



# Practice Recommendations or Not? The LoGeT Model as Empirical Approach to Generate Localized, Generalized, and Transferable Evidence

Andreas Lachner<sup>1</sup> · Leonie Sibley<sup>2</sup> · Salome Wagner<sup>3</sup>

Accepted: 26 February 2024 / Published online: 13 March 2024  
© The Author(s) 2024

## Abstract

In educational research, there is the general trade-off that empirical evidence should be generalizable to be applicable across contexts; at the same time, empirical evidence should be as specific as possible to be localizable in subject-specific educational interventions to successfully transfer the empirical evidence to educational practice. This trade-off is further increased by the fact that the diverse instructional contexts, such as school or student characteristics constrain the applicability of empirical evidence. Several approaches have been proposed to address this issue, however, emphasized the different problems (i.e., localization, generalization, transferability) rather in an isolated manner. To this end, in this article, we introduce a synergistic approach, the LoGeT (localize, generalize, transfer) model, which systematically integrates co-design (localization strategies) and ManyClasses principles (generalization strategies) with co-constructive transfer activities, to generate empirical evidence that may be applicable in educational practice. To illustrate the LoGeT model, we present three long-term projects, covering different granularities and durations of educational interventions across different fields of education (teacher education, adaptive teaching, non-interactive teaching) that successfully applied the LoGeT approach. Finally, we outline further directions for future iterations of the LoGeT model. We hope that the LoGeT approach may be a stimulus to guide researchers as well as practitioners alike to design generalizable and evidence-based educational interventions that are rooted in localized instructional contexts.

**Keywords** Evidence-based education · Experimental research · Science communication · Instructional effectiveness · Transfer

---

✉ Andreas Lachner  
andreas.lachner@uni-tuebingen.de

<sup>1</sup> Institute of Education, Tübingen Center for Digital Education, University of Tübingen, Wilhelmstraße 31, 72074 Tübingen, Germany

<sup>2</sup> Institute of Education, University of Tübingen, Tübingen, Germany

<sup>3</sup> Tübingen Center for Digital Education, University of Tübingen, Tübingen, Germany

## Introduction

There is the common plea that educational interventions should be based on empirical evidence in order to be effective (e.g., Bromme et al., 2014; Slavin, 2002, 2020). To this end, educational researchers are commonly advised to provide implications that are based on the obtained (empirical) evidence for educational practice (Slavin, 2020). At the same time, we as researchers know that it is often difficult to directly transfer findings of a single empirical study to the “wild” (Renkl, 2013; Robinson et al., 2013). Therefore, there are justifiable concerns—particularly fed by the replication crisis (Maxwell et al., 2015)—that such practice recommendations have strong implications for educational practice, as for instance the instructional context (Kaplan et al., 2020; Turner & Nolen, 2015), but also differences in teaching or intra-individual differences in students’ pre-requisites may vary largely across situations, making it difficult to directly transfer the obtained empirical evidence. These effects may be intensified, when implications should be drawn for (younger) school students in the classroom, as often empirical studies on the effectiveness of interventions are investigated within laboratory conditions utilizing (convenience) samples of mature university students in isolated learning settings (Brod, 2021; Jacob et al., 2022; Lachner et al., 2022).

Consequently, it can be argued that educational research constantly faces two essential crises. On the one hand, we are suffering from the replication crisis, which demonstrated that it is often difficult or even impossible to directly reproduce and generalize the findings of scientific studies (Maxwell et al., 2015; Sweller, 2023). On the other hand, we are in a transfer crisis, as it is difficult to localize scientific evidence into different application contexts and make them transferable to educational practice (Fyfe et al., 2021; Renkl, 2013). To tackle the replication and transfer crises, several methodological approaches have been realized to warrant the ecological validity of educational interventions. For instance, co-design approaches explicitly bring educational practitioners and researchers together during the design process (Roschelle et al., 2006; Severance et al., 2016; Slattery et al., 2020). The mutual construction processes within co-design approaches are regarded to contribute to more applicable and effective educational interventions, as they clearly consider teachers as experts of teaching (Lachner et al., 2016; Leinhardt & Putnam, 1986) in the process of improving teaching and learning (Severance et al., 2016). At the same time, it has to be acknowledged that the generalizability of co-design approaches may be limited, as the obtained empirical evidence is localized to the specific instructional context in which the intervention was implemented. Relatedly, ManyClasses studies (Fyfe et al., 2021), a recent quantitative methodological approach to provide generalizable evidence, are implemented to experimentally investigate psychological principles in different instructional field-conditions. Thus, rather than conducting one experimental study in just one setting (e.g., secondary biology education), in a ManyClasses experiment, the same experimental set-up (e.g., feedback versus no-feedback) is implemented across multiple courses spanning a range of topics, teachers, and

student populations. The findings are then aggregated via meta-analytic procedures, and potential boundary conditions can be investigated via moderation analyses (Fyfe et al., 2021). However, in ManyClasses experiments, teacher implementations are rather treated as noise, as the active role of teachers during the design process is considered only to a less pronounced extent.

Finally, from a transfer perspective, it is an open question, how implications of (empirical) research may reach the “wild,” that is, the teachers and schools that were not participating in the original design studies. To this end, several initiatives, such as the What Works Clearinghouse (see Slavin, 2020, for an overview), aim to summarize the (often complex) findings of empirical studies and provide a comprehensible and transferable overview for teachers and teacher educators. Given that the process of transfer is mainly directional via one institution that processes (published) empirical findings, it may be difficult to keep up with the current research activities, however.

To this end, in this article, we propose an integrated approach, the LoGeT (localize, generalize, transfer) model, which systematically combines co-design and ManyClasses principles with transfer activities, to (1) co-constructively design and implement educational interventions, (2) investigate the effectiveness of those interventions across different instructional settings by applying meta-analytic techniques, and (3) provide transferal outlets to comprehensibly communicate the obtained evidence. To illustrate the LoGeT model, we additionally present the processes of three ongoing long-term projects of different educational granularities (i.e., teacher education, adaptive teaching, learning by non-interactive teaching) in which we deliberately followed the LoGeT approach.

## Empirical Evidence and Educational Practice: Two Worlds Apart?

Given that evidence-based medicine is often regarded as a role model for educational research (Bromme et al., 2014; Slavin, 2020), empirical educational research adopted the use of (quantitative) findings as primary sources for practice recommendations. Particularly, meta-analyses, which synthesize findings of primary quantitative studies, are regarded to provide robust estimates on effects of educational interventions (Renkl, 2022; Seidel et al., 2017; Slavin, 2020).

Teachers, however, rarely use empirical evidence to legitimate their decisions. Instead, they rely on anecdotal evidence and prior experiences (of colleagues) to base their educational decisions (e.g., Bråten & Ferguson, 2015; Lortie, 1975; Weinhuber et al., 2019). For instance, recent research demonstrated that mathematical teachers rarely include conceptual information about the underlying processes in their instructional explanations (Lachner et al., 2016; Weinhuber et al., 2019), although conceptual information has demonstrated to be effective particularly in initial phases of skill acquisition, as it may enhance germane processing of procedural information (Bokosmaty et al., 2015; Lachner et al., 2019; van Gog et al., 2008). Research indicated that a crucial reason for the omission of conceptual information is that teachers rather use experiential knowledge during reasoning: In their Study 2, Lachner et al. (2019) asked mathematics teachers ( $N=69$ ) to judge the instructional

quality of pre-validated explanations which systematically varied in the presence/omission of conceptual information. The authors found that the teachers primarily relied on previous experiences while judging the explanations, as they regarded the addition of conceptual information to result in additional students' demands rather than germane processing of the instructional explanations, likely a result of previous experiences with single students. One crucial reason for this finding is that teachers may find it difficult to apply the obtained findings in their teaching. Thus, the perceived utility of empirical findings may be too low, as they may not immediately be able to answer teachers' questions when planning and realizing subject-matter teaching (Renkl, 2022). This assumption is in line with a recent survey study by Farley-Ripple et al. (2022). The authors conducted an online survey with  $N=4415$  educators and asked them to report their attitudes towards the use of empirical evidence in educational practice. The educators indeed acknowledged the general use of empirical evidence for educational practice; at the same time, however, around one-third of them lamented that research is not localized enough to provide constructive guidance regarding the solution of their problems at schools. The authors inferred that teachers require more localized empirical evidence that comes from a context that resembles teachers' own one and that may be adapted to the diverse school contexts (e.g., such as school tracks, subjects, students' prerequisites). In addition to localization and generalization of empirical evidence, recent research also emphasized the role of perceived costs for the application of empirical evidence. Following the theory of planned behavior, Greisel et al. (2023) asked pre-service teachers ( $N=157$ ) to report their motivational prerequisites to engage with empirical evidence. Additionally, the pre-service teachers were required to assess critical classroom situations. The authors found that the quality of assessments was negatively related to self-efficacy ( $b = -0.17$ ) and the perceived costs to engage with scientific evidence ( $b = -0.16$ ), explaining 9.2% of variance, suggesting that feasible measures of direct transfer are required to reduce the perceived costs to engage with scientific evidence.

## Measures to Enhance the Localization, Generalization, and Feasibility of Scientific Evidence

The previous findings highlighted the assumption that current instructional research on educational interventions (see Mayer, 2023, for an overview) do not necessarily answer the myriads of questions of educational practice regarding the effectiveness of educational interventions. Against this background, several independent approaches have been explored to either localize, generalize, or transfer scientific evidence to educational practice.

### Realizing Co-design Approaches to Localize Scientific Evidence

Co-design originated from Scandinavian participatory design traditions (Bødker, 1996) in which stakeholders have been actively involved during the design process. To this end, co-design has a long tradition as a generic principle in human-computer

interaction (Iniesto et al., 2022) and in the learning sciences (Roschelle et al., 2006). Co-design approaches commonly differ from researcher-led “top-down approaches,” as they consider and integrate teachers’ everyday practices by actively involving them as agents in the design process. Teachers are seen as professional contributors and source of educational innovation (Roschelle et al., 2006; Severance et al., 2016). Thus, related terms, such as co-construction or co-creation, have been used interchangeably in these contexts (Iniesto et al., 2022).

Commonly co-design comprises four team-based, reciprocal, and interactive design phases. First, in the contextual inquiry phase, researchers and teachers set a common ground and work on a mutual understanding of the goals, the context, and problems the intervention should be targeting, as well as negotiate the individual contributions of the team members. Second, during the participatory design phase, distinct design principles are derived in close cooperation with the stakeholders, for instance, by conducting design thinking workshops. Third, during the product design phase, a design prototype is developed to define potential use cases. Fourth, in accordance with generic forms of design-based research, in the prototype-as-hypothesis phase, functional rapid prototypes are iteratively tested within the intended learning environment to derive a potential “functioning” educational intervention. For this purpose, often qualitative methods are used to get a rich understanding of the localized boundary conditions of the educational intervention (Iniesto et al., 2022). In a final prototyping phase, often quantitative studies, such as randomized controlled field studies or classroom experiments (see Holstein et al., 2019; Yannier et al., 2022, for methodological examples), are additionally implemented, which test the effectiveness of the particular intervention. The iterative and active involvement of teachers and researchers alike are regarded to contribute to more applicable educational innovations, as they clearly consider teachers in the process of improving teaching and learning (Severance et al., 2016) and concretely build on local strategies to enhance concrete educational interventions. At the same time, one pitfall of these localized design strategies is that the obtained findings may only hold true for a specific context for which the intervention was targeted for (see also Zheng, 2015). Therefore, it is difficult to assess the effectiveness of the educational intervention and generalize the obtained evidence to other contexts (e.g., different subjects, student populations, student prerequisites).

### **Realizing ManyClasses Approaches to Generalize Scientific Evidence**

To provide generalizable and ecologically valid evidence regarding the effectiveness of educational interventions, Fyfe et al. (2021) recently proposed the ManyClasses approach. The ManyClasses approach extends previous experimental methods, such as classroom experiments, as researchers not only test a distinct hypothesis within one context, that is, one education experiment in one course (e.g., 10th grade biology class on osmosis), but rather implement a myriad of experiments, which test the same principle or hypothesis across different contexts (subjects, classes). Thus, the single experiments may function as individual conceptual replications in ecologically valid classroom contexts. To test potential effects of the intervention and

of different boundary conditions, meta-analytic strategies can be applied which use individual participant data (Riley et al., 2021; Veroniki et al., 2023). ManyClasses studies were adopted primarily in applied cognitive psychology contexts.

One of the rare examples of a ManyClasses study is the one by Fyfe et al. (2021). The authors investigated the effect of timing of feedback on students' learning. Therefore, they realized a within-participants design and randomly assigned students to different treatment orders per class assignment (1 delayed feedback, 2 immediate feedback versus 1 immediate feedback, 2 delayed feedback). Additionally, they randomly assigned classes to whether they received an incentive or not, as between-classroom factor. They implemented the experimental setup in 38 different classrooms of various disciplines (e.g., history, chemistry, psychology), comprising 46 different instructors and  $N=2081$  students. Surprisingly, the authors did not find an effect of delayed versus immediate feedback and study incentives. Preregistered moderation analyses of 40 different moderators did not obtain strong evidence for systematic interactions.

Sana and Yan (2022) realized a within-participants design in which they compared the effectiveness of interleaving versus blocking concepts during retrieval practice in eight STEM classrooms (biology, chemistry, science, physics) with 9th to 12th grade students ( $N=155$ ). Consistent with the retrieval practice effect (Roediger & Butler, 2011; Yang et al., 2021), the findings revealed that students performed better on interleaved practice than blocked practice across classes, attesting the robustness of the findings in STEM domains (see also Brunmair & Richter, 2019; Taylor & Rohrer, 2010).

Together, ManyClasses experiments may provide a potential lens for investigating robust and generalizable evidence across different contexts. It is in line with recent movements of transparent and reproducible research practices. Thus, it may help researchers to trace whether an effect is "expectable" in their localized teaching context. At the same time, however, previous examples of ManyClasses approaches were rather realized as a researcher-centered top-down approach, as the main scope was to theoretically test a psychological research question in a set of diverse contexts. This procedure may have the danger of not targeting the current needs of educational practice. That said, additional measures are needed to adequately inform stakeholders and transfer the obtained evidence into educational practice.

### Realizing Educational Outreach to Transfer Scientific Evidence

Due to the increasing demand of making scientific evidence accessible and comprehensible for society and particular stakeholders, in addition to other formats, the Internet has become a rich source of informal transfer activities (see Seidel et al., 2017; Slavin, 2020, for examples). A prominent example of such transfer activities are so-called clearing houses (e.g., <https://ies.ed.gov/ncee/wwc/>; <https://www.clearinghouse.edu.tum.de/>). Clearing houses aim to present current research findings (mostly based on meta-analyses) in the format of compact summaries to ensure the comprehensibility for non-statisticians. Additionally, per summary, the quality of the obtained scientific evidence is benchmarked to provide practitioners with guidelines

regarding the trustworthiness of the obtained findings. Clearing houses were adopted from medical research and have been implemented in applied educational research contexts. To this end, clearing houses review the extant literature to find robust empirical studies that could be processed as compact summaries for educational practice. Thus, clearing houses aim to function as central linking pins between educational research and educational practice to ensure evidence-based transfer.

The What Works Clearinghouse (WWC) is a federal online portal hosted via the Institute of Education Sciences, which provides evidence-based information regarding the effectiveness of educational programs, policies, or interventions in a comprehensible manner (see Slavin, 2020). The empirical basis of the WWC is based on single-intervention studies, as well as (self-conducted) meta-analyses. To enhance the comprehensibility of the findings, graphical representations are integrated to highlight the context and the quality criteria of the presented studies (see Fig. 1).

Relatedly, the Technical University of Munich hosts the Clearing House Unterricht (Clearing House Teaching), which provides compact summaries of published meta-analyses in different formats such as written texts or podcasts to inform (mostly German) teacher educators on the effectiveness of instructional strategies in secondary STEM education (see Seidel et al., 2017). Additionally, the clearing house offers additional web-based trainings (Clearinghouse Unterricht academy) to train teacher educators regarding the basic methods of empirical educational research. In these clearing houses, based on medical research, meta-analyses are regularly taken as primary source and gold standard of empirical evidence in education (Renkl, 2022; Seidel et al., 2017). Together, such clearing houses provide accessible and comprehensible information for practitioners regarding the effectiveness of interventions. At the same time, as discussed previously, there may also emerge potential difficulties regarding the implementation of such evidence-based practices, as the scientific evidence is not localized in applicable interventions to demonstrate and exemplify the underlying principles, as primarily aggregated meta-analyses are taken, which makes it difficult to implement and adopt evidence-based practices. At the same time, as such clearing houses often follow a cascade-transfer strategy, which presuppose a multi-phase research and transfer process from conducting primary studies and synthesizing and aggregating evidence in meta-analyses to translating the obtained evidence for educational practice, there may be a natural bottleneck to provide practitioners with in-time and state-of-the-art empirical evidence.

On the other end of the continuum, due to the open educational resource (OER) movement (see Mullens & Hoffman, 2023, for an overview), several federal platforms exist that provide current and freely available instructional materials that could function as role models and examples for teacher educators, teachers and students (e.g., Academic Materials; Florida Postsecondary Academic Library Network, [sesam@lmz](mailto:sesam@lmz)). However, although OER may provide a sensible infrastructure for the dissemination of evidence-based practices, the quality as well as the consideration of empirical evidence within the published learning materials may vary greatly among the learning materials (Mullens & Hoffman, 2023). One reason for this observation may be that OERs are produced by practitioners for practitioners with little to no measures of quality assurance. Thus, OER materials necessarily do not integrate empirical research within the loop of developing OER to test their effectiveness.

## Good Behavior Game

Systematic Review Protocol for Social, Emotional, and Behavioral Interventions

Good Behavior Game had positive effects on student behavior and teacher practice and potentially positive effects on student writing conventions and writing productivity compared with business-as-usual programs. Good Behavior Game had uncertain effects on literacy and mathematics achievement, intrapersonal competencies, and school climate.

Good Behavior Game is a classroom management strategy that aims to improve social skills, minimize disruptive behaviors, and create a positive learning environment. Teachers place students into teams and reward them for demonstrating appropriate behaviors and following classroom rules.

**Findings** 16 STUDIES THAT MET STANDARDS OUT OF 31 ELIGIBLE STUDIES REVIEWED

Outcome domain	Effectiveness Rating	Grades	Evidence Tier
Alphabets	The WWC could not calculate effect sizes for any of the findings that meet standards in this domain.	K	NO TIER ASSIGNED
General Literacy Achievement	0	1-5	NO TIER ASSIGNED
General Mathematics Achievement	0	1-2	NO TIER ASSIGNED
Intrapersonal Competencies	0	K-5	NO TIER ASSIGNED
School Climate	0	2-5	NO TIER ASSIGNED
Student Behavior	1	K-11	TIER 1 STRONG
Teacher Practice	3	K-4	TIER 3 PROMISING
Writing Conventions	3	1-2	TIER 3 PROMISING
Writing productivity	3	1-2	TIER 3 PROMISING

Last Updated: May 2023

### Related Resources



**WWC Summary of Evidence for Good Behavior Game**



**Systematic Review Protocol for Social, Emotional, and Behavioral Interventions Evidence Review Protocol**



**Find other Children and Youth with Disabilities, Social Emotional Learning and Behavior, Teachers and School Leaders related interventions**



**See related studies**

-  [Download the Brief 211 KB](#)
-  [Download the Full Report 516 KB](#)
-  [Download Intervention Report Appendices 276 KB](#)

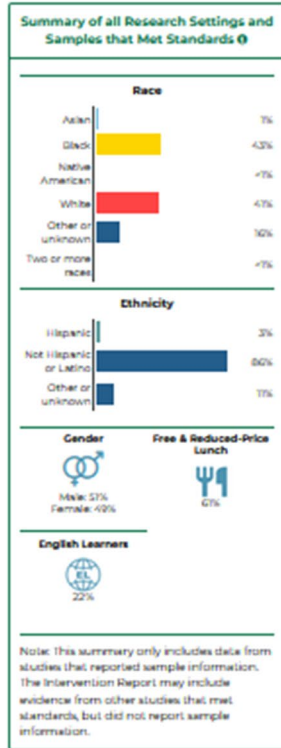


Fig. 1 Screenshot of a protocol of the What Works Clearinghouse (<https://ies.ed.gov/ncee/wwc/InterventionReport/728>)



## The LoGeT Model

The previous considerations suggested that integrative and non-isolated approaches are needed to simultaneously generate localizable, generalizable, and transferable evidence in order to meaningfully inform educational practice regarding the effectiveness of instructional interventions. To this end, we propose the localize, generalize, and transfer (LoGeT) model to localize, generalize, and transfer scientific evidence for educational practice. In the LoGeT model, we synthesized and systematically integrated strategies of co-design, ManyClasses, and transfer approaches. Figure 2 visualizes the different stages of the LoGeT model.

### Localization Stage: Co-design of Instructional Interventions

In the localization stage, the principle-oriented and co-constructive design of instructional interventions is emphasized. As a core-principle of co-design, interdisciplinary design teams (e.g., teachers and researchers) are engaged in a participatory design process to design an instructional intervention. In contrast to common co-design approaches, several subject-matter teachers and instructors are invited simultaneously to include a broad and inclusive set of different subjects and contexts to localize abstract design-principles in authentic subject-specific teaching experiences. To enhance the grounding processes (Clark & Brennan, 1991) among the diverse backgrounds of the design team, a context inquiry is accomplished for instance by guided focus groups in design thinking workshops. These workshops should help trace the different design and implementation conditions and prepare a joint design framework for the instructional interventions. The implementation of such a joint design framework is needed to warrant that the to-be-designed interventions are

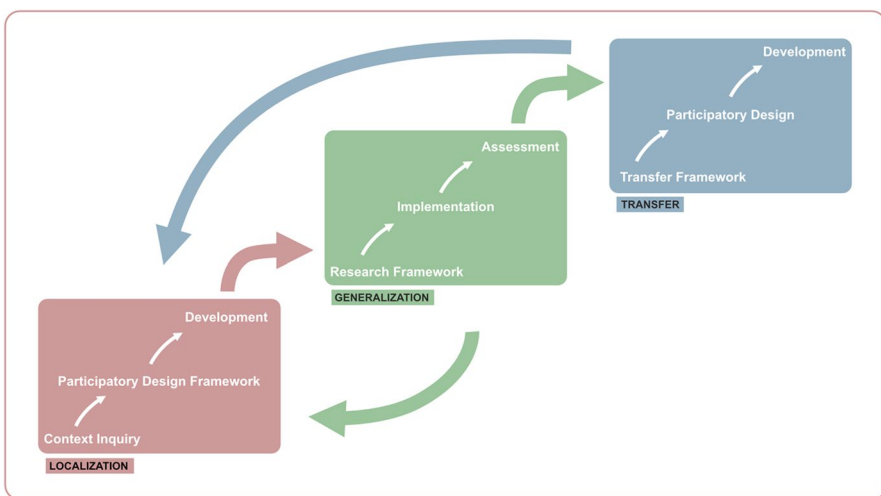


Fig. 2 The three stages of the LoGeT model

comparable with regard to the underlying design principles and likewise enhance the fidelity of the proposed implementations (Carroll et al., 2007; see also Maciver et al., 2021; Palmer et al., 2015). Based on the design framework, subject-specific interventions are implemented. Reciprocal formative feedback workshops should further help to assure that the instructional implementation is implemented according to the to-be-tested principles, as well as that the context of implementation is adequately considered in the interventions.

### **Generalization Stage: Assessing the Effectiveness of Instructional Interventions**

In the generalization stage, the different instructional interventions are implemented in diverse contexts, covering diverse subjects, cohorts, and contexts. Following a ManyClasses approach, the developed interventions are evaluated as implementations of a distinct design principle. Depending on the achievable sample size and the assumed effects, different implementation conditions and different experimental designs (e.g., within-participants experiments, between-participants experiments, cluster-randomized field trial) can be implemented for assessing the effect of an instructional intervention. Depending on the available sample size, also a broader variety of research designs (e.g., correlational, mixed methods) may be considered during the generalization stage. A careful and balanced design of the test instruments is warranted in order to not compare apples with pears. As for the localization stage, iterative design workshops with the different stakeholders should help design assessments that measure the intended construct in the particular context and at the same time enhance the comparability of the different instruments. Additional statistical measures such as standardization could further contribute to the comparability of the instruments (Fyfe et al., 2021). To aggregate the findings, several approaches exist—such as mixed effect models, cluster-robust inference, or hierarchical Bayesian models—that explicitly take the nested data structure (studies are nested within different classrooms) into account and allow to explicitly model potential moderation effects of the instructional contexts (see Fyfe et al., 2021; Sana & Yan, 2022; Sibley et al., 2023a, for examples).

### **Transfer Stage: Transferring the Obtained Evidence**

In the transfer stage, the obtained evidence as well as the designed interventions are processed and published so that they can be adopted by other stakeholders. Transfer activities should be considered early in the previous stages. Based on the targeted audience, different contents (e.g., the compact summaries, descriptions of the intervention, the utilized learning materials as OER) and formats for publication can be considered (e.g., print, multi-media, social media). These media formats can serve as the backbone for further development. Websites such as <https://senseaboutscience.org/> provide potential strategies to transfer scientific evidence. That said, as in the research process, co-constructive, formative design and testing could contribute to the acceptance and adoption of the different transfer products. Besides, the presentation of these products in (national) application-related specialist outlets

(e.g., American Educator, <https://www.learningscientists.org>) as well as talks and round-tables in practice-related conferences could further contribute to transfer the obtained evidence into practice.

### Three Empirical Examples of Applying the LoGeT Model

In the following, we present three empirical examples in which several researchers adopted the LoGeT-model to provide localized, generalizable, and transferable evidence. The three examples constitute different application settings (teacher education, schooling), participants (pre-service teachers, school students) and subjects. Additionally, the examples differed regarding the time scale and the type of intervention (minimal invasive manipulation versus entire intervention).

#### An Example in Teacher Education

The first example was a project in the context of teacher education (Lachner et al., 2021) to support pre-service teachers' technology integration during teaching. One prevailing challenge in the field of teaching with educational technology was the limited availability of well-designed interventions that target teachers' acquisition of technological pedagogical content knowledge. That said, empirical evidence regarding the effectiveness of such interventions was scarce, as most previous studies only had relied on self-reports. To address this desiderate, a theoretical approach grounded in the SQD (Strategies for Quality Development; Tondeur et al., 2012) model was adopted in this project. The following procedure was realized.

**Localization Stage** To localize the generic design principles of the SQD-model, Lachner et al. (2021) employed a comprehensive approach that was centered around the formation of five small-scale design teams. In an initial phase, the design teams comprised local experts from the participating subject-matter didactics (biology, English as a foreign language, German literature, mathematics, philosophy) and two educational technology researchers. Due to successful project funding, in a subsequent phase, the core team was extended with three additional project staff members. The project staff members had a subject-matter teaching background in one or more subjects and considerable experience in adopting educational technology. The staff members were mainly responsible for the design and development of the subject-specific interventions together with the educational technology researchers and the subject-matter didactic experts. The context inquiry was realized within regular project meetings. During these meetings, it became salient that the interventions should be predominantly grounded in subject-specific teaching practices to foster TPACK and be easily implementable in the current courses of the participating subject-matter didactics. Thus, a timeframe of 3 weeks was chosen for the duration of the intervention. In the participatory design framework, the educational researchers suggested the SQD model (Tondeur et al., 2012) as generic design model that was applied across the subject-specific realizations to guarantee their comparability. The

SQD-model comprises different instructional activities (i.e., collaboration, authentic experiences, feedback, role models, reflection activities, instructional design) that should contribute to pre-service teachers' development of TPACK. Given that the SQD model is relatively generic, the model was adapted for each subject by the design teams and enriched by the corresponding theories from the subject-matter didactics (e.g., Ladel, 2009; Surkamp & Viebrock, 2018; Tiedemann, 2019). To ensure that the core framework was comparable across subjects, the design teams met in regular design thinking workshops to ensure the implementation fidelity and comparability of the interventions. Thus, these teams embraced a generic approach that was consistently applied across all design groups, emphasizing collaboration and co-constructive input. The instructional activities were orchestrated in three larger sessions. During the first session, an online learning module was implemented to introduce the students to theories and principles of subject-specific technology integration principles. To model effective technology integration, the students were provided with video-modeling examples that represented good practices for integrating technology into their respective fields (see <https://www.youtube.com/@tubingencenterfordigital8488> for examples). In the second session, the students were engaged in a collaborative design task in which they realized an instructional design of a subject-specific lesson and realized the respective teaching materials (see Backfisch et al., 2024, for the empirical findings). The students additionally received formative feedback from the instructors. In the third session, the students tested their instructional design within micro-teachings to make authentic experiences in an approximation of teaching practices. In these micro-teachings, other students mimicked school students with a pre-given script. The micro-teachings were videotaped. As an additional homework assignment, the students were engaged in a peer-feedback task in which they provided feedback to the other students regarding the quality of their micro-teaching based on pre-validated prompts.

**Generalization Stage** To be able to generalize the findings, a joint research framework was realized that comprised both a cohesive design of test instruments and a comparable experimental procedure across subjects. The realization of a joint research framework allowed to detect and repair potential inconsistencies between the different localized realizations of the localized design frameworks to increase treatment fidelity. The design teams decided to adopt a cluster-randomized design in which classes were randomly assigned to the intervention or a control condition, as a true experimental design was not feasible in such a classroom setting. As for the design of the intervention, the design teams closely worked together to realize test instruments that were subject-specific, but at the same time roughly comparable across the different subjects. For this reason, the design team used vignette-based, open-ended questions to measure TPACK, asking to integrate technology for subject-specific teaching (e.g., prior knowledge activation, testing). The number of items was fixed across subjects.

The obtained data was analyzed via multi-level analyses (varying slope model) to account for variations among the different subject matter courses. This approach helped discern the nuances of how the interventions affected TPACK and self-efficacy across various teaching contexts. The study yielded robust findings, suggesting

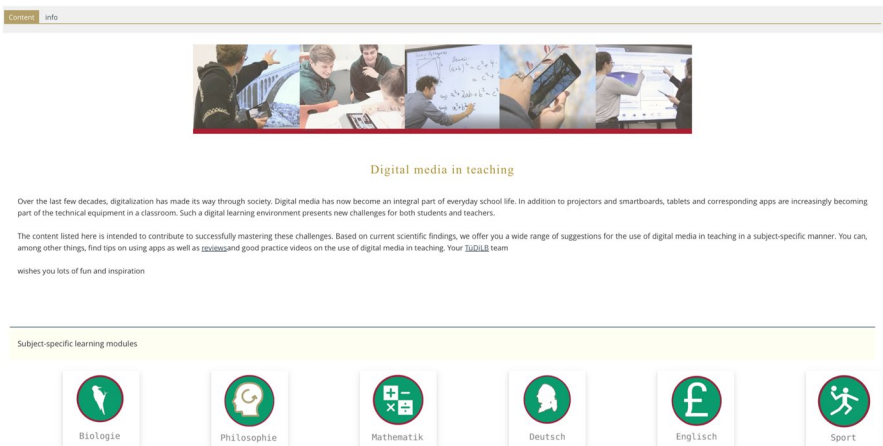
that the interventions indeed contributed to the enhancement of teachers' TPACK and technology-related self-efficacy. This positive impact could be explained by the implementation of SQD-features within the interventions.

**Transfer Stage** To transfer the contents and the obtained findings, the design team followed different dissemination strategies that were planned in regular meetings and framed in a joint transfer framework. In those meetings it was decided that the materials should both be re-usable and adaptable for different users. As such, the design teams realized the different materials as open educational resources (OER) and disseminated the materials in different repositories (see [https://lms-public.uni-tuebingen.de/ilias3/goto\\_pr01\\_cat\\_6596.html](https://lms-public.uni-tuebingen.de/ilias3/goto_pr01_cat_6596.html) for an overview and Fig. 3). To this end, the design teams provided a joint framework for publishing the different materials to establish cohesion. Additionally, the findings were presented at teacher education conferences and published in teacher education journals and book chapters (e.g., Franke et al., 2020).

Moreover, the intervention and the underlying principles served as the foundation for a university-wide curriculum, spanning 25 subjects and benefiting around 4000 pre-service teachers. Additionally, this intervention served as a prototype, inspiring follow-up projects that followed an adopted design and research approach based on the obtained findings. In summary, the research findings have not only improved teacher education but have also sparked collaborative initiatives and innovation in teacher development on a broader scale.

## An Example for Adaptive Teaching at Comprehensive Schools

The second example constitutes a project within the schooling context (Sibley et al., 2023a). In response to the increasing challenge of student heterogeneity, a 4-year



**Fig. 3** An example of the transfer outlet of the TPACK project

instructional development project was initiated to explore adaptive teaching with educational technology as a potential solution. The specific aim was to leverage available technologies to enhance the implementation of adaptive teaching. Over 4 years, researchers, teachers, and stakeholders of the school administration engaged in a collaborative effort to develop and implement adaptive teaching units (duration 3–4 weeks) among central subjects of secondary education.

**Localization Stage** In a first step, the research team met in regular meetings with the school principal as well as the local school administration to inquire the particular context for adaptive teaching with technology. It was decided that the project should focus on upper secondary education at comprehensive schools, as these classes were recently equipped with educational technology and infrastructure by the community and thus could serve as a blueprint for realizing adaptive teaching with technology. After receiving project funding from two private foundations, it was possible to delegate seven teachers that formed the core design teams together with three educational researchers. In the participatory design framework, the design teams developed a common model of adaptive teaching, which served as the basis for subject-specific teaching units. The adaptive teaching model comprised the iterative phases of formative assessment, macro-adaptations, and micro-adaptations (Corno, 2008). The adaptive teaching model was adopted in 12 teaching units that covered 3 to 4 weeks across different subjects (mathematics, physics, chemistry, German literature, English as a foreign language, Spanish as a foreign language, ethics). Again, the delegated teachers met in regular design meetings together with the educational researchers to ensure the comparability of the different realizations of adaptive teaching with technology.

**Generalization Stage** The initial research framework was designed as a mixed-methods study, as the study was mainly realized at one school. A second school joined the project after 2 years. Due to these project restrictions, the main aim of the study was to test the generic model's feasibility. The mixed-methods approach included a quantitative study involving 183 students to measure learning gains (pre-posttest scores) and identify potential moderating factors that could explain differences in learning gains. Therefore, the design teams met in iterative sessions to jointly design the knowledge tests. The design of the knowledge tests also helped the different design teams to adjust their localizations of the design frameworks. Qualitative data were collected via interviews with three of the participating teachers, who obtained high, medium, versus low knowledge gains of the students in their teaching units, focusing on the implementation conditions of the adaptive teaching units. To test the feasibility of the adaptive teaching framework, cluster-robust estimations of fixed effect models were used to account for the correlated error terms within a cluster (students within teaching units). Additionally, moderation analyses were conducted to investigate potential boundary conditions. Overall, the quantitative findings showed significant learning gains across the teaching units regardless of the subject domain of the teaching units. Additionally, larger increases were obtained for students with low prior knowledge and when the implementation fidelity was high. The qualitative data emphasized the importance of formative assessments, micro-adaptations,

and a parsimonious use of technology. In the next step, a 2-year cluster-randomized experimental field study is currently underway to examine the impact of adaptive teaching, enhanced by educational technology, on student outcomes.

**Transfer Stage** To transfer the obtained findings and the resulting materials, the learning materials and teaching units have been published as OER (see Fig. 4). For this purpose, one staff member was responsible to develop a transfer framework together with the participating teachers ([https://lms-public.uni-tuebingen.de/ilias3/goto.php?target=cat\\_6858](https://lms-public.uni-tuebingen.de/ilias3/goto.php?target=cat_6858)). Additionally, continuing workshops as well as articles in educational practice journals (Sibley et al., 2023a) further increased the impact of the project. Most recently, the design idea has been scaled in a multi-site project in which teachers are explicitly trained to realize adaptive teaching with technology and built up a thematic professional network. Based on the insights from the transfer stage, we adjusted both the localization stage and the generalization stage of the multi-site project. To this end, in addition to upper secondary education, also lower secondary education school tracks have been part of the network and the corresponding design team. Additionally, to gain more robust insights into the effectiveness of adaptive teaching with technology, we have been realizing a cluster-randomized field trial with control classes that did not attain the adaptive teaching classes. These transfer measures should additionally increase the transferability of the previous findings.

### An Example for Learning by Non-interactive Teaching

The last example by Sibley, Russ et al. (2023b) also targeted a schooling context but was conducted to investigate the potential of a minimal-invasive instructional strategy,

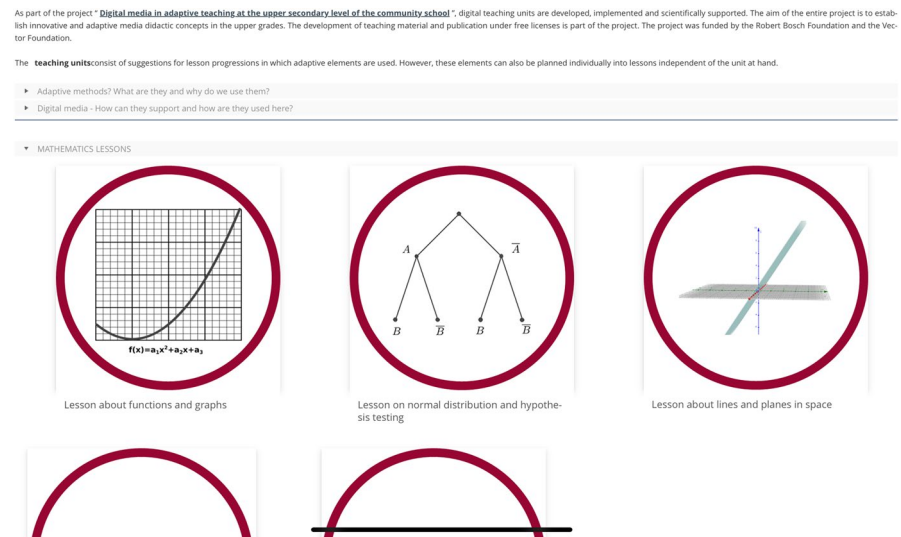


Fig. 4 An example for the transfer outlet in the DiA:GO project

that is, the effectiveness of non-interactive teaching (Lachner et al., 2022). Non-interactive teaching is a generative activity in which students generate an explanation to a non-present or even fictitious student, with the aim to enhance one's own understanding of the previously learned contents (Hoogerheide et al., 2019; Lachner et al., 2022). The effectiveness of learning by non-interactive teaching was predominantly investigated in laboratory contexts and only seldom applied in subject-specific school settings (see Hoogerheide et al., 2019; Jacob et al., 2022, for exceptions). Against this background, the aim of the project was to investigate the effectiveness of implementations of non-interactive teaching in diverse school-settings (Sibley et al., 2023a).

**Localization Stage** Contrary to the aforementioned projects that relied on external funding, this project was integrated into a course on educational technology in an educational master's program at a university in Southwestern Germany. The master's program welcomes both in-service teachers and graduates in educational research, offering them the opportunity to learn from each other and enhance their scientific understanding within the context of schooling. The main goal of the project was to learn how research on educational technology can be conducted in authentic schooling contexts. Ten design teams comprising both in-service teachers and master's students alike collaboratively developed 20 teaching units across a set of diverse subjects (e.g., physics, history, English as a foreign language, economics) and school types (e.g., primary, secondary, and high school, vocational education). In addition, the design teams randomly implemented a non-interactive teaching task (versus control), which was provided to the students at the end of the teaching unit. The teaching units were held by the corresponding teacher of the particular design team. To ensure that the non-interactive teaching tasks were comparable across teaching units and at the same time were adapted to the particular teaching contexts (e.g., primary vs. secondary education, availability of infrastructure), the teams discussed their realizations of the teaching units during the weekly courses.

**Generalization Stage** As the sample size was relatively restricted, in the research framework the design teams decided to realize a within-participants design that likely required a lower number of participants to achieve high levels of test power. Again, the knowledge tests as well as the technical design of the experiment were discussed during the weekly course sessions. Five open-ended questions were designed per teaching unit to measure differences in the learning outcomes. To this end, the students ( $N=191$ ) received different sequences of non-interactive teaching and restudy activities across two comparably difficult lessons. Overall, it could be demonstrated that non-interactive teaching was not generally more effective than restudy (see also Lachner et al., 2021, for meta-analytic evidence). However, additional moderation analyses revealed that non-interactive teaching was more effective in the humanities as well as upper secondary education and when it was graded, as external incentive. The findings highlighted the role of the context in which non-interactive teaching was embedded.



**Transfer Stage** As non-interactive teaching is rather a minimal-invasive learning activity, the transfer strategy was different compared to the previous two project examples. Thus, the findings of the current study in combination with previous empirical evidence of non-interactive teaching (see Lachner et al., 2022, for an overview) were incorporated in online professional development programs of the federal institute of teaching training, which comprised the integration of educational technology in the schooling context in general. Additionally, short explanation videos were developed as practice guides to introduce German teachers and educational practitioners to the evidence-based use of non-interactive teaching that are regularly used in pre-service teacher education ([https://youtu.be/Ohp9x\\_tUar4?si=MStavvqM\\_Aa9oIG2](https://youtu.be/Ohp9x_tUar4?si=MStavvqM_Aa9oIG2); <https://youtu.be/WZYh1aSxozE?si=ZerMa2PlfQiJZlA8>; see also Agarwal, 2024, for related strategies in the context of retrieval practice).

## Conclusions

Educational policymakers are commonly advised to base their educational decisions on empirical evidence. To this end, educational researchers are warranted to provide implications based on the obtained (empirical) evidence for educational practice (Slavin, 2020). Given that empirical evidence is often dependent on the particular context, research methodologies that explicitly take context variables into account are needed to provide evidence, which is located in authentic educational practices and at the same time generalizable across different instructional situations. At the same time, the obtained evidence needs to be processed to be transferable to educational practice. Although different methodological approaches exist, there is no integrative approach to address the previously mentioned problems. In this paper, we have taken one step further toward filling this methodological gap by providing an integrative approach, the LoGeT model. The LoGeT model is a working model that explicitly synthesized co-design, ManyClasses approaches, and transfer strategies to provide localized, generalizable, and transferable knowledge for educational practice. The model is not sequential in nature, but rather should be interpreted as reciprocal.

We see two main strengths of the LoGeT model. First, the LoGeT model bridges disparate approaches that have traditionally been treated in disciplinary isolation: Whereas co-design approaches mainly emerged within research on human–computer-interaction and the learning sciences, ManyClasses approaches have considerably been adopted in applied cognitive psychology research. Meanwhile, transfer activities, such as clearing houses, however, have been mainly adopted in applied educational research settings. By systematically intertwining these “non-mutually exclusive” approaches, the LoGeT model may bridge the boundaries of these approaches and facilitate cross-fertilization, allowing for valuable insights to be gained from each perspective. The integrative character of the LoGeT model may thus also enhance the collaboration and knowledge integration among different research domains during inter- and transdisciplinary work (e.g., educational researchers, methodologists, educational practitioners).

As a second strength, the LoGeT model distinguishes itself through its reciprocal nature, which serves as an additional safeguard and informs the revision activities of preceding phases. This reciprocal approach within the LoGeT model allows combining distinct design and empirical decision-making phases and thus contributes to the generation of robust evidence that is both localizable and generalizable across contexts. This integrative procedure may allow to derive more flexible and interactive decisions, than the single approaches, or simple hierarchical entanglements.

Additionally, we presented three different empirical examples of different granularities, which illustrate the LoGeT procedure. To this end, we hope to provide a stimulus for a research agenda that explicitly takes into account the specific needs of educational practice. Despite the potential benefits, also some challenges need to be addressed. One challenge regards potential biases. Although different instructional contexts have been included in the LoGeT model, it is naturally difficult to sample classes a priori that reflect all potential contextual variables. Thus, similarly to findings from meta-analyses, the obtained contextual findings cannot be interpreted to be causal, but rather inform future experimental studies to directly test contextual effects (see also Renkl, 2013). Such biases may increase, as the empirical evaluation for generalization depends on the willingness of (motivated) teachers and instructors to participate in the co-design process. Therefore, in the worst case, the findings may be confounded rather by teacher motivation and the fidelity of the implementation than by the psychological principle under investigation (Fyfe et al., 2021). To circumvent such biases within-designs that investigate such effects within one classroom and measures of implementation fidelity (Carroll et al., 2007) are needed to infer potential causal effects of the educational interventions.

In addition to methodological threads, there are also practical challenges. LoGeT studies require considerable research efforts and infrastructure. For instance, these studies need a vivid network of participating teachers and instructors that have the additional time to actively work in the projects. Two of the three examples demonstrated that additional funding is required to enable teachers to actively participate in the research process, which is a demanding endeavor in the current lack of skilled labor. That said, multiple variants of the experimental design and test measures have to be validated and developed for the specific classrooms, as well as compared across the different classroom settings. Last, but not least, the studies have to be processed for practitioners via outlets for scientific outreach and public engagement, to reach a broader community that goes beyond the participating teachers and instructors. Such activities require professionals to effectively communicate the obtained evidence (Slavin, 2020). That said, as the LoGeT procedure is relatively demanding and requires considerable research efforts, additional incentives are required to go beyond piecemeal research approaches.

Despite these challenges, we think that the LoGeT model provides an alternative lens to investigate potential effects of instructional interventions and principles in the wild that allows to localize and generalize empirical evidence in education. Thus, we hope that this approach provides a starting point for future research and editorial agendas, as well as methodological advancements to adopt distinct strategies for generating and transferring scientific evidence.

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agarwal, P. K. (2024). Personal reflections on science communication and sharing retrieval practice research with teachers. *Educational Psychology Review*, 36(1). <https://doi.org/10.1007/s10648-023-09839-w>
- Backfisch, I., Franke, U., Ohla, K., Scholtz, N., & Lachner, A. (2024). *Collaborative design practices in pre-service teacher education for technological content knowledge (TPACK): Group composition matters*. Advance online publication. <https://doi.org/10.1007/s42010-023-00192-z>
- Bødker, S. (1996). Creating conditions for participation: Conflicts and resources in systems development. *Human-computer interaction*, 11(3), 215–236
- Bokosmaty, S., Sweller, J., & Kalyuga, S. (2015). Learning geometry problem solving by studying worked examples: Effects of learner guidance and expertise. *American Educational Research Journal*, 52(2), 307–333. <https://doi.org/10.3102/00028312145494>
- Bråten, I., & Ferguson, L. E. (2015). Beliefs about sources of knowledge predict motivation for learning in teacher education. *Teaching and Teacher Education*, 50, 13–23. <https://doi.org/10.1016/j.tate.2015.04.003>
- Brod, G. (2021). Generative learning: Which strategies for what age? *Educational Psychology Review*, 33(4), 1295–1318. <https://doi.org/10.1007/s10648-020-09571-9>
- Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(4), 3–54. 10.1007.
- Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029–1052. <https://doi.org/10.1037/bul0000209>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 1–9. <https://doi.org/10.1186/1748-5908-2-40>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. <https://doi.org/10.1037/10096-006>
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173. <https://doi.org/10.1080/00461520802178466>
- Farley-Ripple, E., Van Horne, S., Tilley, K., Shewchuk, S., May, H., Micklos, D. A., & Blackman, H. (2022). Survey of evidence in education for schools (SEE-S) descriptive report. *Center for Research Use in Education*. <https://files.eric.ed.gov/fulltext/ED628009.pdf>. Accessed 10 Sept 2023
- Franke, U., Fabian, A., Preiß, J., & Lachner, A. (2020). TPACK 4.0 – interdisziplinäre, praxisorientierte und forschungs-basierte Förderung von fachspezifischem methodendidaktischem Wissen bei angehenden Lehrpersonen. In K. Kaspar, M. Becker-Mrotzek, S. Hoffmues, J. König, D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (pp. 182–187). Münster: Waxmann
- Fyfe, E.R., Leeuw, J.R., de Carvalho, P.F., Goldstone, R. L., Sherman, J., Admiraal, D., Alford, L. K., Bonner, A., Brassil, C. E., Brooks, C. A., Carbonetto, T., Chang, S. H., Cruz, L., Czayoniewicz-Klippel, M., Daniel, F., Driessen, M., Habashy, N., Hanson-Bradley, C. L., Hirt, E. R.,..., Motz, B. A. (2021). ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science*, 4(3). <https://doi.org/10.1177/25152459211027575>

- Greisel, M., Wekerle, C., Wilkes, T., Stark, R., & Kollar, I. (2023). Pre-service teachers' evidence-informed reasoning: Do attitudes, subjective norms, and self-efficacy facilitate the use of scientific theories to analyze teaching problems?. *Psychology Learning & Teaching*, 22(1). <https://doi.org/10.1177/14757257221113942>
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics*, 6(2). <https://doi.org/10.18608/jla.2019.6.2.3>
- Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction*, 64, 101226. <https://doi.org/10.1016/j.learninstruc.2019.101226>
- Iniesto, F., Charitonos, K., & Littlejohn, A. (2022). A review of research with co-design methods in health education. *Open Education Studies*, 4(1), 273–295. <https://doi.org/10.1515/edu-2022-0017>
- Jacob, L., Lachner, A., & Scheiter, K. (2022). Do school students' academic self-concept and prior knowledge constrain the effectiveness of generating technology-mediated explanations? *Computers & Education*, 182, 104469. <https://doi.org/10.1016/j.compedu.2022.104469>
- Kaplan, A., Cromley, J., Perez, T., Dai, T., Mara, K., & Balsai, M. (2020). The role of context in educational RCT findings: A call to redefine “evidence-based practice.” *Educational Researcher*, 49(4), 285–288. <https://doi.org/10.3102/0013189X2092186>
- Lachner, A., Fabian, A., Franke, U., Preiß, J., Jacob, L., Führer, C., Küchler, U., Paravicini, W., Randler, T., & Thomas, P. (2021). Fostering pre-service teachers' technological pedagogical content knowledge (TPACK): A quasi-experimental field study. *Computers & Education*, 174, 104304. <https://doi.org/10.1016/j.compedu.2021.104304>
- Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2022). Learning-by-teaching without audience presence or interaction: When and why does it work? *Educational Psychology Review*, 34, 575–607. <https://doi.org/10.1007/s10648-021-09643-4>
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, 44(3), 197–203. <https://doi.org/10.1007/s11251-016-9376-y>
- Lachner, A., Weinhuber, M., & Nückles, M. (2019). To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of principle-oriented explanations. *Contemporary Educational Psychology*, 58, 175–185. <https://doi.org/10.1016/j.cedpsych.2019.03.008>
- Ladel, S. (2009). *Multiple externe Repräsentationen (MERs) und deren Verknüpfung durch Computereinsatz: Zur Bedeutung für das Mathematiklernen im Anfangsunterricht*. [Multiple external representations (MERs) and their linking through computer use: On the importance for mathematics learning in early education]. Dr. Kovač
- Leinhardt, G., & Putnam, R. R. (1986). Profile of expertise in elementary school mathematics teaching. *The Arithmetic Teacher*, 34(4), 28–29. <https://doi.org/10.5951/AT.34.4.0028>
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago press.
- Maciver, D., Hunter, C., Johnston, L., & Forsyth, K. (2021). Using stakeholder involvement, expert knowledge and naturalistic implementation to co-design a complex intervention to support children's inclusion and participation in schools: The CIRCLE framework. *Children*, 8(3), 217.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487–498. <https://doi.org/10.1037/a0039400>
- Mayer, R. E. (2023). How to assess whether an instructional intervention has an effect on learning. *Educational Psychology Review*, 35, 64. <https://doi.org/10.1007/s10648-023-09783-9>
- Mullens, A. M., & Hoffman, B. (2023). The affordability solution: A systematic review of open educational resources. *Educational Psychology Review*, 35, 72. <https://doi.org/10.1007/s10648-023-09793-7>
- Palmer, V. J., Chondros, P., Piper, D., Callander, R., Weavell, W., Godbee, K.,..., Gunn, J. (2015). The CORE study protocol: a stepped wedge cluster randomised controlled trial to test a co-design technique to optimise psychosocial recovery outcomes for people affected by mental illness in the community mental health setting. *BMJ open*, 5(3), e006688
- Renkl, A. (2013). Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educational Psychology Review*, 25, 317–324. <https://doi.org/10.1007/s10648-013-9236-0>
- Renkl, A. (2022). Meta-analyses as a privileged information source for informing teachers' practice? *Zeitschrift Für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000345>

- Riley, R. D., Tierney, J. F., & Stewart, L. A. (2021). *Individual participant data meta-analysis: a handbook for healthcare research*. John Wiley & Sons
- Robinson, D. H., Levin, J. R., Schraw, G., Patall, E. A., & Hunt, E. B. (2013). On going (way) beyond one's data: A proposal to restrict recommendations for practice in primary educational research journals. *Educational Psychology Review*, 25, 291–302. <https://doi.org/10.1007/s10648-013-9223-5>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roschelle, J., Penuel, W., & Shechtman, N. (2006). Co-design of innovations with teachers: Definition and dynamics. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *ICLS '06: Proceedings of the 7th international conference on learning sciences* (pp. 606–612). International Society of the Learning Sciences
- Sana, F., & Yan, V. X. (2022). Interleaving retrieval practice promotes science learning. *Psychological Science*, 33(5), 782–788. <https://doi.org/10.1177/09567976211057507>
- Seidel, T., Mok, S. Y., Hetmanek, A., & Knogler, M. (2017). Metaanalysen zur unterrichtsforschung und ihr beitrage für die realisierung eines clearing house unterricht für die lehrerbildung. *Zeitschrift Für Bildungsforschung*, 7(3), 311–325. <https://doi.org/10.1007/s35834-017-0191-6>
- Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2016). Organizing for teacher agency in curricular co-design. *Journal of the Learning Sciences*, 25(4), 531–564. <https://doi.org/10.1080/10508406.2016.1207541>
- Sibley, L., Lachner, A., Plicht, C., Fabian, A., Backfisch, I., Scheiter, K., & Bohl, T. (2023a). *Feasibility of adaptive teaching with technology: Which implementation conditions matter?* [Manuscript submitted for publication]. University of Tübingen.
- Sibley, L., Russ, H., & Lachner, A. (2023b). Learning by Teaching in the Wild: Domain, Grades, and School Track Matter. <https://doi.org/10.31234/osf.io/d293m>
- Slattery, P., Saeri, A. K., & Bragge, P. (2020). Research co-design in health: A rapid overview of reviews. *Health Research Policy and Systems*, 18(1), 1–13. <https://doi.org/10.1186/s12961-020-0528-9>
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21. <https://doi.org/10.3102/0013189X031007015>
- Slavin, R. E. (2020). How evidence-based reform will transform research and practice in education. *Educational Psychologist*, 55(1), 21–31. <https://doi.org/10.1080/00461520.2019.1611432>
- Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review*, 35(4), 95. <https://doi.org/10.1007/s10648-023-09817-2>
- Surkamp, C. & Viebrock, B. (2018). *Teaching English as a foreign language. An introduction*. Metzler
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837–848. <https://doi.org/10.1002/acp.1598>
- Tiedemann, M. (2019). Der problemorientierte Ansatz [The problem-orientated approach]. In M. Peters & J. Peters (Eds.), *Moderne Philosophiedidaktik [Modern philosophy didactics]* (pp. 213–230). Felix Meiner Verlag.
- Tondeur, J., Van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134–144. <https://doi.org/10.1016/j.compedu.2011.10.009>
- Turner, J. C., & Nolen, S. B. (2015). Introduction: The relevance of the situative perspective in educational psychology. *Educational Psychologist*, 50(3), 167–172. <https://doi.org/10.1080/00461520.2015.1075404>
- van Gog, T., Paas, F., & Van Merriënboer, J. J. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, 18(3), 211–222. <https://doi.org/10.1016/j.learninstruc.2007.03.003>
- Veroniki, A. A., Seitidis, G., Tsigvoulis, G., Katsanos, A. H., & Mavridis, D. (2023). An introduction to individual participant data meta-analysis. *Neurology*, 100(23), 1102–1110. <https://doi.org/10.1212/WNL.000000000207078>
- Weinhuber, M., Lachner, A., Leuders, T., & Nückles, M. (2019). Mathematics is practice or argumentation: Mindset priming impacts principle- and procedure-orientation of teachers' explanations. *Journal of Experimental Psychology: Applied*, 25(4), 618–646. <https://doi.org/10.1037/xap0000227>
- Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 399–435. <https://doi.org/10.1037/bul0000309>

- Yannier, N., Crowley, K., Do, Y., Hudson, S. E., & Koedinger, K. R. (2022). Intelligent science exhibits: Transforming hands-on exhibits into mixed-reality learning experiences. *Journal of the Learning Sciences*, *31*(3), 335–368. <https://doi.org/10.1080/10508406.2022.2032071>
- Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education*, *2*, 399–420. <https://doi.org/10.1007/s40692-015-0036-z>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.