic track and comprehensive schools), whereby teaching history was also possible without having studied the subject.

The second author only accessed the data once prior to pre-registration to gain insights into the nesting structure (necessary for the Monte-Carlo simulation studies conducted to estimate power [see appendix]). We have data from 137 teachers in our sample, of which the majority represented one school each, and 34 teachers (24.8 %) participated in the PD program with one to two colleagues from the same school. In addition to the teachers, their students of the classes[1] in which the participating teachers taught history during the time of the PD program, were invited to participate in the study. Students as well as their parents had to provide written informed consent to be eligible to participate in the study and fill out the questionnaires administered. Each of the teachers participated with 0 to 5 (M = 1.73) classes, yielding N = 173 classes in total. The number of participating students in each of the classes varied between 1 to 28 (M = 9.82; SD = 6.34), resulting in a total of N = 1,670 students. We will include all teachers in our study independently from their number of participating classes and number of students within a class.

Study Design

The study was a randomized controlled field trial with two cohorts and three measurement time points each. Participants of each cohort (cohort 1 starting 08/2020, cohort 2 starting 01/2021) were randomly assigned to either a treatment or waiting control group via a random number generation. Randomized allocation to one of the groups was done on a school-level to ensure participating teachers from the same school were in the same group, additionally considering balanced distribution among groups. The study design included repeated measures at three time points: The first measurement (TP1) happened before the start of the PD program, the second measurement (TP2)

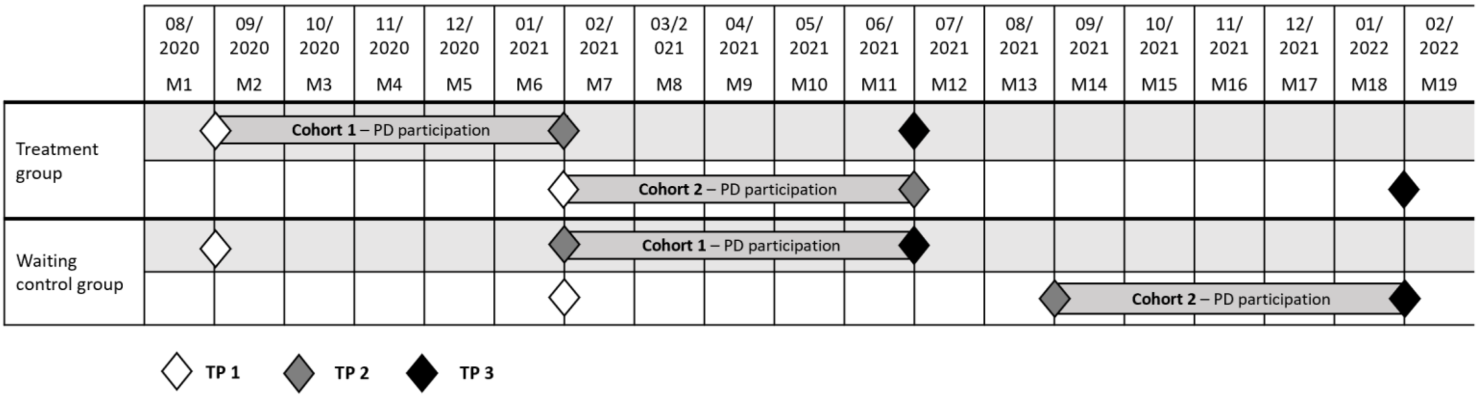
[1] Note. Whereas in U.S. middle schools or junior high schools classmates are usually different in every course/subject throughout the school day and year, in Germany a class of students take all the same courses throughout the school day and year together (Authors, 2022).

took place just after the end of the 5-months PD program in the treatment group, and the third measurement (TP3) took part another 5 months later after also the waiting control group had completed their PD program (see timeline in Figure 3). At all measurement time points both the participating teachers and their students were asked to fill out online questionnaires.

With regards to treatment fidelity, we kept track of participants' attendance rates during the PD program as it is well known that the effectiveness of a PD program depends significantly on teacher fidelity to interventions (Desimone & Stuckey, 2014).

Figure 3

Timeline



Note. M = Month; TP = Measurement time point.

Procedure and Materials

The PD program followed a blended learning concept and consisted of eight sessions that were held either in an in-person or a virtual format taking place every 2-4 weeks (depending on school holidays). The first and the fifth session were in-person meetings (1 day/8 hours each), however the latter of them had to be delivered online due to Covid-19 regulations at that time. Distributed evenly across the 5-months long PD program were six 90-120-minute synchronous interactive web conferences, focusing on a specific challenge of teaching (e.g., selection of materials, differentiation, or structure and goal direction; details see below). In the sense of a flipped classroom, participants had to do 30-45 minutes preparation for each meeting. Moreover, each session required 30-90 minutes individual practice in the following 2-3 weeks to apply the PD program contents (ideally for their own teaching). The individual practice could be done in teams if desired, but collaboration was not required. The products of the individual practice (i.e., teaching materials or preparations thereof) had to be uploaded to a feedback platform and were randomly assigned to two other participants who gave feedback via the platform. In addition, all materials were available on a digital learning platform. The digital learning platform also includes a forum for collaborative and communicative activities like discussions or questions.

The PD program was a joint project of three different disciplines (educational psychology, special education experts, history didactics including former teachers). Contents of the PD program subsequently were informed by different perspectives and each session integrated input from all disciplines on a specific focus topic. In most cases, the overall topic was defined by concrete subject-specific challenges during teaching contextualized in a general educational psychological way (e.g., structuring single lessons and lesson sequences, activating prior knowledge). Based on this, a specific pedagogical method was applied explicitly for teaching history in secondary schools and heterogeneous classrooms (e.g., Advance Organizers to support the activation and consideration of prior knowledge and interests). Moreover, the pedagogical psychological characteristics of the respective method were discussed to enable participants to understand underlying structures of effective teaching beyond applying single methods in their teaching. Every session also integrated the introduction and discussion on different digital tools that support teaching and learning regarding the overall session topic (i.e., tools for visualization and joint creation of Advance Organizers).

Measured Teacher Variables

Teachers' Motivation to Implement Objectives PD Programs

The central outcome teacher variable is teachers' motivation to implement objectives of PD programs. We used scales based on Osman and Warner (2020) to assess four motivational subscales (expectancy of success, intrinsic value, utility value, cost) with four items each (e.g., "I am positive that I will be able to apply what I learned in this PD in my history classes" for expectancy of success and "I have too much to invest to apply what I learned in this PD in my teaching practice" for cost). All items of the scale were measured with a 6-point Likert-type scale from 1 = *do not agree at all* to 6 = *fully agree*.

PD-Targeted Variables on the Teacher-Level

We assessed teachers' pedagogical content knowledge, technological-pedagogical knowledge as well as their self-efficacy for teaching with technology that are targeted by the PD program and potentially influence teachers' motivation to implement objectives of PD programs. We assessed these variables pre- and post PD (TP1 and TP2 respectively) to reflect respective changes during the PD program. We will use these variables as independent variables in our analyses.

Technological-Pedagogical Knowledge (TPK). We measured teachers' conceptual technological-pedagogical knowledge (TPK) with the test by Lachner et al. (2019). The conceptual knowledge test includes 8 multiple-choice tasks where teachers have 4 answer options each for which they have to decide whether they are right or wrong (e.g., "When learning with digital text and image elements ... (a) students learn better when stimulating additional information is integrated into the content presented. / (b) students learn better when the presented content is linked to each other. / (c) students learn better when they are presented with repetitive information. / (d) students learn better when related text and image content is presented one after the other.").

Pedagogical Content Knowledge (PCK). We assessed teachers' pedagogical content knowledge (history) with the test (GeDiKo-T) by Zabold et al. (in print). The test included 8 tasks targeting more procedural knowledge on the planning, implementation, and reflection of a history lesson that aims at the promotion of historical competencies in students (factor 1) and 3 tasks targeting more declarative knowledge of history didactics (i.e., on diagnostics of learning levels and subject-specific terminology; factor 2). All of the 11 tasks were presented in multiple-choice format consisting of a total of 120 subordinate questions (i.e., items) with either 2 or 3 or 4 answer options (depending on the question format) and one correct answer each (e.g., "Please assess the potential that thematic suggestions from a textbook have for references to the present such as those listed above from 'hardly any potential' to 'a lot of potential'" or "Establishing the learning state is a diagnostic measure that can be done before a teaching sequence on a new topic is conducted. What is true and what is not true about this form of diagnostic?"). The test performance regarding both factors of teachers' PCK is expressed as a weighted likelihood estimation (WLE; Warm, 1989).

Note. Currently, the factor structure of the test is being verified with a different data set than used in this study. We will include the test in our analyses according to the results of the factor structure check.

Self-Efficacy for Teaching With Technology. We used an adapted version of Ehmke et al.'s (2004) scale to assess teachers' self-efficacy for teaching with technology pre and post PD. The scale consisted of 9 items (e.g., "I'm too old to learn how to use technology“ [reverse scored] or „I can quickly learn new digital programs"), each scored on a 4-point Likert-type scale from 1 = *do not agree at all* to 4 = *fully agree*.

Additional Teacher and Context-Specific Factors on the Teacher-Level

Furthermore, we assessed teacher-specific and context-specific factors that potentially influence teachers' motivation to implement objectives of PD programs. We will use these variables as independent variables in our analyses.

Innovative Behavior. To account for individual differences with regards to overall openness to changes in own teaching practices, we assessed teachers' innovative behavior (based on Kleysen & Street, 2001) with 14 items (e.g., „How often do you experiment with new ideas and solutions?" or "How often do you search for ways to improve existing procedures or practices in your teaching?"), each scored on a 5-point Likert-type scale ranging from 1 = *never* to 5 = *always*.

Workload. We assessed teachers' overall workload with 5 items (e.g., "I feel overloaded in general" or "I perceive it as a problem to have to teach in so many classes") rated on a 5-point Likert-type scale from 1 = *do not agree at all* to 5 = *fully agree* (adpoted from Jaekel et al., 2020).

Pressure From Above. In order to account for support from teachers' surroundings (or lack thereof), we used the scale by Pelletier et al. (2002) to assess teachers' perceived pressure associated with (a) colleagues (e.g., "When you participate in PD, your colleagues are interested in what you have learned"), (b) school management (e.g., "The school management supports initiatives and further developments in your teaching"), and (c) the curriculum (e.g., "Due to curriculum restrictions, you have little freedom to develop your teaching"). The subscales consisted of five, four and four items respectively and were rated on a 7-point Likert-type scale from 1 = *do not agree at all* to 7 = *fully agree*.

General Motivation for PD. We assessed teachers' general motivation to participate in PD with the scale from Rzejak et al. (2014), covering four subdimensions (social interaction, external expectation, career orientation, self-development orientation) which included 4 items each (e.g., "In general, I take part in PD, because I am looking for collegial exchange" for social interaction or "In general, I take part in PD, because I am obliged to do so" for external expectation or "In general, I take part in PD, because I want to be up-to-date" for self-development orientation). The items were all measured with a 6-point Likert-type scale from 1 = *do not agree at all* to 6 = *fully agree*.

Perceived Quality of PD. We assessed teachers' perceived quality of PD with 4 items (e.g., "When I think back on the PD program, it has changed the way I think about my subject" or "When I think back on the PD program, the PD program was in line with the curricular requirements") rated on a 4-point Likert-type scale from 1 = *do not agree at all* to 4 = *fully agree* (items selected from Soine & Lumpe, 2014). The items are representative of the quality dimensions of high-quality PD as formulated by Desimone (Desimone, 2009; i.e., collaboration, content focus, coherence, collective participation). We will include the items individually in the analyses.

Covariates / Control Variables

We assessed teachers' teaching experience (continuous variable measured in years) and gender (dichotomous variable; 1 = *female*; 0 = *male*). We additionally assessed a number of background variables: We asked participants about the impact of Covid-19 on their teaching (3 items; "Did you teach history in distance learning during school closure?", "Have you taught history in split classes during the past school semester?" and "How high would you rate the limitations and rearrangements in your teaching over the past school semester due to Covid-19?") as well as about other roles at school and additionally taught subjects (next to history). Furthermore, we assessed teachers' treatment fidelity operationalized via their PD attendance rate.

Measured Student Variables

Students' Subject-Specific Engagement

We measured students' *cognitive engagement in history lessons* by adapting 5 items (i.e., subject mathematics was replaced by the subject history; e.g., "How often does it occur in history class that you check your schoolwork for mistakes?) from Jaekel et al. (2020) each rated on a 4-point Likert-type scale from 1 = *never* to 4 = *very often.*

Students' Subject-Specific Motivation

We assessed students' *motivation in history* pre and post PD with 12 items (based on expectancy-value theory; Gaspard et al., 2017) which were rated on a 4-point Likert-type scale from 1 = *do not agree at all* to 4 = *fully agree*. The items covered four motivational factors with 3 items each; intrinsic value (e.g., "I simply like history"), attainment value (e.g., "History is important to me personally"), utility value (e.g., "What we learn in history is useful for me"), and cost (e.g., "Dealing with history is exhausting me").

Students' Subject-Specific Achievement

We measured students' subject-specific achievement with a test of historical competencies (HiTCH) by Authors (2017). The test included 12 tasks (multiple-choice) consisting of a total of 84 subordinate questions (i.e., items) with either 2 or 3 or 4 answer options (depending on the question format) and one correct answer each. The tasks refer to different historical contexts whereby each of the task requires historical questioning skills, methodological knowledge, historical orientation competence, and historical content knowledge. All items required right or wrong decisions (e.g., "Researching the past is not that easy! Fill in the blank so that there is a correct statement about researching the past. You can choose: must not - can - must" or "Anne Frank. One element does not fit into the row. Choose the element that doesn't fit in the row and choose a reason that states how it is different from the other elements.") and assess an overall level of historical competence as one factor. The test performance is expressed as a WLE (Warm, 1989).

Students' Perceptions of Teaching Quality

To gain insights into teachers' teaching practices post PD, we asked students about *perceived instructional quality* in their history lessons. With respect to the specific PD focus on technology-enhanced teaching, we focused on teaching with technology use in history lessons. Quality of technology use was measured with 13 items scored on a 4-point Likert-type scale from 1 = *do not agree at all* to 4 = *fully agree*. The items reflected the level of classroom management, constructive support and cognitive activation (Praetorius et al., 2018; Seidel & Shavelson, 2007) and asked for instance "In lessons with technology, it's clear what you're allowed to do and what you're not" for classroom management or "Our teacher gives me additional support in lessons with technology if I need help" for constructive support.

Students' Perceptions of Teachers' Enthusiasm

Teachers' enthusiasm can be perceived well by students (i.e., easier than, for instance, teacher self-efficacy: Lazarides et al., 2021). We measured students' perceived teacher enthusiasm adapted from (i.e., subject mathematics was replaced by the subject history) a scale by Kunter et al. (2008) using 3 items (e.g., "Our history teacher is an enthusiastic teacher" or "Our history teacher is excited about history as a subject"), rated on a 4-point Likert-type scale from 1 = *do not agree at all* to 4 = *fully agree*.

Covariates / Control Variables

All student variables at TP1 function as covariates (i.e., potential confounding variables). In addition, we will use students' class level, gender, and their latest grades in German and history.

Statistical Analyses

[RQ1] To answer the first research question, we plan to conduct exploratory factor analyses (EFA), confirmatory factor analyses (CFA) and exploratory structural equation modeling (ESEM) to test the factor structure of teachers' motivation to implement objectives of PD programs. The steps are as follows:

Exploratory Factor Analyses (EFA)

In a first step, we determine the item with the highest loading for each factor of the four factors separately. In a second step, we then perform an EFA with all 16 items using a target rotation (i.e., items with the highest loadings in each case are uniquely assigned to the corresponding factor; Asparouhov & Muthén, 2009; Browne, 2001).

If the statistically fit of this model is acceptable (cut-offs see Schermelleh-Engel et al., 2003) for all fit indicators (RMSEA, CFI, SRMR), we will use this model. If the statistical fit of this model is not acceptable (cut-offs see Schermelleh-Engel et al., 2003) for at least one of the fit indicators (RMSEA, CFI, SRMR), we will first exclude the item that shows the highest cross-loading and second perform an EFA again with the 15 remaining items using a target rotation. For the renewed EFA with target rotation, the items with the highest loadings for each factor are not determined again. Instead, the originally identified target items per factor are retained. We will repeat this procedure as often as necessary until the model fit is acceptable. We call the resulting model M1. If the model fit will not become acceptable for all fit indicators (RMSEA, CFI, SRMR) by the described procedure, we will further explore the data to better understand the factor structure. For example, we will test for dimensionality using an EFA by comparing 3 versus 4 factors using all 16 items.

Confirmatory Factor Analyses (CFA)

In addition, we will conduct confirmatory factor analyses (CFA) using nine items to test the respective three-factor model as it was found by Osman and Warner(2020; M2). Furthermore, we will test for a four-factor model in our data (based on theoretical assumptions) using all 16 items (see original item set by Osman & Warner, 2020; M3). We will again evaluate the model fit (RMSEA, CFI, SRMR) following Schermelleh-Engel et al. (2003) and compare all three models (M1, M2, M3) by χ^2-difference tests and comparing model fit. In our analyses for research questions 2 and 3, we will use the model that was found to be the best in the comparison.

Exploratory Structural Equation Modeling (ESEM)

Finally, in case that all CFA models do not show an acceptable fit or as robustness checks for the model with the best fit, we will compute ESEM models (Marsh et al., 2009).

[RQ2] To answer the second research question, we will conduct multiple multivariate regression analyses (analysis of covariance [ANCOVA] approach) in a structural equation modelling (SEM) framework to analyze the influence of PD-targeted aspects (i.e., TPK, PCK, self-efficacy for teaching with technology) as well as additional teacher-specific and context-specific aspects (i.e., innovative behavior, workload, pressure from above, motivation for PD, and perceived PD quality) on different factors of teachers' motivation to implement objectives of PD programs. As we are interested in whether TPK, PCK, and self-efficacy at the end of the PD program—controlling for the respective pretest measures—influences teachers' motivation to implement objectives of PD programs, we will specify mediation models. Figure 4 shows the concept of these models. As potential confounding variables we will include teaching experience and gender. As robustness checks, we plan to include more covariates (e.g., treatment fidelity; see the section on covariates). The number of covariates included depends on the complexity of the models to be realized (i.e., number of parameters to be estimated), which can be estimated given the available sample size.

Given the expected sample size, we plan to reduce the complexity of the statistical models and therefore (a) first analyze the factors of the dependent variable (i.e., teachers' motivation to implement objectives of PD program) separately before analyzing them together and (b) apply a block wise approach, and (c) use scale scores of independent and dependent variables before applying latent variable modeling. That is:

(a) In a first step, we will analyze the factors of the dependent variable (i.e., teachers' motivation to implement objectives of PD programs) separately. In a second step, if the sample size permits (i.e., the models converge), we will analyze all factors of the dependent variable together in one model.
We will group the predictors and set up separate regression models for each group (M4, Figure 4a [block 1]: PD-targeted factors; M5, Figure 4b [block 2]: teacher- and context-specific factors). The following procedure is always chosen to run the models regarding research question 2: In a first step, as some independent variables on teacher level have subdimensions (e.g., *motivation for PD participation*), models are computed of both blocks that consider all subdimensions of an independent variable together. We consider all subdimensions of an independent variable together as we do not expect a risk of multicollinearity (i.e., high intercorrelations). In a second step, if the sample size permits (i.e., the models converge), all independent variables (including all subdimensions) of one block are considered together in one comprehensive model. To examine the relative importance of predictors from both blocks, we will then compute a regression model with predictors from both blocks. Doing this, we will select those predictors used in M4 and M5 on which at least one factor of teachers' motivation to implement objectives of PD programs regresses statistically significantly on. If multiple predictors show statistically significant regression coefficients and a reduction in the number of predictors is necessary due to sample size, we will select predictors based on the strength of their regression

weight. That means that we will first include the predictors that show the strongest regression weight overall within a block (i.e., for all factors of teachers' motivation to implement objectives of PD program). If predictor 1 shows, for instance, the strongest regression weight for factor 3 of teachers' motivation to implement objectives of PD programs and predictor 2 shows the strongest regression weight for factor 1, then we will keep both predictors. If it is not possible to keep both predictors, we will use the predictor with the (absolute) highest standardized regression weight across all factors of the dependent variable. Next, we will include other statistically significant predictors in the overall model if the model converges. To evaluate the explanatory power of the predictors for the variance of teachers' motivation to implement objectives of PD programs, we will use sample-size adjusted R^2_{adj}. To be able to compare the models based on R^2_{adj}, we use a saturated correlates approach (Graham, 2003) to keep the likelihood constant across the sequence of models (Hayes, 2021).

(b) All variables will first be modelled as manifest variables (e.g., scale scores, test scores). Only if the sample size allows it, the variables (besides control variables and test scores) are modeled as latent variables as well. When using latent variables—depending on the teacher-level sample size—we reserve the right to parcel items per construct (Little et al., 2013). We argue that a simplification of the measurement model utilizing scale scores or parceling is appropriate in our study as our research goal is to assess relations among constructs and not to assess the measurement properties (Rhemtulla, 2016).

Note 1: Using scale scores implies a higher power to detect structural misspecification that using latent modeling (Rhemtulla, 2016).

Note 2. As we will use different models (e.g., manifest, latent), we establish a decision rule by which we determine statistical significance. That is, if only manifest modeling of the constructs is possible within a model, the decisions on statistical significance tests will be based on the results when these scale scores are used. If latent modeling of the predictors is possible, the decisions on statistical significance tests will be based on the results using latent predictors. Latent modeling of the predictors would be of particularly advantage (thus will be preferred) when differences in the standardized regression coefficients should be interpreted as different reliabilities of the predictors are relevant in this case. If modeling of both predictors and outcomes is possible, the decisions on statistical significance tests will be based on the results using latent predictors and outcomes. Due to the high number of models computed (cumulative alpha risk), we will use the Benjamini-Hochberg (1995) adjustment for multiple testing to control the false discovery rate (i.e., the expected proportion of Type I errors).

[RQ3] To answer the third research question, we plan to use a SEM framework (specifically correlation models; M6) to investigate how the factors of teachers' motivation to implement objectives of PD programs are associated with student-level outcomes (i.e., engagement, motivation, achievement, perceived teaching quality, and teacher enthusiasm; see Figure 5).

We will include all dependent variables (student variables) at the same time in the models. The inclusion of several dependent variables at the same time make sense if values are missing on individual items. In this case, missing values are estimated via FIML. However, if missing values are systematically missing, e.g., because a participant did not participate at a measurement point or dropped out of the survey, the addition of further dependent variables does not have a major benefit regarding the estimation of missing values. Thus, if the models become too complex (i.e., convergence problems), we model separate models for the dependent variables (student level).

We will calculate intraclass correlation coefficients (ICCs) to estimate the impact of nesting of students (level 1) within classes (level 2) within teachers (level3). As even low intraclass correlations (e.g., .01) can lead to inappropriate standard error estimates and significance tests in regression analyses (Geiser, 2013), we will consider the nested data structure (i.e., students within different classes of different teachers) comprehensively. That is, as the multi-level structure is merely a nuisance factor but not the core of the research interest, we will account for clustering without modeling any cluster-specific random effects by using cluster-robust standard errors (i.e., students in teachers) to avoid assumptions required when modeling with random effects (McNeish et al., 2017). Furthermore, in the analyses the class level should be considered as an additional level (in addition to the student and teacher level). This is because assessments of students from the same classes, e.g., on teacher enthusiasm, might be more similar than assessments of students from another class, for example, because a teacher implements PD content in one class and not in another. Therefore, we will include all student assessments in the model each as two different manifest variables (scale scores) for students from class 1 or class 2, respectively, for each teacher. We assume that a teacher teaches either one or a maximum of two classes that participated in this study. This means one score is modelled for the first class and another score for the potentially second class. The respective scores are allowed to be correlated to capture teacher-level associations. The basic reasoning is that there are only clusters for teachers, but in some cases the class level occurs as an intermediate level in the models because the construct is measured by students from two classes. Classes are treated as interchangeable meaning that no systematic differences are expected between both classes. Therefore, all parameter estimates for each covariate assessed in two different classes are constrained to be equal (means, variances, regression coefficients). Further, covariances among these predictors are constrained to be equal across classes (i.e., the same intercorrelation pattern of the measures is assumed for both classes). Also, covariances of each predictor in one class with all predictors and the outcome in the respective other class are constrained to be equal across classes. Moreover, the residual variances of the student level outcome for each class and their covariances with teacher level measures are constrained to be equal across classes. In cases that only one class of a teacher participated in the study, the missingness of the respective student level measures for the second class are treated with full information maximum likelihood (FIML; see below for details). Based on a Monte-Carlo simulation study (see Appendix C) we assume that this is a plausible approach to adequately account for effects by class level in our statistical models. To ensure a straightforward graphical representation, we exemplify this modeling approach for the variable *students' subject-specific engagement* but we will apply the second-order modeling approach for all variables included (i.e., students' subject-specific motivation, students' subject-specific achievement, perceived teaching quality, perceived teacher enthusiasm). Some constructs on student level have subdimensions. For example, the variable *motivation domain-specific (history)* includes the four subdimensions *intrinsic value*, *attainment value*, *utility value*, and *cost*, each consisting of three items. Given the available sample sizes, the following procedure is always chosen to run the models regarding research question 3: 1. The models are computed separately for each subdimension of each variable. 2. Models are computed that consider all subdimensions of a variable together. 3. If the sample size permits (i.e., the models converge), all variables are considered together in one comprehensive model.

In all models for research questions 2 and 3, we will model teachers' motivation to implement objectives of PD programs using three or four factors, based on the findings of research question 1 (i.e., we will use the factor structure that emerges from the analyses regarding RQ1). In the paper by Ehmke et al., (2004) on construct *self-efficacy for teaching with technology*, it was shown that negatively formulated items did not have as high discriminatory power as positively formulated items. If confirmatory factor analyses reveal that the use of only positively formulated items provides a better model fit (measured by RMSEA, CFI, and SRMR; for cut-offs see Schermelleh-Engel et al., 2003), then we will assess the construct *self-efficacy for teaching with technology* with the positively formulated items.

As we will include TPK, PCK, and self-efficacy twice (measured at two measurement points TP1 and TP2) as independent variables in our SEM, we will test self-efficacy for measurement invariance following Geiser (2013). All hypotheses will be tested with two-tailed tests with a critical *p*-value and confidence intervals set at an alpha level of .05. Power analyses (more precisely Monte Carlo simulations) showed that sufficient power $1-\beta \geq .80$ can be expected in the analyses (see Appendix).

Missing Data

First, due to the longitudinal design of the study, we expect a drop out of participants across the different measurement points. In this case, we will comply with the following rule for missing value treatments: Do participants who show missing values because they dropped out of the study in the longitudinal study (drop-out group) differ from participants who did not drop out (non-drop-out group) regarding the dependent, independent and control variables (e.g., tested via t-tests)?

1) Yes, then we first include participants from both the drop-out and the non-drop-out group in the analyses and include these variables as control variables in the model for which differences between the drop-out group and non-drop-out group are evident (in addition to the variables to be included in the models based on theoretical considerations). Current research shows that approaches in which missing values are handled with FIML provide unbiased estimates if all variables associated with missingness are included in the estimation (Graham, 2009, 2012; Schafer & Graham, 2002). Therefore, we will treat missing values with FIML estimation. We will use FIML for both independent and dependent variables. Furthermore, we will include auxiliary variables (e.g., age) in the FIML estimation using the command *auxiliary (m)* as implemented in Mplus (saturated correlates models: Graham, 2003) unless these variables are already included in the model as control variables. However, as the inclusion of auxiliary variables involves estimating relationships between them and all indicators (manifest variables) in the models, auxiliary variables can only be added if sample sizes permit. Therefore, the addition of auxiliary variables will be more feasible at the student level than at the teacher level. If sample sizes are too small, we will refrain from adding auxiliary variables. Second, as a robustness check, we will conduct the same statistical models by applying a listwise (i.e., case) deletion approach and include only those participants from the non-drop-out group in the analyses.

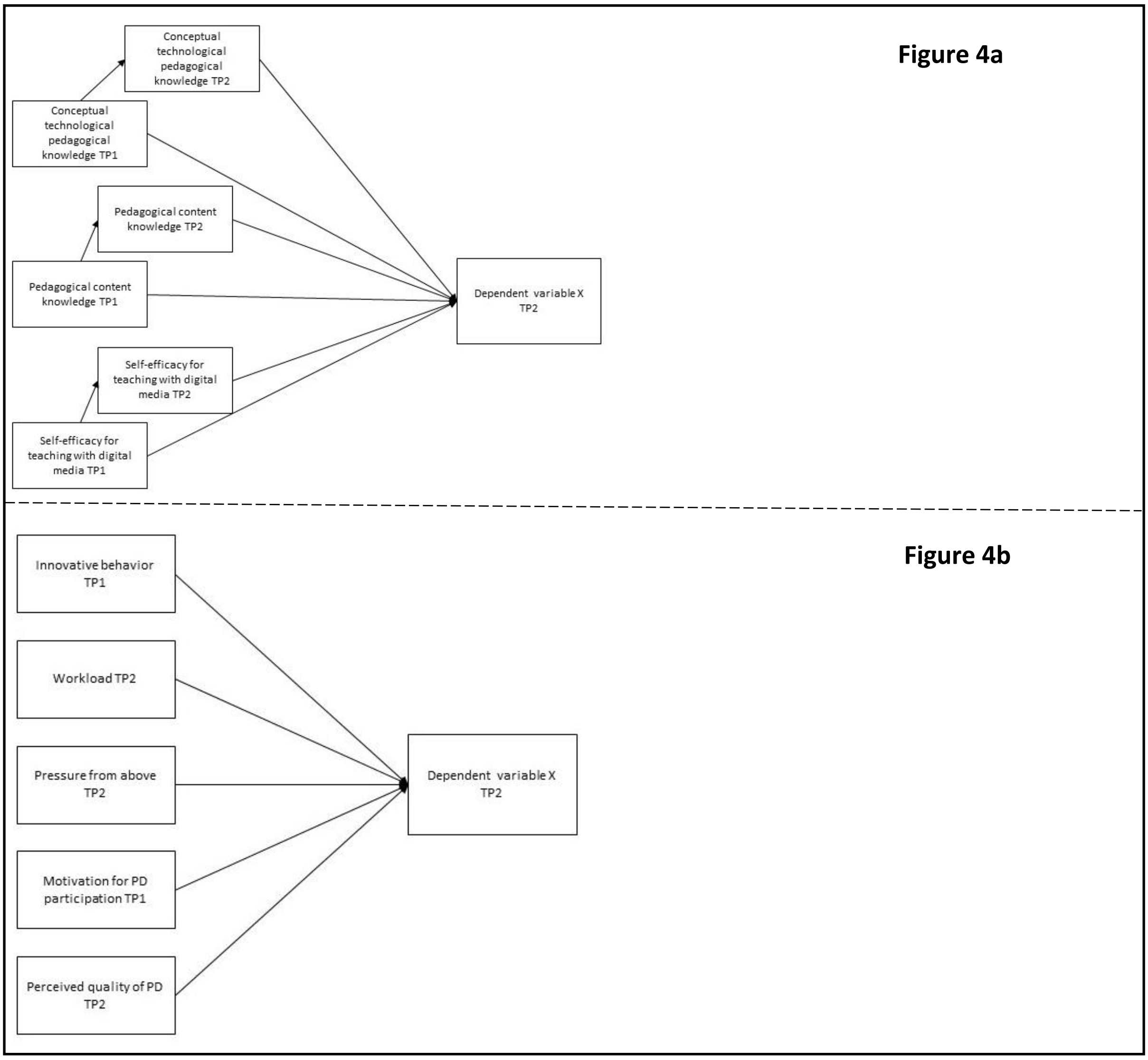
2) No, then we will perform the same analyses as described in 1) without additional auxiliary variables.

Second, we assume that individual values will be missing. However, as the authors who will carry out the analyses currently have no access to the data, we do not know how many values are missing in total for the variables considered in this study. If scattered values of dependent and independent measures are missing, then we will again use FIML to treat missing values for both independent and dependent variables. We will exclude participants (listwise/case deletion) who have missing values on all dependent and independent measures.

To sum up, we assume a smaller sample size than teachers available only due to differences between the drop-out and the non-drop-out group regarding the dependent and independent variables or due to cases who show missing values on all dependent and independent measures.

Figure 4

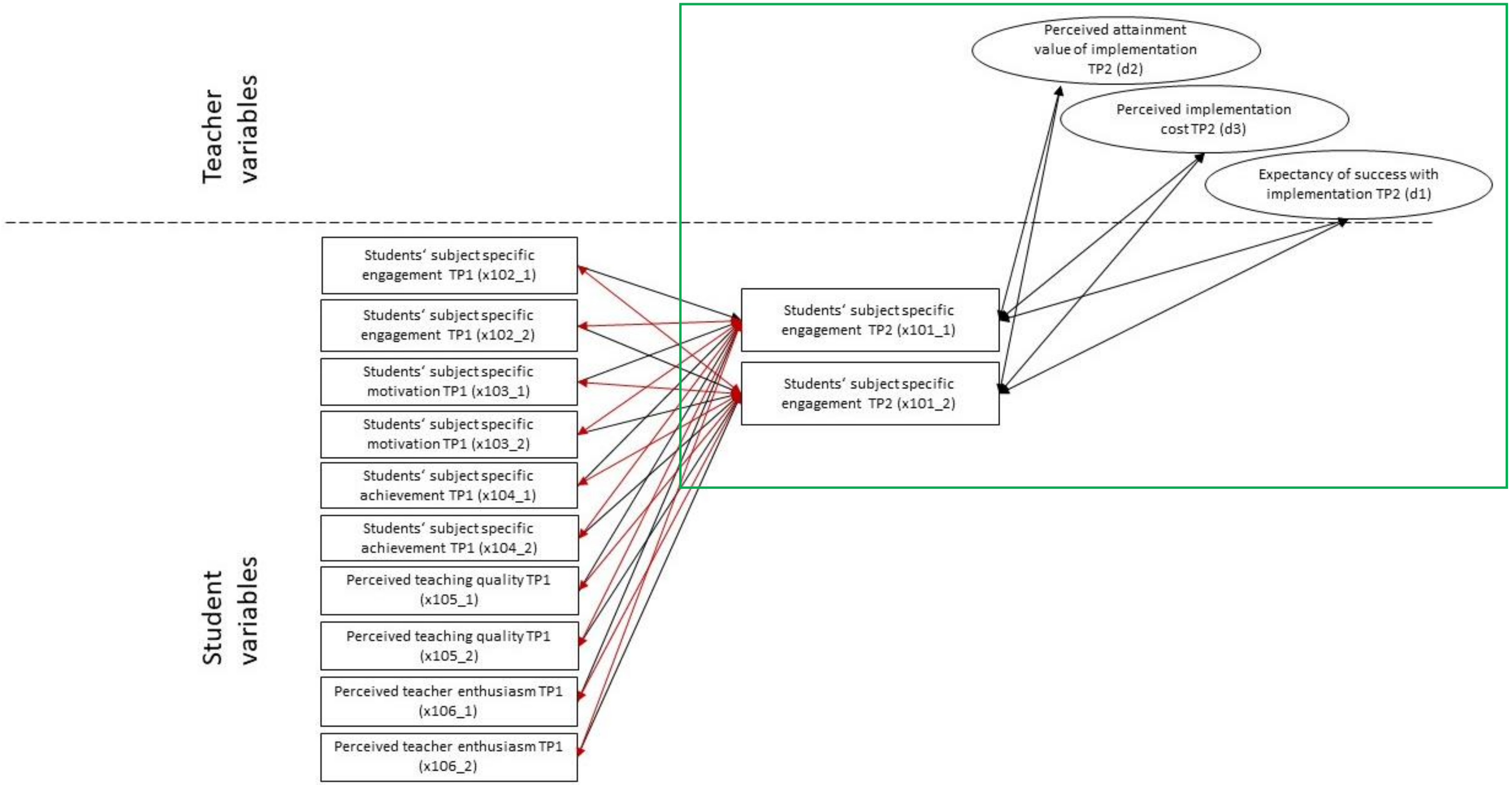
Conceptual Illustration of Structural Equation Models M4 & M5 Used to Answer Research Question 2



Note. Figures 4a shows the first block of independent variables (PD-targeted factors). Figure 4b shows the second block of independent variables (teacher- and context-specific factors). *X* represents one of the four factors of the main dependent variable motivation to implement PD (e.g., expectancy of success with implementation). All variables in the figure are regressed on covariates (e.g., teaching experiences, gender; not illustrated in the figures). Details of the modeling (e.g., correlations of predictors) are not shown in the figures.

Figure 5

Conceptual Illustration of Structural Equation Models M6 Used to Answer Research Question 3



Note. x10i_j, i ∈ {1, ..., 6}, j ∈ {1, 2} refer to the labels used in the simulation study (Appendix C). As an example, student* subject specific engagement was used as a dependent variable on student level. However, this model will also be applied to the other student-level variables. The models are planned as single level models with robust standard errors (Mplus command *type = complex*). The green border indicates research question 3. The student variables at TP1 function as covariates. In addition, all variables in the figure are regressed on further covariates (on student level: e.g., grade, gender; on teacher level: e.g., teaching experiences, gender; not illustrated in the figure). Details of the modeling (e.g., correlations of predictors) are not shown in the figures. Red arrows between student variables represent covariances, black arrows represent regressions.

Teacher Variables				
Construct / Variable	**Function**	**Items (Number / Sample)**	**Timepoint**	**Scale**
Motivation to implement PD (Osman & Warner, 2020)	DV (RQ2) IV (RQ3)	16 items: 4 subscales (expectancy, intrinsic value, utility value, cost) with 4 items each; e.g., *“I am positive that I will be able to apply what I learned in this PD in my history classes”* for expectancy and *“I have too much to invest to apply what I learned in this PD in my teaching practice”* for cost	Posttest (TP2)	continuous on 6-point Likert-type scale (1 = do not agree at all, 6 = fully agree)
Technological-pedagogical knowledge conceptual (TPK; Lachner et al., 2019)	IV	8 tasks (multiple-choice) with 4 answer options each of which one to three are correct; e.g., *“When learning with digital text and image elements ... (a) students learn better when stimulating additional information is integrated into the content presented. / (b) students learn better when the presented content is linked to each other. / (c) students learn better when they are presented with repetitive information. / (d) students learn better when related text and image content is presented one after the other.”*	Pretest (TP1) and Posttest (TP2)	Item level: categorical scoring (0 = wrong, 1 = right) Test score: continuous sum score based on right/wrong coding
Pedagogical content knowledge history (GeDiKo-T; Zabold, in print)	IV	11 tasks (multiple-choice; 2 factors: 1 = reflection of given lesson planning and action; 2 = history didactic categorization competence as a reference point for lesson planning, action and reflection) consisting of a total of 120 subordinate questions (i.e., items) with either 2 or 3 or 4 answer options (depending on the question format) and one correct answer each. *Note*. Currently, the factor structure of the test is being verified with a different data set than used in this study. We will include the test in our analyses according to the results of the factor structure check.	Pretest (TP1) and Posttest (TP2)	Item level: categorical scoring (0 = wrong, 1 = right) Test score: two continuous WLE scores for each of the two factors based on right/wrong coding
Self-efficacy for teaching with technology (adapted from Ehmke et al., 2004)	IV	9 items; e.g., *“I'm too old to learn how to use technology“* (reverse scored) or *„I can quickly learn new digital programs”*	Pretest (TP1) and Posttest (TP2)	continuous on 4-point Likert-type scale (1 = do not agree at all, 4 = fully agree)
Innovative behavior (adapted from	IV	14 items, e.g., *„How often do you experiment with new ideas and solutions?”* or *“How ofter do you search for ways to improve existing*	Pretest (TP1)	continuous on 5-point Likert-type scale

Kleysen & Street, 2001).		*procedures or practices in your teaching?"*		(1 = never, 5 = always)
Workload	IV	5 items (e.g., *"I feel overloaded overall"* or *"I consider it a problem to have to teach in so many classes")*	Posttest (TP2)	continuous on 5-point Likert-type scale (1 = do not agree at all, 5 = fully agree)
Pressure from above, i.e., from (1) colleagues, (2) school administration, (3) curriculum (Pelletier et al., 2002)	IV	13 items: 3 subscales with 5 and 4 and 4 items; subscale 1 e.g., *"When you participate in PD, your colleagues are interested in what you have learned."*, subscale 2 e.g., *"The school management supports initiatives and further developments in your teaching."*, subscale 3 e.g., *"Due to curriculum restrictions, you have little freedom to develop your teaching."*	Posttest (TP2)	continuous on 7-point Likert-type scale (1 = do not agree at all, 7 = fully agree)
Motivation for PD participation (adopted from Rzejak et al., 2014)	IV	16 items: 4 subscales (social interaction, external expectation, career orientation, self-development orientation) with 4 items each; e.g., subscale 1*"In general, I take part in PD, because I am looking for collegial exchange"* or subscale 2 *"In general, I take part in PD, because I am obliged to do so"* or subscale 4 *"In general, I take part in PD, because I want to be up-to-date"*	Pretest (TP1)	continuous on 6-point Likert-type scale (1 = do not agree at all, 6 = fully agree)
Perceived quality of PD (items selected from Soine & Lumpe, 2014)	IV	4 Items (e.g., "When I think back on the PD program, it has changed the way I think about my subject" or "When I think back on the PD program, the PD program was in line with the curricular requirements").	Posttest (TP2)	continuous on 4-point Likert-type scale (1 = do not agree at all, 4 = fully agree)
Gender	CV	1 item	time invariant	categorical (1 = female, 2 = male, 3 = other)
Years of teaching experience	CV	1 open item	time invariant	continuous
Additional subjects (next to history)[a]	CV	13 multi-choice items + 1 open item	time invariant	categorical
Weekly teaching load[a]	CV	1 open item	time invariant	continuous
Other role(s) at the school[a]	CV	1 item + open response	time invariant	categorical (1 = yes, 2 = no)
Attendance rate [Treatment fidelity]	CV	logged by researchers	time invariant	continuous (1 = 2 hours of treatment to 12 = 24 hours of treatment)

Note. DV = Dependent Variable, IV = Independent Variable, CV = Covariate / Control Variable
[a] This variable is included only if the model complexity to be realized (i.e., number of parameters to be estimated) can be estimated with the available sample size.

Student Variables				
Construct / Variable	**Function**	**Items (Number / Sample)**	**Timepoint**	**Scale**
Students' subject specific engagement - Cognitive engagement in history lessons (adapted from Jaekel et al., 2020)	CV (TP1) DV (TP2)	5 items, e.g., *"How often does it occur in history class that you check your schoolwork for mistakes?"*	Pretest (TP1) and Posttest (TP2)	continuous on 4-point Likert-type scale (1 = never, 4 = very often)
Students' subject specific motivation - Motivation in history (adapted from Gaspard et al., 2017)	CV (TP1) DV (TP2)	12 items: based on expectancy-value theory covering 4 subdimensions with 3 items each; intrinsic value (e.g., *"I simply like history"* or attainment value *(e.g., "History is important to me personally"),* utility value *(e.g., "What we learn in history is useful for me"),* and cost *(e.g., "Dealing with history is exhausting me")*	Pretest (TP1) and Posttest (TP2)	continuous on 4-point Likert-type scale (1 = do not agree at all, 4 = fully agree)
Students' subject specific achievement - historical competencies (HiTCH, Authors, 2017)	CV (TP1) DV (TP2)	12 tasks (multiple-choice; 1 factor = historical competencies) consisting of a total of 84 subordinate questions (i.e., items) with either 2 or 3 or 4 answer options (depending on the question format) and one correct answer each; e.g., *"Researching the past is not that easy! Fill in the blank so that there is a correct statement about researching the past. You can choose: must not - can - must"* or *"Anne Frank. One element does not fit into the row. Choose the element that doesn't fit in the row and choose a reason that states how it is different from the other elements."*	Pretest (TP1) and Posttest (TP2)	Item level: categorical scoring (0 = wrong, 1 = right) Test score: Continuous, WLE score based on right/wrong coding
Perceived teaching quality - Quality of technology use (adapted from Jaekel et al., 2020)	DV (TP2)	13 items (reflecting 3 teaching quality subdimensions); e.g., *"In lessons with technology, it's clear what you're allowed to do and what you're not"* for classroom management or	Posttest (TP2)	continuous on 4-point Likert-type scale (1 = do not agree at all, 4 = fully agree)

		"Our teacher gives me additional support in lessons with technology if I need help" for constructive support or "*In lessons with technology, we often apply what we have learned to new things*" for cognitive activation.		
Perceived teacher enthusiasm (adapted from Kunter et al., 2008)	DV (TP2)	3 items, e.g., *"Our history teacher is an enthusiastic teacher"* or *"Our history teacher is excited about history as a subject"*	Posttest (TP2)	continuous on 4-point Likert-type scale (1 = do not agree at all, 4 = fully agree)
Gender	CV	1 item	time invariant	categorical (0 = male, 1 = female)
Latest grades in German and history	CV	2 items with one for each subject	Pretest (TP1) and Posttest (TP2)	categorical (1 = very good to 6 = insufficient)
Class level	CV	1 item	Pretest (TP1)	categorical (1 = 5th grade, 2 = 6th grade, 3 = 7th grade, 4 = 8th grade, 5 = 9th grade, 6 = 10th grade, 7 = 11th grade, 8 = 12th grade)
Note. DV = Dependent Variable, IV = Independent Variable, CV = Covariate / Control Variable				

References

Aas, H. K. (2020). Professional development for inclusive and adaptive education: Lesson Study in a Norwegian context. *Professional Development in Education*, 1–15. https://doi.org/10.1080/19415257.2020.1850509

Abrami, P. C., Poulsen, C., & Chambers, B. (2004). Teacher motivation to implement an educational innovation: Factors differentiating users and non-users of cooperative learning. *Educational Psychology*, *24*(2), 201–216. https://doi.org/10.1080/0144341032000160146

Andersson, C., & Palm, T. (2018). Reasons for teachers' successful development of a formative assessment practice through professional development – a motivation perspective. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 576–597. https://doi.org/10.1080/0969594X.2018.1430685

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, *27*(1), 10–20. https://doi.org/10.1016/j.tate.2010.08.007

Backfisch, I., Lachner, A., Hische, C., Loose, F., & Scheiter, K. (2020). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and Instruction*, *66*, 101300. https://doi.org/10.1016/j.learninstruc.2019.101300

Backfisch, I., Lachner, A., Stürmer, K., & Scheiter, K. (2021). Variability of teachers' technology integration in the classroom: A matter of utility! *Computers & Education*, *166*, 104159. https://doi.org/10.1016/j.compedu.2021.104159

Basma, B., & Savage, R. (2018). Teacher professional development and student literacy growth: A systematic review and meta-analysis. *Educational Psychology Review*, *30*(2), 457–481. https://doi.org/10.1007/s10648-017-9416-4

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Blank, R. K., & de las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement. How meta analysis provides scientific evidence useful to education leaders* (Council of Chief State School Officers, Ed.). Council of Chief State School Officers.

Blömeke, S., & Olsen, R. V. (2019). Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries. *Teaching and Teacher Education*, *77*, 170–182. https://doi.org/10.1016/j.tate.2018.09.018

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Cameron, S., Mulholland, J., & Branson, C. (2013). Professional learning in the lives of teachers: Towards a new framework for conceptualising teacher learning. *Asia-Pacific Journal of Teacher Education*, *41*(4), 377–397. https://doi.org/10.1080/1359866X.2013.838620

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession. A status report on teacher development in the United States and abroad* (National Staff Development Council and The School Redesign Network at Stanford University, Ed.).

https://static1.squarespace.com/static/56b90cb101dbae64ff707585/t/583c7fe720099e25d0b1bd24/1480359912004/nsdcstudy2009.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, & Education*, *7*(3), 252–263.

Desimone, L. M., Smith, T. M., & Ueno, K. (2006). Are teachers who need sustained, content-focused professional development getting it? An administrator's dilemma. *Educational Administration Quarterly*, *42*(2), 179–215. https://doi.org/10.1177/0013161X04273848

Desimone, L. M., & Stuckey, D. (2014). Sustaining professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, PreK-12* (pp. 467–482). Guilford Publications.

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54. https://doi.org/10.1016/j.edurev.2018.02.002

Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness*, *13*(1), 29–66. https://doi.org/10.1080/19345747.2019.1670884

Donnell, L. A., & Gettinger, M. (2015). Elementary school teachers' acceptability of school reform: Contribution of belief congruence, self-efficacy, and professional development. *Teaching and Teacher Education*, *51*, 47–57. https://doi.org/10.1016/j.tate.2015.06.003

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://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

Ehmke, T., Senkbeil, M., & Bleschke, M. (2004). Typen von Lehrkräften beim schulischen Einsatz von Neuen Medien. In F. Schumacher (Ed.), *Innovativer Unterricht mit neuen Medien: Ergebnisse wissenschaftlicher Begleitung von Semik-Einzelprojekten* (pp. 35–66). FWU Institut für Film und Bild in Wissenschaft und Untericht.

Emo, W. (2015). Teachers' motivations for initiating innovations. *Journal of Educational Change*, *16*(2), 171–195. https://doi.org/10.1007/s10833-015-9243-7

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Fletcher-Wood, H., & Zuccollo, J. (2020). *The effects of high-quality professional development on teachers and students. A rapid review and meta-analysis*. Education Policy Institute.

Foley, L. S. (2011). Exploring K–3 teachers' implementation of comprehension strategy instruction (CSI) using expectancy-value theory. *Literacy Research and Instruction*, *50*(3), 195–215. https://doi.org/10.1080/19388071.2010.505680

Forlin, C., Keen, M., & Barrett, E. (2008). The concerns of mainstream teachers: Coping with inclusivity in an Australian context. *International Journal of Disability, Development and Education*, *55*(3), 251–264. https://doi.org/10.1080/10349120802268396

Gaines, R. E., Osman, D. J., Maddocks, D. L. S., Warner, J. R., Freeman, J. L., & Schallert, D. L. (2019). Teachers' emotional experiences in professional development: Where they come from and what they can mean. *Teaching and Teacher Education*, *77*, 53–65. https://doi.org/10.1016/j.tate.2018.09.008

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

Geiser, C. (2013). *Data analysis with Mplus* (1st ed.). The Guilford Press.

Gesel, S. A., LeJeune, L. M., Chow, J. C., Sinclair, A. C., & Lemons, C. J. (2020). A meta-analysis of the impact of professional development on teachers' knowledge, skill, and self-efficacy in data-based decision-making. *Journal of Learning Disabilities*, 002221942097019. https://doi.org/10.1177/0022219420970196

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W. (2012). *Missing data. Analysis and design*. Springer. https://doi.org/10.1007/978-1-4614-4018-5

Hammerness, K., Darling-Hammond, L., Berliner, D. C., Cochran-Smith, M., McDonald, M., & Zeichner, K. (2005). How teachers learn and develop," in teachers for a changing world: What teachers should learn and be able to do. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing Teachers for a Changing World. What Teachers Should Learn and Be Able to Do* (pp. 358–389). Jossey-Bass.

Harris, J., Mishra, P., & Koehler, M. J. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education*, *41*(4), 393–416. https://doi.org/10.1080/15391523.2009.10782536

Hattie, J. (2012). *Visible learning for teachers. Maximizing impact on learning*. Routledge.

Hayes, T. (2021). R-squared change in structural equation models with latent variables and missing data. *Behavior Research Methods*, *53*(5), 2127–2157. https://doi.org/10.3758/s13428-020-01532-y

Hendriks, M. A., Luyten, H., Scheerens, J., Sleegers, P., & Steen, R. (2010). *Teachers' professional development. Europe in international comparison* (J. Scheerens, Ed.). Office for Official Publications of the European Union. https://doi.org/10.2766/63494

Hill, H. C. (2009). Fixing teacher professional development. *Phi Delta Kappan*, *90*(7), 470–476. https://doi.org/10.1177/003172170909000705

Hill, H. C. (2011). *The nature and effects of middle school mathematics teacher learning experiences*. *113*(1), 205–234. https://doi.org/10.1177/016146811111300106

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

Jaekel, A.-K., Göllner, R., & Trautwein, U. (2020). How students' perceptions of teaching quality in one subject are impacted by the grades they receive in another subject: Dimensional comparisons in student evaluations of teaching quality. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000488

Jiang, Y., Rosenzweig, E. Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemporary Educational Psychology*, *54*, 139–152. https://doi.org/10.1016/j.cedpsych.2018.06.005

Kale, U., & Akcaoglu, M. (2018). The role of relevance in future teachers' utility value and interest toward technology. *Educational Technology Research and Development*, *66*(2), 283–311. https://doi.org/10.1007/s11423-017-9547-9

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Kragler, S., Martin, L., & Kroeger, D. C. (2008). Money down the drain: Mandated professional development. *Journal of School Leadership*, *18*(5), 528–550. https://doi.org/10.1177/105268460801800503

Kunter, M., Frenzel, A., Nagy, G., Baumert, J., & Pekrun, R. (2011). Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology*, *36*(4), 289–301. https://doi.org/10.1016/j.cedpsych.2011.07.001

Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, *18*(5), 468–482. https://doi.org/10.1016/j.learninstruc.2008.06.008

Kwakman, K. (2003). Factors affecting teachers' participation in professional learning activities. *Teaching and Teacher Education*, *19*(2), 149–170. https://doi.org/10.1016/S0742-051X(02)00101-4

Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). Teachers' everyday professional development: Mapping informal learning activities, antecedents, and learning outcomes. *Review of Educational Research*, *86*(4), 1111–1150. https://doi.org/10.3102/0034654315627864

Lachner, A., Backfisch, I., & Stürmer, K. (2019). A test-based approach of modeling and measuring technological pedagogical knowledge. *Computers & Education*, *142*, 103645. https://doi.org/10.1016/j.compedu.2019.103645

Lazarides, R., Fauth, B., Gaspard, H., & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations to changes in student-perceived teaching quality at the beginning of secondary education. *Learning and Instruction*, *73*, 101435. https://doi.org/10.1016/j.learninstruc.2020.101435

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *RICERCAZIONE*, *7*(2), 27–51.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300. https://doi.org/10.1037/a0033266

Liu, S., & Phelps, G. (2020). Does teacher learning last? Understanding how much teachers retain their knowledge after professional development. *Journal of Teacher Education*, *71*(5), 537–550. https://doi.org/10.1177/0022487119886290

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

OECD. (2019). *TALIS 2018 results. Teachers and school leaders as lifelong learners (volume I)*. OECD Publishing. https://doi.org/10.1787/1d0bc92a-en

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*(3), 376–407. https://doi.org/10.3102/0034654311413609

Osman, D. J., & Warner, J. R. (2020). Measuring teacher motivation: The missing link between professional development and practice. *Teaching and Teacher Education*, *92*, 103064. https://doi.org/10.1016/j.tate.2020.103064

Pedder, D., Storey, A., & Opfer, V. D. (2008). *Schools and continuing professional development (CPD) in England—State of the Nation research project*. Cambridge University and The Open University.

Pelletier, L. G., Séguin-Lévesque, C., & Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *Journal of Educational Psychology*, *94*(1), 186–196. https://doi.org/10.1037/0022-0663.94.1.186

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221

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

Rhemtulla, M. (2016). Population performance of SEM parceling strategies under measurement and structural model misspecification. *Psychological Methods*, *21*(3), 348–368. https://doi.org/10.1037/met0000072

Rutherford, T., Long, J. J., & Farkas, G. (2017). Teacher value for professional development, self-efficacy, and student outcomes within a digital mathematics intervention. *Contemporary Educational Psychology*, *51*, 22–36. https://doi.org/10.1016/j.cedpsych.2017.05.005

Sanchez, J. E., Williams, S. L., & Ferrara, M. M. (2018). Teachers´ perceptions of professional development experiences: An ongoing concern. *Journal for Educators, Teachers and Trainers*, *9*(1), 102–115.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037//1082-989X.7.2.147

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74. https://doi.org/10.1.1.509.4258

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Soine, K. M., & Lumpe, A. (2014). Measuring characteristics of teacher professional development. *Teacher Development*, *18*(3), 303–333. https://doi.org/10.1080/13664530.2014.911775

Sprott, R. A. (2019). Factors that foster and deter advanced teachers' professional development. *Teaching and Teacher Education*, *77*, 321–331. https://doi.org/10.1016/j.tate.2018.11.001

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. https://doi.org/10.1007/BF02294627

Watson, R., & Manning, A. (2008). Factors influencing the transformation of new teaching approaches from a programme of professional development to the classroom. *International Journal of Science Education*, *30*(5), 689–709. https://doi.org/10.1080/09500690701854881

Watt, H. M. G., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-choice scale. *The Journal of Experimental Education*, *75*(3), 167–202. https://doi.org/10.3200/JEXE.75.3.167-202

Watt, H. M. G., & Richardson, P. W. (2008). Motivations, perceptions, and aspirations concerning teaching as a career for different types of beginning teachers. *Learning and Instruction*, *18*(5), 408–428. https://doi.org/10.1016/j.learninstruc.2008.06.002

Watt, H. M. G., & Richardson, P. W. (2015). Teacher Motivation. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 64–71). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26082-0

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wozney, L., Venkatesh, V., & Abrami, P. C. (2006). Implementing computer technologies: Teachers' perceptions and practices. *Journal of Technology and Teacher Education*, *14*(1), 173–207.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. https://ies.ed.gov/ncee/edlabs

Yoon, S. Y., Diefes-Dux, H., & Strobel, J. (2013). First-year effects of an engineering professional development program on elementary teachers. *American Journal of Engineering Education (AJEE)*, *4*(1), 67–84. https://doi.org/10.19030/ajee.v4i1.7859

Appendix

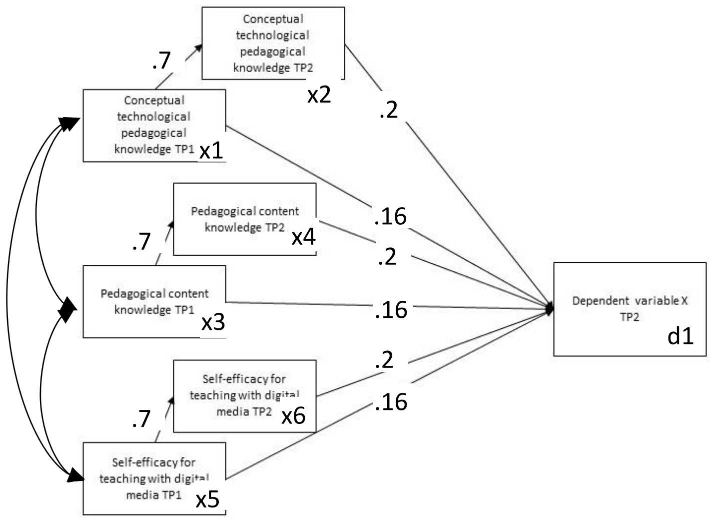
Appendix A. Mplus Power Analysis Regarding RQ2 (Monte Carlo Simulation Study): Example of Mplus Code for *N* = 137 and Effect Size of .3 for Total Effects and a Situation Displayed in Figure A1

Note: In this simulation, we used the most parsimonious model (i.e., the model with the smallest number of items). We did this because we understand the estimated power in this model to be a lower bound on the power of the models (at least with respect to the models shown in Figure 4a), because the power tends to increase when items are added (for example, by explaining a larger proportion of the residual variance of the dependent variable when control variables are added). Therefore, in this simulation, one of the factors of teachers' motivation to implement objectives of PD as operationalized by Osman and Warner (2020) was considered (i.e., task value; 3 items). Furthermore, self-efficacy was assumed to be assessd as a scale score (instead of a latent variable) and no control variables were used. We ran the simulation assuming no missing values. In the analyses, we are interested in total effects.

If while working with the data we find that the values are significantly different from those in the Monte Carlo simulations, then we will perform post hoc power analyses with the values based on the data set used to estimate the power that was achieved.

Figure A1

Conceptual Illustration of Structural Equation Models M4 Used to Answer Research Question 2



Input Code 1 regarding Figure A1

```
MONTECARLO:
        NAMES ARE x1-x6 d1;
        NOBSERVATIONS = 137;
        NREPS = 1000;
        SEED = 53487;

ANALYSIS:
        BOOTSTRAP = 10000;
        MODEL POPULATION:

        !%OVERALL%
        ! independent Variables
```

```
! z-standardized conceptual TPK sum score (TP1: x1; TP2: x2)
! intercepts/mean set to 0
[x1*0];
[x2*0];

! standardized variance
x1*1;
! standardized residual variance
x2*.51;

! z-standardized conceptual PK sum score (TP1: x3; TP2: x4)
! intercepts/mean set to 0
[x3*0];
[x4*0];

! standardized variance
x3*1;
! standardized residual variance
x4*.51;

! z-standardized self-efficacy scale scores (TP1: x5; TP2: x6)

! intercepts/mean set to 0
[x5*0];
[x6*0];

! standardized variance
x5*1;
! standardized residual variance
x6*.51;

! dependent variable
! motivation task value (Osman & Warner): scale score computed
!of 3 items (d1)

! z-standardization (mean set to 0)
[d1@0];

! residual variances of scale score (i.e., dependent variable; standardized)
!1-R^2=1-direct^2+indirect^2=1-((0,16^2*3+0,14^2*3)
d1*.8644;

! regression of predictors (estimated, no reference)
! assumption: relative high temporal stability
x2 ON x1*.70;
x4 ON x3*.70;
x6 ON x5*.70;

! correlation of predictors at first measurement point
x1 WITH x3*.34; !reference: Lachner et al., 2019
x1 WITH x5*.38; !reference: Yildiz Durak, 2021
x3 WITH x5*.4; !assumption based on reference: Yildiz Durak, 2021

! structural equation model (SEM), medium effect sizes (.2 for TP2 and .16
![direct effect] for TP1 => .3 for TP1 [total effect])
d1 ON   x1*.16
        x2*.2
```

```
                x3*.16
                x4*.2
                x5*.16
                x6*.2;
MODEL:
        !%OVERALL%
        ! independent Variables
        ! z-standardized conceptual TPK sum score (TP1: x1; TP2: x2)
        ! intercepts/mean set to 0
        [x1*0];
        [x2*0];

        ! standardized variance
        x1*1;
        ! standardized residual variance
        x2*.51;

        ! z-standardized conceptual PK sum score (TP1: x3; TP2: x4)
        ! intercepts/mean set to 0
        [x3*0];
        [x4*0];

        ! standardized variance
        x3*1;
        ! standardized residual variance
        x4*.51;

        ! z-standardized self-efficacy scale scores (TP1: x5; TP2: x6)

        ! intercepts/mean set to 0
        [x5*0];
        [x6*0];

        ! standardized variance
        x5*1;
        ! standardized residual variance
        x6*.51;

        ! dependent variable
        ! motivation task value (Osman & Warner): scale score computed
        !of 3 items (d1)

        ! z-standardization (mean set to 0)
        [d1*0];

        ! residual variances of scale score (i.e., dependent variable; standardized)
        !1-R^2=1-direct^2+indirect^2=1-((0,16^2*3+0,14^2*3)
        d1*.8644;

        ! regression of predictors (estimated, no reference)
        ! assumption: relative high temporal stability
        x2 ON x1*.70 (bx1x2);
        x4 ON x3*.70 (bx3x4);
        x6 ON x5*.70 (bx5x6);

        ! correlation of predictors at first measurement point
        x1 WITH x3*.34; !reference: Lachner et al., 2019
```

```
x1 WITH x5*.38; !reference: Yildiz Durak, 2021
x3 WITH x5*.4; !assumption based on reference: Yildiz Durak, 2021

! structural equation model (SEM), medium effect sizes (.2 for TP2 and .16
![direct effect] for TP1 => .3 for TP1 [total effect])
d1 ON   x1*.16(bx1d1)
        x2*.2(bx2d1)
        x3*.16(bx3d1)
        x4*.2(bx4d1)
        x5*.16(bx5d1)
        x6*.2(bx6d1);

MODEL CONSTRAINT:
NEW(bx1tot*.3 bx3tot*.3 bx5tot*.3);
bx1tot=bx1x2*bx2d1+bx1d1;
bx3tot=bx3x4*bx4d1+bx3d1;
bx5tot=bx5x6*bx6d1+bx5d1;

OUTPUT: RESIDUAL TECH9;
```

Summarized Results

POWER INFORMATION

Power (=1-β) = .894 d1 on x1 (total effect)

Power (=1-β) = .898 d1 on x3 (total effect)

Power (=1-β) = .902 d1 on x5 (total effect)

MODEL FIT INFORMATION

Number of replications

Requested	1000
Completed	1000

95%-CI-coverage

d1 on x1 (total effect): .941

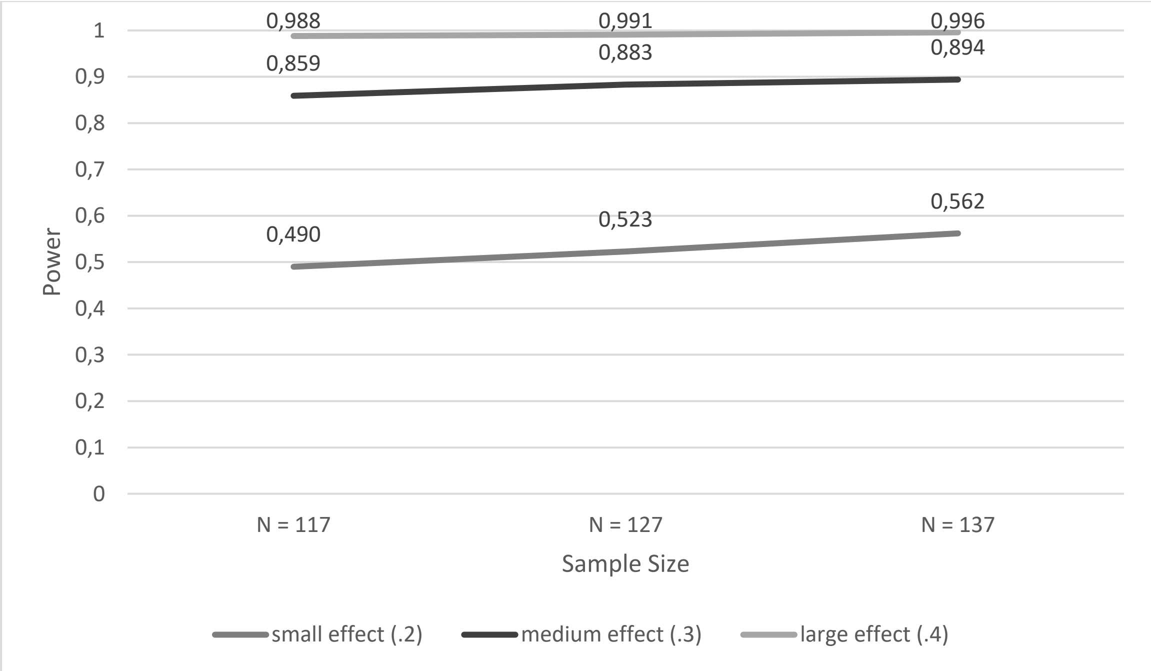
d1 on x3 (total effect): .946

d1 on x5 (total effect): .950

Number of Free Parameters	25
Akaike (AIC)	
Mean:	2396.759
Std Dev	44.819
Number of successful computations	1000
Bayesian (BIC)	
Mean	2469.759
Std Dev	44.819
Number of successful computations	1000
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	
Mean	2390.670
Std Dev	44.819
Number of successful computations	1000
Chi-Square Test of Model Fit	
Degrees of freedom	10
Mean	10.369
Std Dev	4.626
Number of successful computations	1000
RMSEA (Root Mean Square Error Of Approximation)	
Mean	0.024
Std Dev	0.030
Number of successful computations	1000
CFI (Comparative Fit Index)	
Mean	0.994
Std Dev	0.010
Number of successful computations	1000
SRMR (Standardized Root Mean Square Residual)	
Mean	0.085
Std Dev	0.016
Number of successful computations	1000

Figure A2

Overview of the Power per Sample Size (Power Analysis Regarding RQ2 [Monte Carlo Simulation Study], Example for the Total Effect of d1 on x1) – Manifest Modeling of Dependent Variable

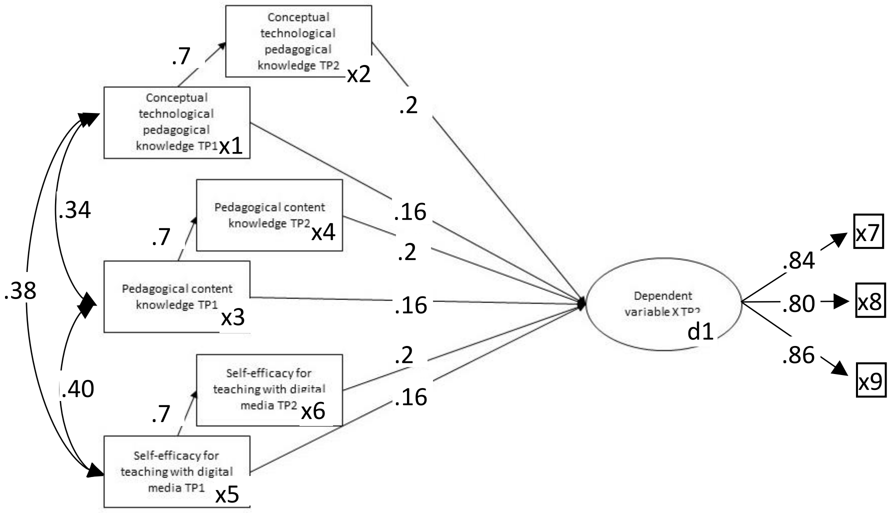


Appendix B. Mplus Power Analysis Regarding RQ2 (Monte Carlo Simulation Study): Example of Mplus Code for *N* = 137 and Effect Size of .3 for the Total Effect and a Situation Displayed in Figure A2

Note: Everything in this simulation is identical to that in Appendix A except that the dependent variable was modeled latently (i.e., task value; 3 items).

Figure A2

Conceptual Illustration of Structural Equation Models M4 Used to Answer Research Question 2



Mplus Input Code 2 regarding Figure A2

```
MONTECARLO:
        NAMES ARE x1-x9;
        NOBSERVATIONS = 137;
        NREPS = 1000;
        SEED = 53487;

ANALYSIS:
        BOOTSRTAP = 10000;
        PROCESSORS = 64;

        MODEL POPULATION:

        !%OVERALL%
        !Independent Variables
        !z-standardized conceptual TPK sum score (TP1: x1; TP2: x2)
        !intercepts/mean set to 0
        [x1*0];
        [x2*0];

        ! standardized variance
        x1*1;
        ! standardized residual variance
        x2*.51;

        ! z-standardized conceptual PK sum score (TP1: x3; TP2: x4)
        !intercepts/mean set to 0
        [x3*0];
```

```
[x4*0];

! standardized variance
x3*1;
! standardized residual variance
x4*.51;

! z-standardized self-efficacy scale scores (TP1: x5; TP2: x6)

! intercepts/mean set to 0
[x5*0];
[x6*0];

! standardized variance
x5*1;
! standardized residual variance
x6*.51;

! dependent variable
! motivation task value (Osman & Warner): 3 items (d1 – x7-x9)
d1 BY x7*.84 x8*.80 x9*.86;

! intercept set to 0
[d1@0];

! residual variances of factors (i.e., dependent variable; standardized)
!1-R^2=1-direct^2+indirect^2=1-((0,16^2*3+0,14^2*3)
d1*.8644;

! intercepts and resiudals of items
!Using information on means: Osman & Warner, 2020, p. 5
[x7*4.10];
[x8*4.20];
[x9*4.75];

! using informations on factor loadings
x7*.2944; !using 1-(loading)^2 for measurement residuals;
x8*.36;
x9*.2604;

! regression of predictors (estimated, no reference)
! assumption: relative high temporal stability
x2 ON x1*.70;
x4 ON x3*.70;
x6 ON x5*.70;

! correlations of predictors at first measurement point
x1 WITH x3*.34; !reference: Lachner et al., 2019
x1 WITH x5*.38; !reference: Yildiz Durak, 2021
x3 WITH x5*.4; !assumption based on reference: Yildiz Durak, 2021

! structural equation model (SEM), medium effect sizes (.2 for TP2 and .16
![direct effect] for TP1 => .3 for TP1 [total effect])
d1 ON x1*.16
      x2*.2
      x3*.16
      x4*.2
```

```
        x5*.16
        x6*.2;

MODEL:
    !%OVERALL%
    ! independent Variables
    ! z-standardized conceptual TPK sum score (TP1: x1; TP2: x2)
    ! intercepts/mean set to 0
    [x1*0];
    [x2*0];

    ! standardized variance
    x1*1;
    ! standardized residual variance
    x2*.51;

    ! z-standardized conceptual PK sum score (TP1: x3; TP2: x4)
    ! intercepts/mean set to 0
    [x3*0];
    [x4*0];

    ! standardized variance
    x3*1;
    ! standardized residual variance
    x4*.51;

    ! z-standardized self-efficacy scale scores (TP1: x5; TP2: x6)

    ! intercepts/mean set to 0
    [x5*0];
    [x6*0];

    ! standardized variance
    x5*1;
    ! standardized residual variance
    x6*.51;

    ! dependent variable
    ! motivation task value (Osman & Warner): 3 items (d1 – x7-x9)
    d1 BY x7@.84 x8*.80 x9*.86;

    ! intercept set to 0
    [d1*0];

    ! residual variances of factors (i.e., dependent variable; standardized)
    !1-R^2=1-direct^2+indirect^2=1-((0,16^2*3+0,14^2*3)
    d1*.8644;

    ! intercepts and resiudals of items
    ! using information on means: Osman & Warner, 2020, p. 5
    [x7*4.10];
    [x8*4.20];
    [x9*4.75];

    ! using informations on factor loadings
    x7*.2944; !using 1-(loading)^2 for measurement residuals;
    x8*.36;
```

```
x9*.2604;

! regression of predictors (estimated, no reference)
! assumption: relative high temporal stability
x2 ON x1*.70(bx1x2);
x4 ON x3*.70(bx3x4);
x6 ON x5*.70(bx5x6);

! correlation of predictors at first measurement point
x1 WITH x3*.34; !reference: Lachner et al., 2019
x1 WITH x5*.38; !reference: Yildiz Durak, 2021
x3 WITH x5*.4; !assumption based on reference: Yildiz Durak, 2021

! residual correlatons
x2 WITH x4*.0 x6*0;
x4 WITH x6*0;

! structural equation model (SEM), medium effect sizes (.2 for TP2 and .16
[direct effect] for TP1 => .3 for TP1 [total effect])
d1 ON   x1*.16(bx1d1)
        x2*.2(bx2d1)
        x3*.16(bx3d1)
        x4*.2(bx4d1)
        x5*.16(bx5d1)
        x6*.2(bx6d1);

MODEL CONSTRAINT:
NEW(bx1tot*.3 bx3tot*.3 bx5tot*.3);
bx1tot=bx1x2*bx2d1+bx1d1;
bx3tot=bx3x4*bx4d1+bx3d1;
bx5tot=bx5x6*bx6d1+bx5d1;

OUTPUT:
    RESIDUAL TECH9;
```

Summarized Results

POWER INFORMATION
Power (=1-β) = .854 d1 on x1 (total effect)
Power (=1-β) = .837 d1 on x3 (total effect)
Power (=1-β) = .845 d1 on x5 (total effect)

MODEL FIT INFORMATION
Number of replications

Requested	1000
Completed	1000

95%-CI-coverage
d1 on x1 (total effect): .950
d1 on x3 (total effect): .944
d1 on x5 (total effect): .940

Number of Free Parameters	27
Akaike (AIC)	
Mean:	2974.182
Std Dev	50.069

Number of successful computations	1000

Bayesian (BIC)

Mean	3053.022
Std Dev	50.069
Number of successful computations	1000

Sample-Size Adjusted BIC (n* = (n + 2) / 24)

Mean	2967.606
Std Dev	50.069
Number of successful computations	1000

Chi-Square Test of Model Fit

Degrees of freedom	27
Mean	27.765
Std Dev	7.427
Number of successful computations	1000

RMSEA (Root Mean Square Error Of Approximation)

Mean	0.020
Std Dev	0.023
Number of successful computations	1000

CFI (Comparative Fit Index)

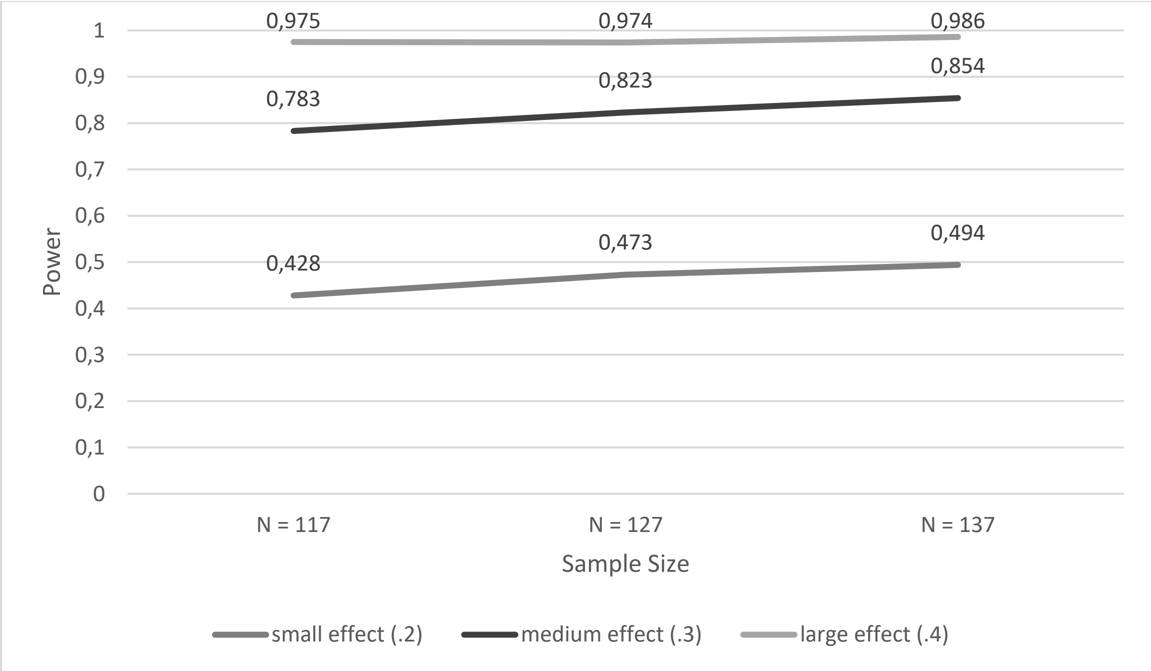
Mean	0.994
Std Dev	0.008
Number of successful computations	1000

SRMR (Standardized Root Mean Square Residual)

Mean	0.101
Std Dev	0.027
Number of successful computations	1000

Figure A3

Overview of the Power per Sample Size (Power Analysis Regarding RQ2 [Monte Carlo Simulation Study], Example for the Total Effect of d1 on x1) – Latent Modeling of Dependent Variable



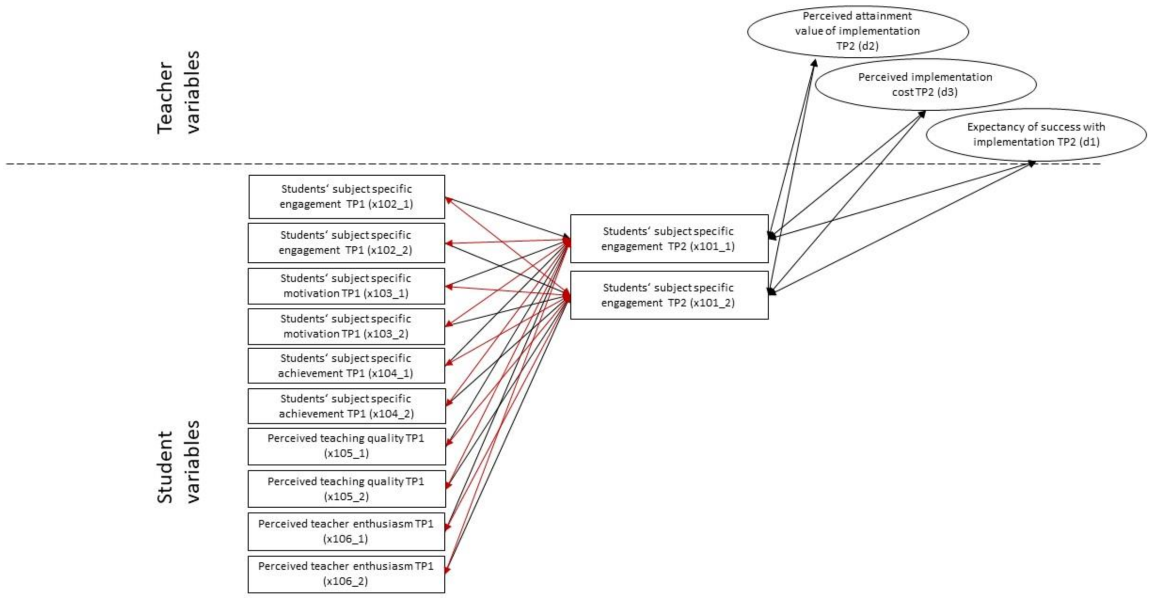
Appendix C. Mplus Power Analysis Regarding RQ3 (Monte Carlo Simulation Study): Example of Mplus Code for *N* = 140 Teachers, 210 Classes, 10 Students per Class, an Effect Size of .3, and a Situtation Displayed in Figure A2

Note: In this simulation, all student variables were assumed to be assessd as a scale scores (instead of a latent variables) and no control variables were used. We ran the simulation assuming no missing values. Correlations of predictors and dependent variables are not shown in Figure A5. As can be seen from the syntax, correlations are allowed and equality constraints are applied. Outcomes per class (x101_1 and x101_2 in Figure A5) are predicted with all predictors from the respective class (black unidirectional arrows) and are correlated with predictors from the respective other class (red bidirectional arrows). In addition, the residuals of the outcome per class are correlated with each other and with the three teacher variables. The teacher variables are also correlated with each other.

If while working with the data we find that the values are significantly different from those in the Monte Carlo simulations, then we will perform post hoc power analyses with the values based on the data set used to estimate the power that was achieved.

Figure A5

Conceptual Illustration of Structural Equation Models M6 Used to Answer Research Question 3



Data Generation With SAS and Mplus Simulation Study

Input Code (data generation with SAS)

```
%LET replicates=1000;

%LET teach1=70; * number of teachers with a single class;
%LET teach2=70; * number of teachers with two classes;

%LET clsize=10; * class size;

%LET vart=0.20; * variance predictors/outcome at teacher level;
%LET varc=0.10; * variance predictors/outcome at class level;

%LET seed=12345;
```

```
DATA test;
	DO rep=1 TO &replicates;
		DO teach=1 TO &teach1+&teach2;
			fspt=SQRT(.10)*NORMAL(&seed); * factor to generate correlations among
			predictors (r=.10) at teacher level;
			x102t=SQRT(&vart)*(fspt+SQRT(.90)*NORMAL(&seed));
			x103t=SQRT(&vart)*(fspt+SQRT(.90)*NORMAL(&seed));
			x104t=SQRT(&vart)*(fspt+SQRT(.90)*NORMAL(&seed));
			x105t=SQRT(&vart)*(fspt+SQRT(.90)*NORMAL(&seed));
			x106t=SQRT(&vart)*(fspt+SQRT(.90)*NORMAL(&seed));

			* r=.10, 5 predictors => variance predicted = 0.2**2*(5*&vart + 20*(&vart*.10)) =
			0.2**2*7*&vart;
			resx101t=SQRT(&vart-0.2**2*7*&vart)*NORMAL(&seed);
			x101t=0.2*x102t+0.2*x103t+0.2*x104t+0.2*x105t+0.2*x106t+resx101t;

			*** teacher questionnaire factors ***;
			* correlation of independent variables on teacher level

				d1	d2	d3
			d2	.66
			d3	.37	.44
			x101t	.30	.30	.30

			residual correlations for d, var(explained by resx101t)=.3**2=0.09
			factor res1 => variance explained=0.28 for d1-d3
			factor res2 => variance explained=0.07 for d2, d3
			factor res3 => variance explained=0.29 for d1, d2;

			res1=NORMAL(&seed);
			res2=NORMAL(&seed);
			res3=NORMAL(&seed);

			d1=.30*resx101t/SQRT(&vart-
			0.2**2*7*&vart)+SQRT(.28)*res1+SQRT(.29)*res3+SQRT(1-.09-.28-
			.29)*NORMAL(&seed);
			d2=.30*resx101t/SQRT(&vart-
			0.2**2*7*&vart)+SQRT(.28)*res1+SQRT(.07)*res2+SQRT(.29)*res3+SQRT(1-.09-.28-
			.07-.29)*NORMAL(&seed);
			d3=.30*resx101t/SQRT(&vart-
			0.2**2*7*&vart)+SQRT(.28)*res1+SQRT(.07)*res2+SQRT(1-.09-.28-
			.07)*NORMAL(&seed);

			* measurement models for d1-d3;
			* Motivation Expectancy (Osman & Warner): 3 items (d1 - x13-x15);
			x13=.72*d1+SQRT(1-.72**2)*NORMAL(&seed);
			x14=.65*d1+SQRT(1-.65**2)*NORMAL(&seed);
			x15=.69*d1+SQRT(1-.69**2)*NORMAL(&seed);

			* Motivation Task Value (Osman & Warner): 3 items (d2 - x16-x18);
			x16=.84*d2+SQRT(1-.84**2)*NORMAL(&seed);
			x17=.80*d2+SQRT(1-.80**2)*NORMAL(&seed);
			x18=.86*d2+SQRT(1-.86**2)*NORMAL(&seed);

			* Motivation Cost (Osman & Warner): 3 items (d3 - x19-x21);
			x19=.84*d3+SQRT(1-.84**2)*NORMAL(&seed);
```

```
x20=.85*d3+SQRT(1-.85**2)*NORMAL(&seed);
x21=.75*d3+SQRT(1-.75**2)*NORMAL(&seed);

IF teach LE &teach1 THEN nc=1; ELSE nc=2;
fspc1=SQRT(.10)*NORMAL(&seed); * factor to generate correlations among
predictors (r=.10) at class level #1;
x102c1=SQRT(&varc)*(fspc1+SQRT(.90)*NORMAL(&seed));
x103c1=SQRT(&varc)*(fspc1+SQRT(.90)*NORMAL(&seed));
x104c1=SQRT(&varc)*(fspc1+SQRT(.90)*NORMAL(&seed));
x105c1=SQRT(&varc)*(fspc1+SQRT(.90)*NORMAL(&seed));
x106c1=SQRT(&varc)*(fspc1+SQRT(.90)*NORMAL(&seed));

* r=.10, 5 predictors => variance predicted = 0.2**2*(5*&varc + 20*(&varc*.10)) =
0.2**2*7*&varc;
x101c1=0.2*x102c1+0.2*x103c1+0.2*x104c1+0.2*x105c1+0.2*x106c1+SQRT(&varc-
0.2**2*7*&varc)*NORMAL(&seed);

x101c2=.; x102c2=.; x103c2=.; x104c2=.; x105c2=.; x106c2=.;
IF nc=2 THEN DO;
        fspc2=SQRT(.10)*NORMAL(&seed); * factor to generate correlations among
        predictors (r=.10) at class level #2;
        x102c2=SQRT(&varc)*(fspc2+SQRT(.90)*NORMAL(&seed));
        x103c2=SQRT(&varc)*(fspc2+SQRT(.90)*NORMAL(&seed));
        x104c2=SQRT(&varc)*(fspc2+SQRT(.90)*NORMAL(&seed));
        x105c2=SQRT(&varc)*(fspc2+SQRT(.90)*NORMAL(&seed));
        x106c2=SQRT(&varc)*(fspc2+SQRT(.90)*NORMAL(&seed));

        * r=.10, 5 predictors => variance predicted = 5*&varc + 20*(&varc*.10) =
        7*&varc;
        x101c2=0.2*x102c2+0.2*x103c2+0.2*x104c2+0.2*x105c2+0.2*x106c2+SQR
        T(&varc)*NORMAL(&seed);
END;

DO stud=1 TO &clsize;
        fspc1s=SQRT(.10)*NORMAL(&seed); * factor to generate correlations
        among predictors (r=.10) at student level (class #1);
        x102c1s=SQRT(1-&vart-&varc)*(fspc1s+SQRT(.90)*NORMAL(&seed));
        x103c1s=SQRT(1-&vart-&varc)*(fspc1s+SQRT(.90)*NORMAL(&seed));
        x104c1s=SQRT(1-&vart-&varc)*(fspc1s+SQRT(.90)*NORMAL(&seed));
        x105c1s=SQRT(1-&vart-&varc)*(fspc1s+SQRT(.90)*NORMAL(&seed));
        x106c1s=SQRT(1-&vart-&varc)*(fspc1s+SQRT(.90)*NORMAL(&seed));

        * r=.10, 5 predictors => variance predicted = 0.2**2*(5*(1-&vart-&varc) +
        20*(1-&vart-&varc)*.10) = 0.2**2*7*(1-&vart-&varc);
        x101c1s=0.2*x102c1s+0.2*x103c1s+0.2*x104c1s+0.2*x105c1s+0.2*x106c1s
        +SQRT((1-&vart-&varc)-0.2**2*7*(1-&vart-&varc))*NORMAL(&seed);
        x101c2s=.; x102c2s=.; x103c2s=.; x104c2s=.; x105c2s=.; x106c2s=.;

        IF nc=2 THEN DO;
                fspc2s=SQRT(.10)*NORMAL(&seed); * factor to generate
                correlations among predictors (r=.10) at student level (class #2);
                x102c2s=SQRT(1-&vart-
                &varc)*(fspc2s+SQRT(.90)*NORMAL(&seed));
                x103c2s=SQRT(1-&vart-
                &varc)*(fspc2s+SQRT(.90)*NORMAL(&seed));
                x104c2s=SQRT(1-&vart-
                &varc)*(fspc2s+SQRT(.90)*NORMAL(&seed));
```

```
                        x105c2s=SQRT(1-&vart-
                        &varc)*(fspc2s+SQRT(.90)*NORMAL(&seed));
                        x106c2s=SQRT(1-&vart-
                        &varc)*(fspc2s+SQRT(.90)*NORMAL(&seed));

                        * r=.10, 5 predictors => variance predicted = 0.2**2*(5*(1-&vart-
                        &varc) + 20*(1-&vart-&varc)*.10) = 0.2**2*7*(1-&vart-&varc);
                        x101c2s=0.2*x102c2s+0.2*x103c2s+0.2*x104c2s+0.2*x105c2s+0.2
                        *x106c2s+SQRT((1-&vart-&varc)-0.2**2*7*(1-&vart-
                        &varc))*NORMAL(&seed);
                    END;

                    x102_1=x102t+x102c1+x102c1s;
                    x103_1=x103t+x103c1+x103c1s;
                    x104_1=x104t+x104c1+x104c1s;
                    x105_1=x105t+x105c1+x105c1s;
                    x106_1=x106t+x106c1+x106c1s;

                    x101_1=x101t+x101c1+x101c1s;

                    x102_2=x102t+x102c2+x102c2s;
                    x103_2=x103t+x103c2+x103c2s;
                    x104_2=x104t+x104c2+x104c2s;
                    x105_2=x105t+x105c2+x105c2s;
                    x106_2=x106t+x106c2+x106c2s;

                    x101_2=x101t+x101c2+x101c2s;

                    OUTPUT;
                END;
            END;
        END;
RUN;

%MACRO export;
    %DO rep=1 %TO &replicates;
        DATA _NULL_; SET test(WHERE=(rep=&rep));
            FILE "E:\mc&rep..dat";
            PUT teach
                x101_1 x102_1 x103_1 x104_1 x105_1 x106_1
                x101_2 x102_2 x103_2 x104_2 x105_2 x106_2
                x13-x21;
        RUN;
    %END;
    DATA _NULL_;
        FILE "E:\mc.dat";
        %DO rep=1 %TO &replicates;
            PUT "mc&rep..dat";
        %END;
    RUN;
%MEND;
%export;
```

Mplus Input Code

```
DATA:
        FILE=mc.dat;
        TYPE=MONTECARLO;

VARIABLE:
        NAMES=teach
        x101_1 x102_1 x103_1 x104_1 x105_1 x106_1
        x101_2 x102_2 x103_2 x104_2 x105_2 x106_2
        x13-x21;
        CLUSTER=teach;
        MISSING=.;

ANALYSIS:
        TYPE=COMPLEX;
        ITERATIONS=10000;

MODEL:
        ! *** COVARIATES ***
        ! variances
        x102_1*1(vx102);
        x102_2*1(vx102);
        x103_1*1(vx103);
        x103_2*1(vx103);
        x104_1*1(vx104);
        x104_2*1(vx104);
        x105_1*1(vx105);
        x105_2*1(vx105);
        x106_1*1(vx106);
        x106_2*1(vx106);

        ! class-specific correlations;
        x102_1 WITH x103_1*0.1 x104_1*0.1 x105_1*0.1 x106_1*0.1(cov23-cov26);
        x103_1 WITH x104_1*0.1 x105_1*0.1 x106_1*0.1(cov34-cov36);
        x104_1 WITH x105_1*0.1 x106_1*0.1(cov45-cov46);
        x105_1 WITH x106_1*0.1(cov56);

        x102_2 WITH x103_2*0.1 x104_2*0.1 x105_2*0.1 x106_2*0.1(cov23-cov26);
        x103_2 WITH x104_2*0.1 x105_2*0.1 x106_2*0.1(cov34-cov36);
        x104_2 WITH x105_2*0.1 x106_2*0.1(cov45-cov46);
        x105_2 WITH x106_2*0.1(cov56);

        ! across class correlations (teacher-specific variance);
        ! class-specific correlations are .10
        ! variances at teacher level are 0.20
        ! => pairwise covariances are 0.2 (e.g. x102_1, x102_2)
        ! all other covariances are 0.2*.1=0.02 with equality constraints:
        ! e.g. x102_1 WITH x103_2 equal to x103_1 WITH x102_2 etc.
        x102_1 WITH x102_2*0.2
        x103_2*0.02(cx102_3)
        x104_2*0.02(cx102_4)
        x105_2*0.02(cx102_5)
        x106_2*0.02(cx102_6);
        x103_1 WITH x102_2*0.02(cx102_3)
        x103_2*0.2
        x104_2*0.02(cx103_4)
```

```
x105_2*0.02(cx103_5)
x106_2*0.02(cx103_6);
x104_1 WITH x102_2*0.02(cx102_4)
x103_2*0.02(cx103_4)
x104_2*0.2
x105_2*0.02(cx104_5)
x106_2*0.02(cx104_6);
x105_1 WITH x102_2*0.02(cx102_5)
x103_2*0.02(cx103_5)
x104_2*0.02(cx104_5)
x105_2*0.2
x106_2*0.02(cx105_6);
x106_1 WITH x102_2*0.02(cx102_6)
x103_2*0.02(cx103_6)
x104_2*0.02(cx104_6)
x105_2*0.02(cx105_6)
x106_2*0.2;

! means
[X102_1*0](mx102);
[X103_1*0](mx103);
[X104_1*0](mx104);
[X105_1*0](mx105);
[X106_1*0](mx106);

[X102_2*0](mx102);
[X103_2*0](mx103);
[X104_2*0](mx104);
[X105_2*0](mx105);
[X106_2*0](mx106);

! *** OUTCOME ON COVARIATES ***
x101_1 ON x102_1*0.2(b102)
x103_1*0.2(b103)
x104_1*0.2(b104)
x105_1*0.2(b105)
x106_1*0.2(b106);
x101_2 ON x102_2*0.2(b102)
x103_2*0.2(b103)
x104_2*0.2(b104)
x105_2*0.2(b105)
x106_2*0.2(b106);

! residual variance x101
! predictor variances var(x)=1
! r=.10, 5 predictors
! => variance predicted = 0.2**2*(5*var(x) + 20*(var(x)*.10))
! = 0.2**2*7*var(x) = 0.28;
! => residual variance = 1-0.28 = 0.72
x101_1*0.72(rx101);
x101_2*0.72(rx101);

! residual correlation x101
! residual variance = 0.72
! r = 0.2
! => cov = 0.2*0.72 = 0.144
x101_1 WITH x101_2*0.144;
```

```
! intercepts
[X101_1*0](ix101);
[X101_2*0](ix101);

! *** TEACHER MEASURES ***
! measurement models
d1 BY x13*.72 x14*.65 x15*.69;
d2 BY x16*.84 x17*.80 x18*.86;
d3 BY x19*.84 x20*.85 x21*.75;

d1-d3@1;

! residuals based on factor loadings
x13*.4816; !using 1-(loading)^2 for measurement residuals;
x14*.5775;
x15*.5239;

x16*.2944;
x17*.36;
x18*.2604;

x19*.2944;
x20*.2775;
x21*.4375;

! intercepts indicators of teacher factors
[X13*0 X14*0 X15*0 X16*0 X17*0 X18*0 X19*0 X20*0 X21*0];

! intercorrelations teacher factors;
d1 WITH d2*.66 d3*.37;
d2 WITH d3*.44;

! *** CORRELATION TEACHER FACTORS WITH OUTCOME (RESIDUAL) ***;
! r=.30 at teacher level;
! var(resx101t) = 0.144, var(x101) = 1;
! => r=.30=cov(x101_t, d_i)/SQRT(0.144*1);
! => cov(x101, d_i)=.30*SQRT(0.144)=0.11384;
d1-d3 WITH x101_1*0.11384(cx101d1-cx101d3);
d1-d3 WITH x101_2*0.11384(cx101d1-cx101d3);

! *** CORRELATIONS TEACHER FACTORS WITH PREDICTORS ***;
d1 WITH x102_1*0(cd1x102)
x103_1*0(cd1x103)
x104_1*0(cd1x104)
x105_1*0(cd1x105)
x106_1*0(cd1x106);
d2 WITH x102_1*0(cd2x102)
x103_1*0(cd2x103)
x104_1*0(cd2x104)
x105_1*0(cd2x105)
x106_1*0(cd2x106);
d3 WITH x102_1*0(cd3x102)
x103_1*0(cd3x103)
x104_1*0(cd3x104)
x105_1*0(cd3x105)
x106_1*0(cd3x106);
```

```
d1 WITH x102_2*0(cd1x102)
x103_2*0(cd1x103)
x104_2*0(cd1x104)
x105_2*0(cd1x105)
x106_2*0(cd1x106);
d2 WITH x102_2*0(cd2x102)
x103_2*0(cd2x103)
x104_2*0(cd2x104)
x105_2*0(cd2x105)
x106_2*0(cd2x106);
d3 WITH x102_2*0(cd3x102)
x103_2*0(cd3x103)
x104_2*0(cd3x104)
x105_2*0(cd3x105)
x106_2*0(cd3x106);

OUTPUT:
STANDARDIZED;
```

Summarized Results

POWER INFORMATION
Power (=1-β) = .696 d1 WITH x101_1
Power (=1-β) = .696 d1 WITH x101_2
Power (=1-β) = .736 d2 WITH x101_1
Power (=1-β) = .736 d2 WITH x101_2
Power (=1-β) = .743 d3 WITH x101_1
Power (=1-β) = .743 d3 WITH x101_2

MODEL FIT INFORMATION
Number of replications

Requested	1000
Completed	1000

95%-CI-coverage
d1 WITH x101_1: .938
d1 WITH x101_2: .938
d2 WITH x101_1: .938
d2 WITH x101_2: .938
d3 WITH x101_1: .942
d3 WITH x101_2: .942

Number of Free Parameters	91
Akaike (AIC)	
Mean:	64602.966
Std Dev	542.683
Number of successful computations	1000
Bayesian (BIC)	
Mean	65080.191
Std Dev	542.683
Number of successful computations	1000
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	

Mean	64791.118
Std Dev	542.683
Number of successful computations	1000
Chi-Square Test of Model Fit	
Degrees of freedom	161
Mean	182.398
Std Dev	25.677
Number of successful computations	1000
RMSEA (Root Mean Square Error Of Approximation)	
Mean	0.009
Std Dev	0.006
Number of successful computations	1000
CFI (Comparative Fit Index)	
Mean	0.984
Std Dev	0.015
Number of successful computations	1000
SRMR (Standardized Root Mean Square Residual)	
Mean	0.041
Std Dev	0.004
Number of successful computations	1000