AQUA-d
Aufgabenqualität im digital gestützten Unterricht

# AQUA-d Conference 2025
# Advancing Digital Task Efficacy: International Perspectives on Technology- Enhanced Education

October 8th-9th, 2025 | Karlsruhe University of Education

The integration of digital technologies in education offers new opportunities to improve the quality of teaching, learning, and assessment. However, the effective design and use of digital tasks require careful consideration of educational principles, subject-specific teaching approaches, skills of educators, and proper technological solutions. This challenge is addressed by the interdisciplinary research group AQUA-d at the Karlsruhe University of Education. Over the last years, the graduate program has conducted empirical work dedicated to this field of research. This international conference provides a platform to discuss key aspects of digital task quality and its role in technology-enhanced education.

Conference themes include:

- **Educational Foundations**: Examining how learning theories guide the development of effective digital tasks.
- **Practical Applications**: Sharing subject-specific strategies for integrating digital tasks into classrooms.
- **Professional Development**: Identifying how educators can develop the competencies needed to design and use digital tasks effectively.
- **Technology Support**: Exploring the role of digital tools and data in supporting teaching and learning.

The program will include symposia, keynote talks and a poster session aimed at fostering collaboration and sharing insights.

**Link to conference: https://en.ph-ka.de/aqua-d-2025**

University of Education Pädagogische Hochschule karlsruhe

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

AQUA
Aufgabenqualität im digital gestützten Unterricht

# Conference Schedule

<table>
<tr><th></th><th>October 8, 2025</th><th>October 9, 2025</th></tr>
<tr><td>9:00<br>9:30</td><td colspan="2">Registration (3.110)</td><td>Keynote “Monitoring and Regulation of Effort and Learning. What does technology have to do with it?” by Prof. Dr. Anique De Bruin (1.013)</td></tr>
<tr><td>10:00<br><br>10:30<br>11:00</td><td colspan="2">Welcome & Keynote "The Role of Science Education in an Age of Misinformation and AI" by Prof. Jonathan Osborne (PhD) (1.013)</td><td>Coffee break (Forum)</td></tr>
<tr><td>11:30<br>12:00</td><td colspan="2">Lunch break (Forum)</td><td>S5: Strategies for Enhancing Performance with Perceptually Rich Displays (3.101)<br>S6: Profiling writing tasks in terms of literary and media didactics (german) (3.107)<br>S7: Formative Assessment - Noticing, Interpreting and Constructing Implications Based on Learning Progress Data (3.109)</td></tr>
<tr><td>12:30<br>13:00<br>13:30<br>14:00</td><td>S1: Digitally Supported Teaching and Learning Scenarios in the Field of Electricity (german) (3.101)</td><td>S2: Evidence-informed School Practice: Challenges, Capacities, and Catalysts for Transformation (3.107)</td><td>Lunch break (Forum)<br>S8: Metacognition and Desirable Difficulties: Enhancing Self-Regulated Learning and Judgment Accuracy in Digital Learning Environments (3.101)<br>S9: Digital learning in mathematics education (3.107)</td></tr>
<tr><td>14:30</td><td colspan="2">Coffee break (Forum)</td><td>Coffee break (Forum)</td></tr>
<tr><td>15:00<br>15:30<br>16:00<br>16:30</td><td>S3: Seeking Ways to Individualize and Tailor Motivation Interventions (3.101)</td><td>S4: Innovating Task Design in Economic Education: Data-Driven Insights in Task Quality and Practical Applications (3.107)</td><td>Closing Keynote "Digital Learning Tasks: Subject-Specific and Generic" by Prof. Dr. Timo Leuders & Goodbye (1.013)</td></tr>
<tr><td>17:00</td><td colspan="2">Coffee break (Forum)</td><td rowspan="3"></td></tr>
<tr><td>17:30<br>18:00<br>18:30<br>19:00</td><td colspan="2">Poster Session (3.101 and Forum)</td></tr>
<tr><td>19:30</td><td colspan="2">Conference Dinner (Badisch Brauhaus)</td></tr>
</table>

# Keynotes

## K1: The Role of Science Education in an Age of Misinformation and AI

Keynote speaker: Prof. Jonathan Osborne (PhD) *(Stanford University)*

October 8, 2025, 10:30-11:30
HS 1.013

**Abstract:**
The ability to "research, evaluate and use scientific information for decision making and action" is one of three core competencies to be measured in the 2025 PISA test. Likewise, "obtaining, evaluating and communicating information" is one of the 8 scientific practices of the innovative Next Generation Science Standards (NGSS).

Today, we are living in an age of misinformation. Yet many science standards were developed over decade ago before it became so pervasive. What competencies and knowledge are needed to address this threat and evaluate scientific information. And, given that much of misinformation is scientific, how should this be addressed in school science?

This talk will argue that students must be taught how to first evaluate the source and not the content. However, school science sustains the myth that students can undertake the task of evaluating the content – something which only expert scientists can do. What then is the basic knowledge and competency required to evaluate sources of scientific information? And what competencies are needed to interact meaningfully with AI? This talk will explore what this implies for science standards and what should be essential components of any school science education.

## K2: Monitoring and Regulation of Effort and Learning. What does technology have to do with it?

Keynote speaker: Prof. Dr. Anique de Bruin *(Maastricht University)*

October 9, 2025, 9:00-10:00
HS 1.013

**Abstract:**

This keynote explores how learners monitor and regulate effort in technology-enhanced learning environments. Drawing on recent empirical work and theoretical advances, including the S2D2 framework (De Bruin et al., 2023) and findings on refutation instructions (Onan et a., 2024), I examine when and why learners persist with effortful yet effective learning strategies. I argue for a nuanced understanding of effort regulation as both cognitively and motivationally driven. Our research shows that educational technology can support learners' metacognitive monitoring and effort regulation to foster durable learning.

## K3: Digital Learning Tasks: Subject-Specific and Generic

Perspectives on Design and Research

Keynote speaker: Prof. Dr. Timo Leuders *(Freiburg University of Education)*

October 9, 2025, 16:00-17:00
HS 1.013

**Abstract:**

Tasks are the core vehicles of learning—this holds true in both traditional and digital settings. In the context of digital learning environments, however, the design of tasks becomes even more pivotal due to the complex interplay of technological features, pedagogical goals, and learner variability. This keynote explores the design and research of digital learning tasks from both subject-specific and generic perspectives, drawing on empirical insights and theoretical frameworks across disciplines.

Central to the presentation is the question: What makes a digital task effective for learning? Building on the premise that tasks are not neutral carriers of content but active drivers of cognitive engagement, the talk examines the features of digital tasks that guide learning processes—specifically utilizing the constructs of constraints and affordances.

The keynote introduces the CODIL-Framework (Reinhold, F., Leuders, T., Loibl, K., Nückles, M., Beege, M., & Boelmann, J. M., 2024). Learning mechanisms explaining learning with digital tools in educational settings: A cognitive process framework. Ed. Psych. Review), which supports systematic analysis and development of digital learning tasks with a focus on the cognitive activities they elicit—such as comparing, modeling, justifying, or reflecting. Examples from various subject domains illustrate how digital tasks can be designed to foster deep learning, and how they can be researched.

## S1: Digitally Supported Teaching and Learning Scenarios in the Field of Electricity

**Presenters:** Thomas Schubatzky *(University of Innsbruck),* Florian Frank (*University of Würzburg*), Lena Lenz *(Karlsruhe University of Education*), Wolfgang Lutz *(University of Würzburg)*

October 8, 2025, 12:30-14:30
3.101

Teaching fundamental concepts in electricity presents specific challenges for both students and educators. The abstract nature of electrical phenomena, their invisibility in everyday contexts, and widespread misconceptions often hinder the development of solid conceptual understanding. Digital media offer a wide range of opportunities to support learning—by making invisible processes visible, offering individualized feedback, and enriching experimentation. This symposium brings together four current research contributions that explore the use and effects of digital tools in electricity education from various perspectives.

Wolfgang Lutz introduces the Würzburg teaching concept, which builds on the Electric Pressure Curriculum and aims to provide a clear, digital, and interactive learning experience. It includes animations, videos, simulations, and interactive exercises to support both classroom learning and flipped-classroom formats. Findings from two studies—one focusing on usage and perception, the other on conceptual understanding among over 2,000 students—highlight the positive effects of the approach.
Florian Frank reports on an experimental study with 436 eighth-grade students exploring the use of Augmented Reality (AR) in electricity education. The AR application visualizes didactic models and measured values during real experiments. The study investigates cognitive load, time on task, and conceptual learning, while also considering the moderating effect of students' affinity for technology.
Thomas Schubatzky presents a digital self-assessment tool designed to raise metacognitive awareness of household energy use. The tool focuses on qualitative understanding rather than numerical estimates and offers personalized feedback. Validated through empirical research, it supports reflection and energy literacy—important components of education for sustainable development.
Lena Lenz presents how digital learning tasks were designed and evaluated based on the *Electricity Education with Potential* (Elektrizitätslehre mit Potenzial) framework. In addition to teaching fundamental concepts such as electric potential difference and Ohm's Law, the focus lies on data-based argumentation and the use of digital data measurement tools. Empirical studies examine both students' learning gains and their argumentation behavior.
This symposium demonstrates how digital learning environments can enrich electricity education: by addressing misconceptions, promoting conceptual understanding, encouraging argumentation based on data, and fostering self-reflection.

**The Würzburg Teaching Concept: A clear, digital, and interactive approach to the Electric Pressure Curriculum**

Wolfgang Lutz (*University of Würzburg*)

The subject Physics is often perceived as particularly difficult, especially in the case of abstract topic areas like electric circuit teaching. This is, among others, due to the fact that experimental effects can often not be proven directly but only through the use of measuring devices, and because the analogies used by teachers rarely match learners‘ experiences. Numerous misconceptions derived from daily life hinder learners from developing a sound conceptual understanding of the relationships between voltage, resistance, and current.
To tackle these challenges, a teaching concept building upon the Electric Pressure Curriculum was developed in a design-based research project. One of the main focuses was to increase the clearness through self-developed animations, which visualised dynamic physical processes like electric current. In addition, different practice-oriented offers for knowledge application were created, e.g. a broad range of exercises, pupil experiments, digital quizzes, as well as interactive screen experiments and simulations. There also was an explanatory video for each unit to give students the opportunity to independently work through the learning contents at home following the flipped classroom method. The digital materials are available free of charge on the website www.phytet.de.
Within the course of the project, two studies were conducted. The first study targeted the learners‘ user behaviour and their subjective perception of the learning material. The second study involved more than 2,000 pupils. It analysed to what extent the developed materials could benefit a deeper conceptual understanding of electric circuit teaching in traditional teaching and using the flipped classroom method. In the presentation, central finding on the use and effect of the teaching materials as well as results on the effect of the teaching concept are presented.

**The benefits of using digital support material in teaching electricity and the influence of the student's affinity for technology on them**

Florian Frank *(University of Würzburg),* Christoph Stolzenberger*)*, & Thomas Trefzger

Augmented Reality (AR) is a technology which has started to receive increased attention from researchers and practitioners alike for its potential use in teaching physics, especially in the context of electricity and simple circuits, where AR can be used to enhance real-life experiments by superimposing additional information atop them like visualizations of didactic models of electricity or measured values of voltage and current. According to multimedia learning theories, this should ease the cognitive burden caused by processing the learning material and enable the learner to invest more of their cognitive capacity towards understanding the content and assimilating it into their long-term memory. To investigate this proposed interaction, an experimental study was conducted with N = 436 8$^{th}$ grade students to investigate the effects of an AR-application on the learning process (i.e., how much cognitive capacity and time the students had to invest) and on the development of conceptual knowledge. The participants worked through a 5-hour-student laboratory, containing four units pertaining to central electrical concepts and laws: "Current and Voltage", "Electrical Resistance", "Parallel Circuits" and "Serial Circuits". The research design encompassed and therefore allowed the comparison of four settings: using traditional informational graphics-based support materials, using a simulation to visualize didactic models of electricity, using an AR-application to visualize didactic models of electricity, and using an AR-application to visualize didactic models of electricity as well as measured values. It was found that using the AR-application to visualize the didactic models led to a significantly higher conceptual knowledge gain, and that using it to visualize measured values led to a significantly lower time on task. Regarding the impact on the cognitive load that the processing of the learning materials impose, digital materials were found to impose a comparatively lower cognitive burden, an effect that is particularly pronounced with tasks of a higher inherent difficulty. In addition to the conceptual knowledge, the time on task, and the cognitive load, each student's affinity for technology was measured via the sub-factors "enthusiasm for technology" and "self-perceived expertise for using technology". The purpose was to investigate if students exhibiting different levels of affinity for technology also differ in the impact that using the digital support material has on them and their learning process. The analysis of the data regarding this research interest is currently ongoing. The talk will present the experimental study, the application used in the study and the results of the analysis described above.

**A Digital Self-Assessment Tool to Support Metacognitive Awareness on Household Energy Use**

Thomas Schubatzky *(University of Innsbruck)* & Lisa Eisele (*Physics Education Research Group, University of Innsbruck*)

Although electricity is key to modern life, its everyday use often remains invisible—not only for students. Misconceptions about which appliances and actions consume significant amounts of energy can hinder both conceptual understanding and climate-relevant behaviour. To address this gap, we developed a digitally supported self-assessment tool that helps learners reflect on their intuitive knowledge of household electricity consumption. The tool is based on a newly developed and validated test instrument comprising 23 items across four thematic domains: heating water, heating solid objects (e.g., ovens), charging batteries, and lighting. These categories reflect common contexts in which electricity is used, but often misunderstood. Instead of requiring numerical estimates, the items ask students to make qualitative comparisons (e.g., "Which consumes more

energy: boiling 1l of water or running a Wi-Fi router for a day?"). This approach is supposed to allow a focus on conceptual understanding. This underlying instrument was empirically validated in a German-speaking sample (N = 447), showing satisfactory reliability (α = .69) and internal structural validity. On average, participants answered 63.9% of items correctly; however, students under 21 scored only 51%, demonstrating the need for targeted educational interventions. Thematic analysis of item-level responses revealed significant misconceptions, particularly about heating-related energy use. These findings informed the structure and content of the feedback tool.
In its digital version, the tool offers individualized, automated feedback: Students receive both their overall performance and category-specific results. This feedback is designed to foster metacognitive reflection ("Which areas do I already understand?"). The tool can be used independently, as a starting point for energy usage in households, or as a basis for project-based learning. In our talk, we will demonstrate the digital tool, share results from classroom piloting, and discuss its integration into electricity-related teaching scenarios. We argue that such low-threshold, feedback-driven tools can support both knowledge acquisition and critical reflection—key components in fostering energy literacy and sustainability-oriented agency in students.

**Digital Learning Tasks in „Electricity with Potential"**
Lena Lenz *(Karlsruhe University of Education)*, Tobias Ludwig & Jan-Philipp Burde

The teaching concept *Electricity Education with Potential* (Burde, 2018) has proven to be highly effective in fostering fundamental concepts related to simple electric circuits in lower secondary education. Building on this concept, digital learning tasks on the topics of *electric potential difference* and *Ohm's Law* were developed and piloted with approximately 40 students to evaluate their learning effectiveness. Through interactive tasks that include formative assessment and inquiries using digital data acquisition systems, students work through the subject content independently and at their own pace. They have the option to seek support from the teacher, who takes on the role of a learning facilitator.
In addition to the topics covered by the original teaching concept, a complementary digital learning task was developed focusing on the *magnetic field of a current-carrying conductor*. This task emphasizes digital measurement and data-based argumentation. It was implemented and evaluated during a project day with around 40 students from two 8th-grade classes at the PhyLa teaching and learning laboratory. The evaluation of the learning task was twofold: first, a pre-post design with eight items was used to assess students' knowledge gains related to the core content of the task. Second, students were given argumentation tasks based on their experimental work. According to Sampson & Clark (2008), arguments can be analyzed from three perspectives: their structure (Toulmin, 2003) their *Nature of Justification* (Ludwig et al., 2019), and their content quality (Kok & Priemer, 2023). Previous studies have shown that students—despite collecting and partially analyzing quantitative experimental data—rarely use such data to justify their hypotheses (Lenz, submitted). The use of digital measurement tools now allows data to be made available to students during experimentation in processed forms such as graphs. This changes the representation of data compared to earlier studies. This raises the question of whether this shift in data representation encourages students to adapt their argumentation behavior and make more frequent use of data-based justifications to support their hypotheses, rather than manually evaluating and comparing data series (Lenz & Ludwig, 2021). Therefore, students' arguments are analyzed particularly by their *Nature of Justification*.
The presentation will showcase both approaches to implementing digital learning tasks within the *Electricity with Potential* framework and initial findings on students' argumentation behavior when working on the *magnetic field of a current-carrying conductor* task.

# S2: Evidence-Informed School Practice: Challenges, Capacities, and Catalysts for Transformation

**Presenters:** Florian Kühlwein *(Karlsruhe University of Education),* Despoina Georgiou *(Utrecht University)*, Audrey Michal *(University of Michigan-Flint)*
**Chair:** Samuel Merk (*Karlsruhe University of Education*)
**Discussant:** Rainer Bromme *(University of Münster*)

October 8, 2025, 12:30-14:30
3.107

**Advancing Digital Task Efficacy: International Perspectives on Technology-Enhanced Education**

**Symposium overview abstract:**

Using evidence from educational research has the potential to improve both teaching and school quality. Given this potential, and the growing body of evidence—especially in light of the recent rise in randomized controlled trials in education (White, 2019)—it is not surprising that an evidence-informed school approach has been promoted increasingly at the international level (Brown et al., 2017; Council of the European Union, 2024; Davies, 1999; Groß Ophoff et al., 2023; Slavin, 2002; White, 2019).
As evidence-informed processes start influencing educational practices worldwide (e.g., Higgins et al., 2022; Ion et al., 2024; Malin et al., 2020), it is crucial to examine how findings from research can be successfully disseminated and integrated into educational settings to realize the potential. To this end, this symposium features current empirical research from quantitative and qualitative studies related to the communication and use of evidence within educational contexts. It explores overarching (international) challenges, capacities, and catalysts for evidence-informed decision-making. The aim is to provide insights into ongoing empirical research, spark fruitful discussions among the attendees through an interactive Q&A session, and foster collaborations among researchers.

**Keywords:**

*Evidence-informed school practice, Communicating evidence, Teachers' decision-making, Teachers' evaluations of evidence, Teachers' evidence use*

**Teachers' Evaluations of Educational Research and Their Relevance for Evidence-Informed Decision-Making**
Florian Kühlwein et al. (*Karlsruhe University of Education*)

**Co-authors (Study 1):** Kirstin Schmidt *(Karlsruhe University of Education),* Tom Rosman *(Leibniz Institute for Psychology Information (ZPID))*, Hugues Lortie-Forgues (*Loughborough University*) & Samuel Merk (*Karlsruhe University of Education*)
**Co-authors (Study 2):** Samuel Merk *(Karlsruhe University of Education),* Jürgen Schneider *(DIPF | Leibniz Institute for Research and Information in Education, Teacher and Teaching Quality)* & Kirstin Schmidt *(Karlsruhe University of Education)*

**Abstract:**

Teachers must make numerous instructional decisions as part of their work. For example, during lesson planning, they must select materials that are both relevant and engaging, while also choosing an instructional strategy that effectively conveys the content and motivates students to learn. Integrating findings from educational research into teachers' decision-making has the potential to lead to more objective and better-informed decisions that can improve teaching quality and, as a result, enhance student learning.
Given this context, I will present findings from two recent experimental studies investigating how teachers' evaluations of educational research shape their appreciation and (intended) implementation of different instructional strategies. Specifically, I will examine the roles of teachers' competencies (e.g., their ability to estimate effect sizes) and their attitudes (e.g., their perceived relevance of research findings) in engaging with educational research findings that include effect size information (Kühlwein et al., under review). Is teachers' (intended) use of educational research findings for instruction shaped more by their competencies or their attitudes?
Furthermore, I will share information on the differential effects of communicating effect sizes verbally (e.g., as Cohen's $U_3$ phrases) or visually (e.g., in the form of half-eye plots) to teachers (Kühlwein et al., 2025). In other words, what are the respective advantages and disadvantages of each presentation mode for communicating scientific evidence to teachers in a user-friendly manner? Both studies underscore the need to better understand how teachers, as scientific laypersons, make instructional decisions, and how these decisions are influenced by statistical information, particularly effect sizes, and their (preexisting) attitudes towards educational research.

**Shifting the Lens: Toward Evidence-Informed Practice in Teacher Development**
Despoina Georgiou et al. *(Utrecht University)*

**Co-authors (Study 1):** María Rodríguez Alcolea *(Departament d'Educació, Barcelona)*
**Co-authors (Study 2):** Annika Diery *(youKnow Gmbh Munich)*, Sog Yee Mok *(University of Teacher Education of the Grisons)*, Frank Fischer *(LMU Munich, Germany)*, Tina Seidel *(TUM Munich)*
**Co-authors (Study 3):** Wieke de Kruyf *(Utrecht University)*

**Abstract:**

While international initiatives continue to promote the use of research in education (Council of the European Union, 2024), the ways in which teachers engage with evidence remain underexplored and vary across different stages of professional development (Ion et al., 2024). This presentation draws on the Evidence-Informed Teaching Implementation Model (Georgiou et al., 2023) as a conceptual lens to explore how different groups of educators engage with research evidence. Rather than following the entire implementation process, the model is used to frame the drivers, barriers, and developmental needs that shape evidence use across the teacher development continuum. A survey of $N$ = 120 primary school teachers reveals that information literacy and teaching experience shape engagement with evidence, with younger, digitally fluent teachers reporting more positive attitudes (Rodriguez & Georgiou, under review). In a complementary qualitative study involving $N$ = 12 university-based teacher educators, we find that although evidence-informed teaching is broadly valued, its implementation is constrained by institutional challenges, limited time, and barriers in translating research into practice (Georgiou et al., 2023). Preliminary findings from a third study on pre-service teachers' source selection practices highlight the importance of fostering critical engagement with research early in teacher education. Together, these studies provide a multi-level perspective on how educators relate to evidence at different career stages. We argue that supporting

evidence-informed practice requires a holistic, differentiated strategy, one that recognizes the unique roles, capacities, and challenges of all stakeholders involved in teacher education.

**Understanding the Assumptions Laypeople Make When Applying Scientific Evidence to Education Decisions**

Audrey Michal et al. (*College of Arts, Sciences and Education, The University of Michigan-Flint)*

**Co-authors (Study 1):** Priti Shah (University of Michigan, Ann-Arbor)
**Co-authors (Study 2):** Priti Shah (University of Michigan, Ann-Arbor)

**Abstract:**

Educators are increasingly relying on scientific evidence to inform their decision making. However, because scientific findings are often presented in oversimplified or sensationalized ways, and because laypeople may generally expect science findings to have strong implications and be widely generalizable to everyone, educators may struggle to apply science findings appropriately to school policy and classroom decisions – that is, they may over-extrapolate from study findings.

However, little is known about which study factors educators consider (if any) when deciding whether study findings are applicable to a given decision. In this talk, I will present recent work investigating the extent to which laypeople are influenced by specific application-related factors of education research studies, including the sample population used and the practical significance of the findings. For example, are people less willing to extrapolate from a study based on a sample population (e.g., high school students) that differs meaningfully from a target student population (e.g., elementary students), compared to a sample population that is more similar (e.g., other elementary students) (Michal & Shah, in preparation)?

Do laypeople assume that science findings are generalizable to a given decision when these application-related study factors are only described in vague, general terms (e.g., “students performed better in Group A than Group B”) (Michal & Shah, 2024)? Having a better understanding of the assumptions that laypeople make about the applicability of scientific evidence to education decisions can ultimately clarify how best to communicate relevant research findings to educators.

**References:**

Bauer, J., Prenzel, M., & Renkl, A. (2015). Evidenzbasierte Praxis - im Lehrerberuf?! Einführung in den Thementeil [Evidence-based practice in teacher education - Introduction to the topic]. *Unterrichtswissenschaft*, *43*(3), 188–192.

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational Research*, *59*(2), 154–172. https://doi.org/10.1080/00131881.2017.1304327

Council of the European Union. (2024). *Council conclusions on promoting evidence-informed policy and practice in education and training to achieve the european education area* [Website]. https://eur-lex.europa.eu/eli/C/2024/3642/oj

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *2*, 108–121. https://doi.org/10.1111/1467-8527.00106

Georgiou, D., Diery, A., Mok, S. Y., Fischer, F., & Seidel, T. (2023). Turning research evidence into teaching action: Teacher educators’ attitudes toward evidence-based teaching.

*International Journal of Educational Research Open*, *4*, Article 100240. https://doi.org/10.1016/j.ijedro.2023.100240

Groß Ophoff, J., Brown, C., & Helm, C. (2023). Do pupils at research-informed schools actually perform better? Findings from a study at English schools. *Frontiers in Education*, *7*, Article 1011241. https://doi.org/10.3389/feduc.2022.1011241

Higgins, S., Katsipataki, M., Villanueva Aguilera, A. B., Dobson, E., Gascoine, L., Rajab, T., Kalambouka, A., Reardon, J., Stafford, J., & Uwimpuhwe, G. (2022). The teaching and learning toolkit: Communicating research evidence to inform decision-making for policy and practice in education. *Review of Education*, *10*(1), Article e3327. https://doi.org/10.1002/rev3.3327

Ion, G., Kowalczuk-Walędziak, M., & Brown, C. (2024). Unlocking the potential of research-informed practice: Insights into benefits, challenges, and significance among teachers in Catalonia, Poland, and England. *European Journal of Education*, *59*(3), Article 12639. https://doi.org/10.1111/ejed.12639

Kühlwein, F., Merk, S., Schneider, J., & Schmidt, K. (2025). How to communicate evidence to teachers: Comparing the effects of verbal and visual effect size representation [Registered Report - Stage I]. *Learning and Instruction*, 100, Article 102179. https://doi.org/10.1016/j.learninstruc.2025.102179

Malin, J. R., Brown, C., Ion, G., Van Ackeren, I., Bremm, N., Luzmore, R., Flood, J., & Rind, G. M. (2020). World-wide barriers and enablers to achieving evidence-informed practice in education: what can be learnt from Spain, England, the United States, and Germany? *Humanities and Social Sciences Communications*, *7*(1), 99. https://doi.org/10.1057/s41599-020-00587-8

Michal, A. L., & Shah, P. (2024). A practical significance bias in laypeople's evaluation of scientific findings. *Psychological Science*, *35*(4), 315–327. https://doi.org/10.1177/09567976241231506

Rodriguez, M., & Georgiou, D. (under review). *Understanding the drivers of primary school teachers' attitudes toward evidence-informed practices*.

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

White, H. (2019). The twenty-first century experimenting society: The four waves of the evidence revolution. *Palgrave Communications*, *5*(1), Article 47. https://doi.org/10.1057/s41599-019-0253-6

## S3: Seeking Ways to Individualize and Tailor Motivation Interventions

**Presenters:** Hanna Gaspard *(University of Konstanz)*, Liene Brandhuber *(Karlsruhe University of Education)*, Emily Q. Rosenzweig *(Teachers College at Columbia University, NY)*, Patrick N. Beymer *(University of Cincinnati, OH)*
**Chairs:** Cora Parrisius & Liene Brandhuber *(Karlsruhe University of Education)*
**Discussant:** Jacquelynne S. Eccles *(University of California-Irvine, CA)*

October 8, 2025, 15:00-17:00
3.101

**Abstract:**

To date, motivation interventions in education typically target a single motivational construct, following a "one-size-fits-all" approach (e.g., utility-value interventions; Hulleman & Harackiewicz, 2009; social belonging interventions; Walton & Cohen, 2011; growth mindset interventions; Yeager

et al., 2016). Although motivation interventions yield positive effects *on average* (Lazowski & Hulleman, 2016), tailoring interventions to individual learner differences—rather than aiming to improve outcomes on average—may enhance motivation more effectively (Rosenzweig & Wigfield, 2022). While recent studies have begun exploring different approaches to individualization in educational settings, research on tailoring motivation interventions to individual needs remains scarce. The symposium "Seeking Ways to Individualize and Tailor Motivation Interventions" serves as a platform to discuss potential approaches and challenges for tailoring motivation interventions that benefit as many people as possible. It brings together international research from four groups, each addressing a different aspect of this complex task.

The first study explores personalizing the timing of motivation interventions to enhance their effectiveness. By systematically monitoring the effects of a utility-value intervention (UVI) over the course of a semester, this study found that the positive effects on students' perceived course value and interest peaked around exam periods, suggesting that students may be especially receptive to such interventions at these times. The second study also focuses on increasing the effectiveness of UVIs, but approaches personalization differently. Derived from a re-analysis of a UVI, this study identified individual characteristics of students (e.g., gender, migration background, and vocational interests) based on which the individualization of UVIs seems promising, given that students' reception of the intervention material varied systematically depending on these characteristics. The third study highlights the interconnection between students' motivational beliefs and examines the role of autonomy support in tailoring interventions grounded in situated-expectancy value theory (SEVT). The results show that allowing students to choose how to reflect on and weigh multiple motivational beliefs can be a promising approach when customizing SEVT-interventions. The fourth study examines various approaches to identifying individual motivational needs when addressing them with interventions grounded in different motivational frameworks. Importantly, the study highlights the methodological challenge of accurately identifying these needs, as different approaches may produce different intervention effects.

All four contributions underscore that individualizing motivation interventions is a complex endeavor requiring further research, with multiple factors needing careful consideration. The symposium ends with a discussion on how motivation interventions can be further improved in the future by making use of individualization.

**References:**

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting Interest and Performance in High School Science Classes. *Science*, *326*(5958), 1410–1412. https://doi.org/10.1126/science.1177067

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Walton, G. M., & Cohen, G. L. (2011). A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students. *Science*, *331*(6023), 1447–1451. https://doi.org/10.1126/science.1198364

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., Roberts, A., Trott, J., Greene, D., Walton, G. M., & Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. https://doi.org/10.1037/edu0000098

**Timing Matters: Lessons from a Weekly Utility-Value Intervention**

Patrick N. Beymer *(University of Cincinnati, OH)*

**Abstract:**

When considering motivational interventions, researchers often face the challenge of determining the most optimal time for intervention. Utility-value interventions (UVIs) are typically implemented at 1-3 predetermined timepoints across a semester; however, these predetermined timepoints may not align with periods when students are most receptive to intervention. As a first step in considering how to personalize the timing of motivational interventions, we developed, implemented, and evaluated a weekly UVI in which students reflected on the value of course content for 13 consecutive weeks in two separate courses: physics ($N$ = 631) and chemistry ($N$ = 1,348). Using a randomized controlled trial, we examined effects on weekly course importance value, weekly course interest, final course interest, STEM career intentions, and performance.

In both courses, results suggested that those who participated in the weekly UVI had higher average weekly importance value ($\beta_{chemistry} = 0.04$, $p < 0.05$; $\beta_{physics} = 0.08$, $p < 0.05$) and weekly interest ($\beta_{chemistry} = 0.06$, $p < 0.01$; $\beta_{physics} = 0.08$, $p < 0.01$) compared to those in the control group. Notably, these effects were more pronounced closer to exam weeks, especially the first exam.

In chemistry, the weekly UVI was also found to be effective in increasing final course interest ($\beta = 0.06$, $p < 0.05$) and STEM career intentions ($\beta = 0.06$, $p < 0.05$). In physics, the weekly UVI increased course grades for racially marginalized students ($\beta = 0.12$, $p < 0.01$), when compared to racially marginalized students in the control group. Finally, in chemistry, the weekly UVI increased course grades for women ($\beta = 0.06$, $p < 0.05$), when compared to women in the control group.

The positive effects on weekly importance value and weekly interest in both courses highlight the potential of brief weekly motivational interventions and provide insights into optimal timing for such interventions. However, mixed long-term effects, course-specific outcomes, and inconsistent moderation patterns suggest important boundary conditions that warrant further examination. Our results offer researchers a framework for considering how the timing of interventions can be personalized to better meet the needs of individual students.

**Promoting Motivation in Mathematics Through Personalized Relevance Interventions**

Hanna Gaspard *(University of Konstanz)*

**Abstract:**

Many adolescents do not see the relevance of mathematics for their lives, and math utility value has been found to decrease throughout secondary school (e.g., Gaspard et al., 2017). Relevance interventions have shown a great potential to foster students' motivation and achievement in math and sciences (Lazowski & Hulleman, 2016; Rosenzweig & Wigfield, 2016). Yet, some studies have also yielded a mixed pattern of effects (e.g., Gaspard et al., 2021) and the effects were found to vary depending on students' characteristics (Rosenzweig & Wigfield, 2016). In the MoMa-PR project, we therefore aim to study how relevance interventions can be personalized for students with different characteristics and whether personalizing these interventions yields stronger effects on motivation and achievement.

In the first step, data from a prior cluster-randomized trial testing a non-personalized relevance intervention in mathematics classrooms (Gaspard et al., 2021) was reanalyzed. To this end, we focused on the $N$ = 1,198 ninth-grade students out of 52 classes that were part of the intervention conditions. The 90-min intervention consisted of an instructor-led psychoeducational presentation

and an individual task that asked students to rate six quotes from young adults referring to the relevance of mathematics. Students' ratings of the personal relevance of these six quotes were used as indicators of student responsiveness. A broad set of student characteristics was considered as predictors, including gender, migration background, prior motivation, STEM career aspirations, and vocational interests. Student characteristics predicted students' ratings of the personal relevance of the different quotes in line with expectations. Furthermore, a higher personal relevance across the six quotes was associated with positive changes in perceived utility value from pretest to posttest and follow-up in the intervention conditions. This suggests that personalization of the presented quotes might make these interventions more effective.
In the next step, we will therefore systematically test whether personalizing these quotes to students' characteristics yields stronger intervention effects compared with one-size-fits all interventions. Digital intervention materials will be developed that allow for such a personalization. Rapid, iterative A/B testing will first be used to compare personalized vs. non-personalized interventions based on different factors for personalization, including students' gender, migration background, and vocational interests. A randomized controlled study will then be conducted to examine the efficacy of a personalized relevance intervention (in terms of all factors that result in substantial differences in the A/B testing) compared with a "one size fits all" intervention and a control condition.

**References:**

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

Gaspard, H., Parrisius, C., Piesch, H., Kleinhansl, M., Wille, E., Nagengast, B., Trautwein, U., & Hulleman, C. S. (2021). The potential of relevance interventions for scaling up: A cluster-randomized trial testing the effectiveness of a relevance intervention in math classrooms. Journal of Educational Psychology, 113(8), 1507–1528. https://doi.org/10.1037/edu0000663

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. Review of Educational Research, 86(2), 602–640. https://doi.org/10.3102/0034654315617832

Rosenzweig, E. Q., & Wigfield, A. (2016). STEM motivation interventions for adolescents: A promising start, but further to go. Educational Psychologist, 51(2), 146–163. https://doi.org/10.1207/s15326985ep2003

**Should We Customize to the Belief, or the Belief System? Testing Two Approaches for Customizing Motivational Supports During Expectancy-Value Interventions**

Emily Q. Rosenzweig *(Teachers College at Columbia University, NY)*, Xiao-Yin Chen *(University of Tennessee)*, Yichi Zhang *(University of Georgia, GA)*

**Abstract:**

Motivational interventions are more effective if customized (Rosenzweig et al., 2019), raising the question of how best to adapt interventions to students' unique motivational needs. In two pre-registered experiments, we tested two new prototypes for customizing interventions grounded in situated expectancy-value theory (Eccles & Wigfield, 2020): (1) Customizing which motivational

beliefs were supported, and (2) customizing how students were able to weigh multiple motivational beliefs together.
In Study 1 ($N$ = 324), students in a U.S. university subject pool completed a chemistry lesson including either an adaptive motivational intervention (supports addressed students' unique motivational needs), a multifaceted intervention (students received multiple motivational supports), or a control condition (no supports). The adaptive condition used an online tool to assess students' motivational beliefs about chemistry and provided only supports for constructs on which students scored low (e.g., A student with low competence-related beliefs but high value received competence support). Yet only the multifaceted intervention promoted students' value beliefs and chemistry knowledge, $\beta$ > 0.13, $p$ < .022. In the adaptive condition, competence-related beliefs and task values were positively correlated ($r$ =.63), and few students (25%) had only one low motivational belief. Results suggest that instead of having isolated motivational needs, students weighed multiple motivational beliefs as closely interrelated in their minds.
In Study 2 ($N$ = 105), U.S. college computer science students completed an online intervention designed to help them weigh multiple motivational beliefs together, or a control condition. The intervention provided all students with the same content, about weighing challenges of computer science learning against benefits (related to both competence-related beliefs and task values). However, we added elements of autonomy support to the intervention, which allowed students to customize how they reflected on weighing beliefs together. The intervention showed positive overall effects on self-efficacy for learning computer science, $\beta$ = 0.25, $p$ = .007 and it reduced cost perceptions among students who were initially more likely to switch their majors.
Together, findings suggest that it may be fruitful to test intervention customization strategies that acknowledge systems of motivational beliefs working together.

**Implementing Individualized Motivation Interventions: Lessons Learned from Two Intervention Studies**

Liene Brandhuber *(Karlsruhe University of Education)*, Nilani Suriakumar (*Karlsruhe University of Education*), Samuel Merk (Karlsruhe University of Education), Benjamin Nagengast *(University of Tübingen),* Cora Parrisius *(Karlsruhe University of Education)*

**Abstract:**

To support all learners effectively, it may be beneficial to tailor motivation interventions to individual needs (individualized interventions) rather than offering the same intervention to everyone (one-size-fits-all approach) (Rosenzweig & Wigfield, 2022). However, research on tailoring individualized motivation interventions within educational contexts is still scarce. To test the effectiveness of individualized interventions, we conducted two consecutive motivation intervention studies (Study 2 currently ongoing). In both studies, we implemented five interventions grounded in different motivational frameworks (i.e., growth mindset intervention, social belonging intervention, achievement goal intervention, attributional retraining, and utility value intervention). In the *needs-matching* condition students received one of these five motivation interventions that matched their individual motivational need; in the *needs non-matching* condition, they received one of the other four. Needs were identified based on students' responses to a baseline questionnaire. The interventions were implemented in methodology lectures for first-year students enrolled in teacher training programs ($N_1$ = 387, $N_2$ = tbd). The results of Study 1 provided the first empirical evidence of the superiority of individualizing motivation interventions, as implemented in this study, over the typical one-size-fits-all approach. In addition to the positive effects of the individualization process on various motivational outcomes, we also observed positive effects on more global outcomes such as dropout intentions and perceived costs.

Despite their benefits, individualized motivation interventions have so far largely relied on researchers' subjective decisions to determine individual needs. Given that need identification is fundamental to the individualization process, the chosen approach could meaningfully affect the results. Therefore, the question of how to appropriately determine individual motivational needs remains unresolved: Who needs which intervention the most? To address this question, we will re-analyze the data from both studies—once the full dataset is available—by simulating various scenarios of need identification and examining their impact on potential intervention effects. Specifically, we will apply alternative need identification methods post hoc, differing from those used in the original studies, and assess how these variations influence the observed intervention effects. Thus, the present study makes a methodological contribution to future research, providing insights into potential ways of identifying different motivational needs and highlighting where challenges may arise. The results of the analyses will be presented at the conference.

**References:**

Rosenzweig, E. Q., & Wigfield, A. (2022). Motivational Interventions in Education: Five Big Questions to Consider When Designing Effective Interventions. In A. O'Donnell, N. C. Barnes, & J. Reeve (Eds.), *The Oxford Handbook of Educational Psychology* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199841332.013.43

## S4: Innovating Task Design in Economic Education: Data-Driven Insights in Task Quality and Practical Applications

**Presenters:** Stephan Friebel-Piechotta *(Institute for Economic Education, University of Oldenburg),* Jonathan Heitzler *(University of Education Freiburg),* Julian Polly *(University of Education Karlsruhe)*
**Chair & moderation:** Claudia Wiepcke *(Karlsruhe University of Education)*

October 8, 2025, 15:00-17:00
3.107

Life in the early 21st century is shaped by massive and interconnected transformation processes. Climate change, demographic shifts, digitalization, globalization, and geopolitical instability point to profound societal, ecological, and economic change. While structural transformation is not new, the pace and magnitude of current developments are unprecedented. In response, economic education must evolve to equip learners with the competencies needed to navigate complexity, uncertainty, and ethical ambiguity.
In the face of accelerating global transformations, economic education must adapt to foster the skills and competencies that future generations of learners need to navigate complexity, uncertainty, and ethical ambiguity. High-quality digital learning tasks are central instruments for fostering such competencies. This symposium addresses the design of digitally enriched tasks that promote conceptual understanding, behavioral insight, and ethical reflection—linking instructional design, competency modeling, behavioral interventions, and technological implementation.
**Stephan Friebel-Piechotta** examines **how behavioral economics can support Education for Sustainable Development (ESD)**, with a particular focus on the attitude-behavior gap. He presents digital classroom experiments and accompanying instructional materials, alongside findings from an intervention study used to design and validate them. The materials and experiments aim to reduce behavioral biases and promote sustainable consumption.
**Julian Polly** introduces the **Behavioral Economic Competence (BEC) framework**, which integrates behavioral economic literacy and numeracy. Drawing on preliminary findings from a quasi-experimental study, he compares the effects of digital and analog experiments with textbook-

based behavioral tasks. This contribution bridges instructional design and competency modeling, highlighting the potential of digital experiments to foster economic decision-making skills.
**Jonathan Heitzler** presents a **digital instructional design for teaching pricing theory in economic education**, structured in two pedagogical phases. Anchored in cognitive learning theory, the approach uses digital market simulations followed by abstract conceptualization through "concreteness fading". This design aims to fostering deep conceptual understanding with digital simulations and self-explanation prompts.
The session is chaired by **Claudia Wiepcke**, who will guide the discussion following each contribution.

**Session 1: Behavioural Economics Approaches Towards Education for Sustainable Development: Teaching materials on biases and debiasing**

Stephan Friebel-Piechotta *(Institute for Economic Education, University of Oldenburg)*

Education for Sustainable Development (ESD) is a central concern of school education. Educational approaches are primarily aimed at changing attitudes and awareness, which in turn should lead to changes in behaviour. While this approach is necessary, it is not sufficient to achieve the desired environmental behaviour, as attitudes do not translate into behaviour in a linear way. The so-called Attitude Behaviour Gap (also: Values Action Gap, cf. Kollmuss & Agyeman, 2002) describes precisely this gap between the attitudes of economic actors (e.g. consumers) and their actual behaviour (e.g. their consumption decisions). Thus, contrary to the assumptions of planned behaviour theories (cf. Ajzen, 1985), sustainable behavioural intentions are not always directly translated into corresponding consumer actions.
Approaches from behavioural economics provide a starting point for explaining and mitigating the attitude-behaviour gap. Unlike traditional economic theory, behavioural economics assumes that people have limited cognitive and affective capabilities. A key element of the behavioural approach is therefore the psychological underpinning and analysis of individual behaviour. For example, cognitive processing, attitudes, expectations, attributions and heuristics are analysed. Behavioural research enables us to better understand the reactions of individuals in economically influenced life situations and to uncover and illustrate causal relationships. In this way, it can help explain the attitude-behaviour gap. Secondly, it can identify starting points for reducing the attitude-behaviour gap.
Based on these considerations, the approach of behavioural economics-based ESD was developed as part of a project funded by the German Federal Environmental Foundation (DBU) (cf. Allbauer-Jürgensen et al. 2024). Based on this theoretical framework, teaching materials for economics lessons have been developed. Digital economic experiments are at the core of these materials. These will be presented in the lecture, as well as the digital experimental intervention study (cf. Friebel-Piechotta & Loerwald 2023), from which the materials were developed.

**References:**

Allbauer-Jürgensen et al. (2024): Behavioural Economics Approaches Towards Education for Sustainable Development: The Role of Biases and Debiasing in Sustainable Consumption, V. Holz et al. (Esd.), Bildung für eine nachhaltige Entwicklung im Umbruch? Beiträge zur Theorieentwicklung angesichts ökologischer, gesellschaftlicher und individueller Umbrüche (123-140). Opladen, Berlin, Toronto: Verlag Barbara Budrich.

Ajzen, I. (1985). From intensions to actions: A theory of planned behavior. J. Kuhl & J. Beckmann (Eds.), Action control (11-39). Berlin/Heidelberg: Springer.

Friebel-Piechotta, S. & Loerwald, D. (2024): Ansatzpunkte für eine verhaltensökonomisch fundierte

Bildung für nachhaltige Entwicklung. F. Birke et al. (Eds.), Ökonomische Bildung als Allgemeinbildung (51-69). Berlin/Heidelberg: Springer.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8, 239–260.

Symposium “Innovating task design in economic education: Data-driven insights in task quality and practical applications”

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8, 239–260.

**Session 2: Fostering Behavioral Economic Literacy and Numeracy through Digital Classroom Experiments: Towards a Competence-Based Model of Behavioral Economics in Education**

Julian Polly *(University of Education Karlsruhe),* Claudia Wiepcke, *(University of Education Karlsruhe),* Taiga Brahm, *(University Tübingen - Economic Education)*

In an increasingly complex world, fostering young people’s capacity for informed and reflective decision-making is a key objective of economic education (Blanck 2025). While traditional curricula are rooted in neoclassical models of rational choice, recent decades have witnessed the emergence of behavioral economics as a complementary paradigm, both in research and educational settings. Behavioral economics broadens the understanding of decision-making by examining systematic deviations from rational behavior.

Behavioral classroom experiments are a promising task format for developing decision-making competencies in secondary education. (Weyland 2019; Warwas & Schmidt 2024). These tasks simulate real-life decision-making situations, positioning students as both participants and observers of behavioral regularities. This dual role enhances metacognitive reflection, enabling students to recognize their own decision-making biases and reflects the epistemological core of behavioral economics, where knowledge is generated through controlled experimentation. The experiments align with key quality criteria in economic didactics, including proximity to real-life contexts, authenticity of decision situations, and connection to scientific practices (Gross & Weyland 2022; Kleinknecht et al. 2013). By engaging with key behavioral phenomena—such as loss aversion, social preferences, or temporal discounting—students gain insight into how economic behavior deviates from neoclassical assumptions and understand the relevance of behaviorally informed decision-making in everyday contexts.

This presentation introduces the framework of Behavioral Economic Competence (BEC), which integrates cognitive, motivational, volitional, and social dimensions of decision-making within behavioral economic contexts. BEC comprises two core dimensions: Behavioral Economic Literacy (BEL), the ability to recognize and critically reflect on behavioral phenomena, and Behavioral Economic Numeracy (BEN), the capacity to analyze and evaluate these phenomena using quantitative reasoning and empirical evidence. The BEC framework extends the Integrative Model of Economic Competence by Seeber et al. (2012), by incorporating insights from behavioral economics into the broader discourse on economic competences in education.

The contribution presents a quasi-experimental study investigating how different instructional approaches influence the development of BEC among secondary students. Three experimental conditions are compared: a digital intervention group engaging with interactive behavioral experiments, an analog group using paper-based experiments, and a control group working with conventional textbook tasks.

**References:**

Blanck, Bettina (2025): „Sag' du mir, was ich machen soll!" - Reflexive Entscheidungen und Förderung von Entscheidungskompetenzen aus erwägungstheoretischer Sicht Zeitschrift für ökonomische Bildung, (14), 57-80.

Gross, A. & Weyland, M. (2022). Aufgaben in der ökonomischen Bildung – ein systematischer Literaturüberblick. Zeitschrift für ökonomische Bildung, (Sondernummer, Jahresband DeGÖB 2021), 42–77. Center of Economic Education in Siegen.

Kleinknecht, M., Bohl, T., Maier, U. & Metz, K. (2013). Lern- und Leistungsaufgaben im Unterricht: fächerübergreifende Kriterien zur Auswahl und Analyse. Bad Heilbrunn: Verlag Julius Klinkhardt.

Seeber, G., Retzmann, T., Remmele, B. & Jongebloed, H.-C. (2012). Bildungsstandards der ökonomischen Allgemeinbildung: Kompetenzmodell – Aufgaben – Handlungs-empfehlungen. Wochenschau Verlag.

Weyland, M. (2019). Experimentelles Lernen im Ökonomieunterricht – wirksam, effizient, unverzichtbar? Zeitschrift für ökonomische Bildung, (8), 1–26. Center of Economic Education in Siegen.

**Session 3: Bridging Experience and Abstraction: A Digital Two-Phase Approach to Teaching Pricing Theory**

Jonathan Heitzler *(University of Education Freiburg),* Franziska Birke, *(University of Education Freiburg)*, Katharina Loibl, *(University of Education Freiburg)* Timo Leuders, *(University of Education Freiburg)*

Understanding how pricing mechanisms work under idealized conditions is essential for students to analyze the limitations of real-world markets and assess the implications of policy decisions. However, understanding pricing as a systemic interaction is challenging for students. Studies show that students often have naïve conceptions of price formation.

In economics education, instruction often progresses from concrete, experience-based examples to idealized representations of a model or principle (Euler and Hahn 2014, S. 158–159) Classroom experiments –such as the so-called Apple Market Experiment– offer students the opportunity to actively engage in simulated market exchanges. When complemented by subsequent explicit instruction, it can promote the development of deeper conceptual understanding (Fyfe et al. 2014).

The study explores a digitally enriched, two-phase instructional design (Loibl et al. 2024) aimed at supporting conceptual learning in pricing theory. In the first phase, students acquire experience-based situational knowledge through interacting in an Apple Market Experiment and reflecting on the market outcomes. Although this knowledge does not encompass the full scope of the target concept, it serves as a stepping stone in the subsequent learning process. Accordingly, we refer to it as intermediate knowledge, highlighting its role as a bridge between concrete experience and abstract conceptualization. In the second phase, a concreteness fading strategy is implemented to gradually transition from concrete experiences to the abstract representation of pricing in the graphical supply-and-demand model (Fyfe and Nathan 2019). A digital simulation of the Apple Market Experiment serves as an intermediate representation step.

To exploit the benefits of concreteness fading, the instructional design includes features that explicitly link the three stages: classroom interaction during the Apple Market Experiment, digital simulation, graphical supply- and-demand model (Fyfe und Nathan 2019). Research on learning with multiple representations suggests that self-explanation prompts and signaling techniques can support learners in identifying structural similarities between representations (Rau et al. 2015). Identifying these similarities, in turn, facilitates the abstraction processes from concrete experiences to the abstract supply-and-demand model. Based on previous research, scaffolded self-explanation prompts are assumed to be most beneficial for students with low intermediate

knowledge (Berthold et al. 2009) whereas open self-explanation prompts may engage students with high intermediate knowledge in more advanced elaboration processes (Lee et al. 2014).
The study uses a quasi-experimental 2x2 pre-posttest design, with students randomly assigned to conditions with either scaffolded or open self-explanation prompts, with or without signaling. It investigates the main and interaction effects of these supports, as well as whether effects are moderated by students' intermediate knowledge and mediated by the abstraction processes. Preliminary findings will be presented at the symposium "Innovating Task Design in Economic Education."

**References:**

Berthold, Kirsten; Eysink, Tessa H. S.; Renkl, Alexander (2009): Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. In: Instr Sci 37 (4), S. 345–363. DOI: 10.1007/s11251-008-9051-z.

Euler, Dieter; Hahn, Angela (2014): Wirtschaftsdidaktik. Stuttgart, Deutschland: utb GmbH.

Fyfe, Emily R.; McNeil, Nicole M.; Son, Ji Y.; Goldstone, Robert L. (2014): Concreteness Fading in Mathematics and Science Instruction: a Systematic Review. In: Educ Psychol Rev 26 (1), S. 9–25. DOI: 10.1007/s10648-014-9249-3.

Fyfe, Emily R.; Nathan, Mitchell J. (2019): Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. In: Educational Review 71 (4), S. 403–422. DOI: 10.1080/00131911.2018.1424116.

Loibl, Katharina; Leuders, Timo; Glogger-Frey, Inga; Rummel, Nikol (2024): CID: a framework for the cognitive analysis of composite instructional designs. In: Instr Sci. DOI: 10.1007/s11251-024-09665-9.

Lee, C. H., Kalyuga, S., & Wales, S. (2014). Expertise reversal effect and its instructional implications. Applying science of learning in education: Infusing psychological science into the curriculum, 31-44.

Rau, Martina A.; Aleven, Vincent; Rummel, Nikol (2015): Successful learning with multiple graphical representations and self-explanation prompts. In: Journal of Educational Psychology 107 (1), S. 30–46. DOI: 10.1037/a0037211.

## Poster Session

October 8, 2025, 17:30-19:00
3.101 & Forum

**P1: Fostering the Revision of Science Explanations Through Critique in Technology-Enhanced Inquiry Learning**

Katharina Bach (*Ludwig-Maximilians-University Munich)*, Sarah Bichler *(Ludwig-Maximilians- University Munich),* Marcia C. Linn *(University of California, Berkeley)*

**Keywords:**
*Inquiry Learning, Revision, Peer Critique, Technology-Enhanced Learning, Science Explanation, Knowledge Integration*

**Abstract:**

Students' intuitive ideas about scientific phenomena often contradict theories taught in school. Technology-enhanced learning environments allow for interactive engagement, supporting students in discovering scientific explanations, for example, through dynamic visualizations (Linn et

al., 2018). Expressing these explanations in their own words and revising them helps students integrate different ideas (Linn & Eylon, 2011). One way to promote substantial revision of explanations is peer critique (Donnelly et al., 2015). We investigated how the characteristics of the critique artifacts (a peer's explanation) impact the integration of ideas during revision of one's own explanation and whether specific designs might work better for learners depending on how well they can already explain the phenomenon.
We used the Web-based Inquiry Science Environment and analyzed how peer critique helps 1305 middle school students revise their science explanations about thermodynamics. We compared three designs of the critique activities that differed in the characteristics of the fictitious peer's ideas. After random assignment to the conditions, students either critiqued a peer's explanation that describes observations (intuitive), combines intuitive and normative ideas (partial), or connects a claim with scientific evidence (link). Afterward, students were asked to improve their explanation. Initial and revised explanations were scored using a knowledge integration (KI) rubric that rewarded linking claims and evidence. A linear mixed model showed that students in the link condition received the highest KI scores in the revision. Moreover, students who started with a higher KI score had an advantage over those with a lower KI score in all conditions (b = 0.77). This difference was strongest in the intuitive condition, indicating that the intuitive condition was the least beneficial for students with lower prior knowledge. Qualitative analyses revealed that many students adopted ideas from the example, regardless of their correctness. This might explain why students in the link condition performed best.
Rather than distinguishing between the ideas in the peer explanation and their own ideas, students often seemed to trust their peers' ideas, even when these were less integrated than their own. Thus, we suggest that a more effective critique activity is comparing two explanations with contrasting ideas and offering (adaptive) guidance that suits students' prior knowledge.

**P2: Supporting Teaching and Learning Through an AI-Enhanced Interactive Chat Activity**

Sarah Bichler *(University of California Berkeley & Ludwig-Maximilians-University Munich),* Allison Bradford, *(University of California Berkeley)*, Kenneth Steimel, *(ETS),* Marcia C Linn, *(University of California Berkeley),* Brian Riordan, *(ETS)*

**Keywords:**
*AI-supported learning, Natural Language Processing, Science Learning, Adaptive instruction, Classroom integration, Mixed-methods*

**Abstract:**

Students encounter science in their everyday lives, using their experiences to make sense of scientific phenomena. Recognizing and integrating students' prior ideas is not only beneficial for learning but also enhances motivation, makes science personally relevant, and shifts teachers and students toward a resource-oriented perspective, one that values and productively builds on students' existing knowledge rather than focusing on correcting misconceptions. However, because students' experiences and prior exposures vary widely, not all individual resources can easily be leveraged in the classroom.

This study examines an AI-supported chat activity embedded within a middle school online inquiry science curriculum on the physics of sound waves. The activity leverages a natural language processing (NLP) model to detect students' specific science ideas within their written explanations and chat conversations. Using predefined teacher-evaluated guidance questions, the AI adaptively provides targeted prompts to help students refine and develop their conceptual understanding.

We investigate (1) how students' science ideas evolve from before to after the chat activity, (2) interaction patterns within the chat activity, and (3) teachers' perspectives on the potential of integrating the chat activity in their instruction. The study was conducted with two science teachers and their 162 middle school students. A mixed-methods approach was employed, including pre- and post-chat written explanations, teacher interviews, and interaction analysis of chat logs.

Preliminary findings suggest that the chat activity helped students elaborate on and refine their initial ideas. Teacher interviews highlight the promise of AI-driven conversations in supporting students individually and providing all students with the opportunity to express their ideas, particularly inviting those who are typically shy to share what they think. Limitations for classroom implementation include the challenge of reliably detecting infrequent ideas, as well as the trade-off inherent in the NLP model: while its specificity to a particular scientific phenomenon and associated prompts enables a targeted analysis of student thinking, this same specificity also limits its broader applicability. Future questions include whether more general language processing models can achieve the same domain-specific depth while being less costly to develop in terms of financial, computational, and time resources.

**P3: The Interplay between Expectancy Beliefs and Task Values in Simulation-Based Learning Environments: An Individual Participant Data Meta-Analysis**

Doris Holzberger *(Technical University of Munich)*, Anika Radkowitsch *(IPN - Leibniz Institute for Science and Mathematics Education)*, Laura Brandl *(LMU Munich)*, Olga Chernikova *(LMU Munich)*, Andreas Obersteiner *(Technical University Munich)*, Amadeus J. Pickal *(Augsburg University)*, Michael Nickl *(LMU Munich)*, Constanze Richters *(LMU Munich)*, Stephanie Kron *(LMU Munich)*, Christof Wecker *(University of Hildesheim*), Nicole Heitzmann *(LMU Munich)*, Marie Irmer *(LMU Munich)*, Caroline Corves *(LMU Munich)*, Birgit Jana Neuhaus *(LMU Munich)*, Martin R. Fischer *(LMU Munich)*, Stefan Ufer *(LMU Munich)*, Frank Fischer *(LMU Munich)*, Tina Seidel *(Technical University Munich)*, Daniel Sommerhoff *(IPN - Leibniz Institute for Science and Mathematics Education)*

**Keywords:**
*Expectancy-Value, Simulation-based Learning, Diagnostic Competence, Simulation-Based Learning, Individual participant Data - Meta-Analysis*

**Abstract:**

This study examines how expectancy beliefs and task values predict diagnostic performance in simulation-based learning environments for prospective teachers and physicians. Using a meta-analytic approach, we analyzed individual participant data from 16 studies (N = 1,492) conducted within a single research unit. Despite the homogeneity in measures and methodologies, measurement invariance analyses revealed two clusters of studies (cluster 1: nine studies, cluster 2: five studies) with differing patterns. For expectancies for success and diagnostic accuracy, the aggregated Fisher's z correlation was .07 ($p = .03$, 95% CI = [.01; .13]) in cluster 1 and -.04 ($p = .63$, 95% CI = [-.29;.22]) in cluster 2. For utility value, the correlations were .14 ($p < .05$; 95% CI = [.02; .26]) and .08 ($p = .10$; 95% CI = [-.01; .18]), respectively. The interaction between expectancy beliefs and utility value showed weak, inconsistent, and non-significant relations with diagnostic accuracy in both clusters. Findings further differed depending on whether the motivational variables were assessed task-specific or more general. No differences emerged between teacher and medical education contexts. These findings highlight the context-dependence of motivational processes and suggest that variations may stem from how constructs are operationalized. The results emphasize the need

to tailor interventions to specific learning contexts and caution against generalizing from single studies. Overall, our study underscores the situated nature of expectancy-value frameworks and the importance of multi-study syntheses in understanding its role in professional education.

**P4: Affective Experience and Learning Behavior in Digital Learning: Linking Introspection and Learning Process Data**

Julia Hilpert *(Karlsruhe University of Education)*, & Marc Philipp Janson *(Karlsruhe University of Education)*

**Keywords:**
*affective experience, self-regulated learning, intelligent tutoring system*

**Abstract:**

Intraindividual perspectives on psychological processes are receiving increasing attention in educational research (Murayama et al., 2017), including learners' emotion and affect (Dietrich et al., 2022), which are central to self-regulated learning (e.g. Efklides, 2011). Digital learning environments enable situational, objective assessments of self-regulated learning behavior, but often lack a theory-driven approach based on psychological processes (Baker et al., 2020).

This contribution addresses this gap by linking objective learning process data from an intelligent tutoring system (ITS) with experience sampling on learners' affective states. Two studies examined intraindividual differences in affect on a daily (Study 1) and item (Study 2) level and the relation to students' learning behavior.

Study 1, based on Control-Value Theory (Pekrun et al., 2023), examined how expectancy of success (as control appraisal), value and cost appraisals and affective experiences relate to planned and actual learning behavior. We analyzed 1,075 daily self-reports and ITS logfile data from 95 students. Multilevel analyses (L1: day, L2: person) revealed that higher expectancy ($b = 0.38$, $p < .001$) and value appraisals ($b = 0.30$, $p < .001$) were associated with more positive affect, while high costs predicted negative affect ($b = -0.25$, $p < .001$). Expectancy ($b = 2.95$, $p = .034$) and value appraisals ($b = 6.82$, $p < .001$) were also positively related to students' motivation to learn, but only costs predicted both planned ($b = 3.25$, $p = .017$) and actual learning ($b = 8.95$, $p < .001$).

Study 2 examined whether recent learning success influenced learners' momentary affective state. Specifically, we tested if the proportion of correctly solved ITS items predicted valence and arousal (Pekrun et al., 2023), assessed every seven items. In 2,144 assessments from 60 students (L1: assessment, L2: person) higher learning success was linked to higher valence ($b = 21.41$, $p < .001$) and lower arousal ($b = 4.87$, $p = .006$), but affect did not predict subsequent learning behavior.

The studies provide insights into the interplay of affective experience and digital learning behavior. They reveal substantial intraindividual variability in affect and self-regulated learning dependent on the learning process. While it remains unclear whether affect drives or follows learning behavior, combining objective learning process data and experience sampling holds promise for theory-driven learning analytics.

**P5: Human-Machine Improvisation Processes in AI-Augmented Music Education**

Florian Öttl *(Karlsruhe University of Education),* Kai Koch *(Karlsruhe University of Education),* Marcus Syring *(Eberhard Karls University Tübingen)*

**Abstract:**

The integration of Generative Artificial Intelligence (GenAI) in the field of music education not only offers great potential for planning and evaluating music lessons but also enables new pathways in all areas of musical learning (Krämer & Hecht, 2024). GenAI-based learning arrangements provide space for individualized approaches where the traditional roles of learners and teachers change fundamentally (Bade, 2023; Holster, 2024). Additionally, the dialogic structure of human-AI-interaction makes it possible to use GenAI-tools as collaborators in musical improvisation, as they are able to "create real-time accompaniment or improvised responses to learners' playing" (Yue & Jing, 2025, p. 143). Arising questions about the role of human creativity in such AI-based improvisation processes can be framed by the concept of a Human-Machine "co-creativity" (Assayag, 2021; Fiorini, 2025; O'Toole & Horvát, 2024). Although the integration of GenAI in music education contexts offer innovative perspectives on musical learning, AI-collaborators are facing the "Habitus of the Machine" (Treß, 2025), including biases that may reproduce stereotypic structures and data security-issues.

This research project aims to investigate the following questions:
(1) What influence does the use of Generative Artificial Intelligence have on learners' musical improvisation? What interaction patterns can be differentiated? Can different types of co-creativity be observed?
(2) What might design tasks look like that promote musical improvisation using GenAI?
It is planned to conduct a Design-Based-Research (Prediger et al., 2012) with a first cycle investigating learners' approaches in human-machine improvisation processes and examining different types of co-creativity and a second cycle developing different tasks to foster learners' AI-augmented musical improvisation.

Assayag, G. (2021). Human-Machine Co-Creativity. In B. Lubat, G. Assayag, & M. Chemillier (Eds), *Artisticiel / Cyber-Improvisations*. Dialogiques d'Uzeste. https://hal.science/hal-03542986

Bade, F. (2023). Rollen und Potenziale von ChatGPT1 in musikpädagogischen Kontexten. *Diskussion Musikpädagogik*, *100/23*, 48–57.

Fiorini, M. (2025). Being the Artificial Player: Good Practices in Collective Human-Machine Music Improvisation. In A. L. Brooks, D. Banakou, & S. Ceperkovic (Eds), *ArtsIT, Interactivity and Game Creation* (Vol. 650, pp. 310–330). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97254-6_22

Holster, J. (2024). Augmenting Music Education through AI: Practical Applications of ChatGPT. *Music Educators Journal*, *110*(4), 36–42. https://doi.org/10.1177/00274321241255938

Krämer, O., & Hecht, B. (2024). *Anwendungsszenarien von KI im Musikunterricht*. https://doi.org/10.13140/RG.2.2.32590.06722

O'Toole, K., & Horvát, E.-Á. (2024). Extending human creativity with AI. *Journal of Creativity*, *34*(2), 100080. https://doi.org/10.1016/j.yjoc.2024.100080

Prediger, S., Link, M., Hinz, R., Hußmann, S., Thiele, J., & Ralle, B. (2012). *Lehr-Lernprozesse initiieren und erforschen: Fachdidaktische Entwicklungsforschung im Dortmunder Modell*. https://wwwold.mathematik.tu-dortmund.de/~prediger/veroeff/12-Prediger_et_al_MNU_FUNKEN_Webversion.pdf

Treß, J. (2025). Tracking Down the Musical Habitus of the Machine. *Action, Criticism, and Theory for Music Education*, *24*(3), 79–108.

Yue, Y., & Jing, Y. (2025). Artificial Intelligence in Music Education: Exploring Applications, Benefits, and Challenges. *2025 14th International Conference on Educational and Information Technology (ICEIT)*, 141–146. https://doi.org/10.1109/ICEIT64364.2025.10976152

**P6: AI-Based Feedback for Argumentation Learning in Teacher Education**

Antonia Ohliger *(RPTU University Kaiserslautern-Landau)*, Miriam Leuchter *(RPTU University Kaiserslautern-Landau)*

**Keywords:**
*Argumentation Skills, Artificial Intelligence, Feedback, Teacher Education*

**Abstract:**

Given the complex and multi-layered structure of teaching, argumentation skills (Toulmin, 2003) provide an essential foundation for reflective decision-making and evidence-based lesson planning (Nückles & Schuba, 2019). Despite their importance, many teachers struggle with argumentation (Lytzerinou & Iordanou, 2020), partly because argumentation learning is often overlooked in teacher education (Topi, 2018) and limited by a lack of individualized learning opportunities (Santhanam et al., 2016). Moreover, arguments are often prone to fallacies, where premises do not adequately support claims (Walton, 2004).

Therefore, fostering argumentation skills should be more systematically integrated into teacher education. Building on findings that AI-based feedback can effectively support the improvement of such skills (Neshaei et al., 2025), we developed an AI-based feedback tool. In this study, we ask whether AI-based feedback enhances pre-service-teachers' argumentation skills.

250 pre-service teachers participated in a 6-week online course. Their ability to recognize argument structures and argumentation fallacies was assessed pre-post in weeks 1 and 6 using A-FaST (Berkle et al., 2023). In weeks 2 to 5, students wrote argumentations based on job-relevant video scenarios. The experimental group (EG; N=134) received AI-based feedback on their written argumentations, while the control group (CG; N=116) did not. All participants were given the opportunity to revise their argumentations.

Both groups improved significantly in recognizing argumentation fallacies. However, within the EG, cluster analysis revealed that N=58 requested feedback once per argumentation task (minimum required) and made no revisions. N=74 requested feedback more frequently (Max=37) and made revisions (Min Jaccard Similarity=0.54). Results showed that higher initial test performance predicted higher feedback usage, resulting in a significant improvement in fallacy recognition compared to EG students without feedback usage.

These findings highlight the potential of AI-based feedback to enhance pre-service teachers' argumentation skills by providing scalable and individualized support. However, its effectiveness strongly depends on students' willingness to engage with the feedback provided (Chan & Hu, 2023). This study emphasizes the need for strategies that promote pre-service teachers' engagement to ensure the benefits of AI-based feedback on argumentation learning within teacher education.

**P7: Fostering Preservice Teachers' Values and Expectancies for Teaching CT Through a Programming Intervention**

Dennis Rudolf Pouhe *(RPTU University Kaiserslautern-Landau)*, Veronika Barkela *(RPTU University Kaiserslautern-Landau)*, Miriam Leuchter *(RPTU University Kaiserslautern-Landau)*

**Keywords:**
*computational thinking, expectancy-value theory, preservice teachers, Scratch, computer-based learning environment*

**Abstract:**

The integration of computational thinking (CT) into primary education is considered increasingly relevant, particularly in light of the growing societal importance of digital technologies (Fraillon et al., 2020). CT encompasses skills essential for programming, such as logical reasoning and algorithmic design (Angeli et al., 2016). To prepare preservice teachers for teaching CT, it is essential that they acquire the necessary skills themselves. However, preservice teachers often feel reluctant toward engaging with and teaching CT (Barkela et al., 2024). According to expectancy-value theory, learning success can positively influence preservice teachers' values and expectancies (Weber et al., 2022), thereby potentially increasing their behavioral intention to teach CT. Based on this, we asked:

Do preservice teachers change their values and expectancies to integrate programming into primary school when attending an eight-week intervention?

We piloted an intervention that was built around a Scratch-based learning environment (https://scratch.mit.edu), enriched with scaffolds, like semantically structured instructions, video-based demonstrations of solutions and starter files to support the creation of correct algorithms. A total of N=103 preservice teachers participated, and values and expectancies for teaching CT were measured before and after the course.

At the beginning, preservice teachers reported medium cost and above-average values related to teaching CT, however, they reported limited self-efficacy. Paired t-tests revealed significant decreases from pre- to post-test in preservice teachers' intrinsic, utility, and attainment values while no significant changes were found for cost and self-efficacy.

The findings suggest that preservice teachers are generally open to teaching CT. However, the intervention reduced rather than enhanced their values related to teaching it. It appears that the scaffolds were insufficient to support successful task completion, which may have led to frustration, overwhelm, or disengagement instead of fostering positive learning experiences, ultimately contributing to the decline in values. In a follow-up study, we will revise the computer-based learning environment by incorporating stronger, structure-oriented scaffolds, such as modelling and a more focused presentation of programming options, to better support learners. At the conference, we will present the computer-based learning environment and discuss the results in detail.

**P8: When Does Learning by Non-Interactive Teaching Work? A Large-Scale Analysis of Learner Characteristics in a Classroom Setting**

Heike Russ *(University of Tübingen)*, Andreas Lachner *(University of Tübingen),* Nicolas Hübner *(University of Bonn)*, Leonie Sibley *(Zurich University of Teacher Education),* Katharina Scheiter *(University of Potsdam)*

**Keywords:**
*Non-interactive teaching, Generative learning, Aptitude-treatment interaction, Learner characteristics, Secondary physics education*

**Abstract:**

Background: Non-interactive teaching, in which students explain previously learned contents to a non-present peer, is a generative learning activity that has gained increasing attention in recent years. While meta-analyses indicate small-to-moderate benefits, findings have been inconsistent, suggesting that its effectiveness depends on contextual factors.
Aims: Drawing on the aptitude-treatment interaction framework, this study examines how learner characteristics—specifically (meta-)cognitive, motivational, personality, and demographic prerequisites—moderate the effects of non-interactive teaching on immediate and lasting learning outcomes.
Sample: By adopting a one-stage individual participant meta-analytic approach, we synthesized data from three analogous classroom experiments (N = 1,074) in secondary physics education.
Methods: Students were randomly assigned to either explain the contents to a non-present peer via instant messaging (non-interactive teaching) or to restudy the materials. Scientific knowledge and monitoring accuracy were assessed immediately after the intervention and again after eight weeks. Learner characteristics covering cognitive, metacognitive, motivational, personality-related, and demographic prerequisites were assessed prior to the intervention and tested simultaneously as moderators, allowing for estimation of each effect while controlling for potential confounding factors.
Results: Results showed that non-interactive teaching resulted in higher immediate scientific knowledge and stronger under-confidence compared to restudy but did not enhance long-term retention. However, the effectiveness of non-interactive teaching varied by learner characteristics: Immediate monitoring accuracy depended on language proficiency. Long-term retention was moderated by students' interest in physics, as low- and medium-interest students, but not high-interest students, demonstrated superior performance eight weeks after the intervention. This effect was explained by increased mental effort allocation.
Discussion: The findings highlight that the effectiveness of non-interactive teaching depends on individual learner characteristics, particularly language proficiency and subject-specific interest. This suggests that tailoring instructional activities to student profiles may enhance learning outcomes and supports more nuanced applications of aptitude-treatment interaction models in classroom settings.

**P9: The Influence of Perceived Control, Perceived Value, and Enjoyment on Self-Regulated Learning from Text**

Celina Safferthal *(Karlsruhe University of Education)*, Anja Prinz-Weiß *(Karlsruhe University of Education)*

**Keywords:**
*Enjoyment, judgment accuracy, metacognition, regulation effectiveness, self-regulated learning, text comprehension*

**Abstract:**

Everyone knows the feeling of reading a text and experiencing enjoyment. But how does enjoyment while reading influence self-regulated learning from text? Knowledge is often acquired through self-regulated learning from text. To succeed, learners must accurately judge their comprehension and develop strategies to improve it. Although the positive impact of enjoyment on performance is well documented (e.g., Zaccoletti et al., 2020), little is known about how achievement-related emotions serve as cues (e.g., Koriat, 1997) in metacognitive processes—especially regarding the accuracy of judgments and subsequent regulation of learning. A recent correlational study (Prinz-Weiß et al., 2022) found that higher enjoyment or hope was linked to an overestimation of comprehension. However, it only examined the link between achievement emotions and metacognitive judgments, not regulation processes or outcomes. In this study (N = 324), we experimentally induced enjoyment to explore its effects on text comprehension, the accuracy of judgments, and regulation for better understanding. The experimental group received an enjoyment-enhancing intervention; the control group received a neutral one. In line with Pekrun's (2006) Control-Value Theory, students' enjoyment was increased by promoting value and control beliefs about the text's topic ($CO_2$ emissions). Three hypotheses were tested: Compared to the control group, the enjoyment group would show H1) better comprehension before regulation, H2) greater overestimation of comprehension, H3) smaller improvement after regulation. Inducing enjoyment was not effective, but perceived control increased significantly. The experimental group had better pre-regulation comprehension (small effect), but did not overestimate more or regulate less effectively than the control group. Replication is needed to draw valid conclusions about enjoyment's role in judgment accuracy and regulatory effectiveness.

**P10: Signals of Struggle: Using Digital Self-Assessments to Detect At-Risk Students**

Zoe Maj Sander *(Heidelberg University)*, Vivien Rieder *(Heidelberg University)*, Birgit Spinath *(Heidelberg University)*

**Keywords:**
*Digital Self-Assessment, Motivational Development, At-Risk Students, Disengagement*

**Abstract:**

Mandatory courses in higher education frequently experience high dropout rates, with many students requiring multiple attempts to complete them successfully. This contributes to student frustration, extended time to graduate, and inefficient allocation of resources (McKinney et al., 2019). In addition, a significant number of students remain formally enrolled but disengage – failing to adopt self-regulated learning strategies and consequently achieving poor outcomes (Bosch et al., 2021). Previous studies identified motivation and learning satisfaction as predictors of dropout and disengagement (Bosch et al., 2021; Dresel & Grassinger, 2013; Scheunemann et al., 2022). However,

motivation is not static but tends to decline during the first weeks of the term before stabilizing, though not uniformly across all students (Benden & Lauermann, 2022; Sander et al., 2024). Prior research emphasizes heterogeneous motivational trajectories, with a substantial subgroup falling into unfavorable profiles showing low motivational baselines and/or steep declines (Musu-Gilette et al., 2015; Robinson et al., 2022). While summative assessment and evaluation fail to detect these issues early, digital formative (self-)assessments offer a promising approach to monitoring critical developments in a scalable and accessible manner (Johnson & Davies, 2014). In this study, we aim to identify early signals of dropout intention and disengagement based on changes in motivation and satisfaction using digital self-testing. Therefore, we examine longitudinal data from 883 teacher students in an educational psychology lecture course. Based on the theoretical framework of Situated Expectancy-Value Theory (Eccles & Wigfield, 2020), we assessed students' expectancies and task values at 5 measurement points (t1-t5) across one academic term. Furthermore, students reported their satisfaction and dropout intention starting from t2. We operationalized disengagement as ceased participation in digital self-testing. Panel logistic regression models with lagged predictors will be conducted once data collection is completed. The potential of digital formative (self-)assessment to detect and address maladaptive motivational processes during the academic term will be discussed.

**P11: EffecTive Project: Enhancing Pedagogical Digital Competence to Transform Teaching and Learning**

Tina Seufert *(Ulm University)*, Tabea Rosenkranz *(Ulm University)*, Aileen Herold *(Ulm University)*, Anita Radi-Pentz *(Ulm University)*

**Keywords:**
*Pedagogical Digital Competence, Teacher Professional Development, Self-Regulated Learning, Technology-Enhanced Learning, Instructional Design*

**Abstract:**

This poster presents the EffecTive project, a Horizon Europe initiative aimed at strengthening Pedagogical Digital Competence (PDC) among teachers across Europe. Grounded in competence development theory (Desimone, 2009; Blömeke et al., 2015), the project addresses a central research question: Which training approaches most effectively support teachers in developing PDC and, in turn, promote student self-regulated learning in digitally enriched environments? It builds on the assumption that teachers' situation-specific digital-pedagogical skills are key to creating inclusive, engaging learning experiences and improving student outcomes such as cognitive engagement and self-regulation.
To investigate this, the project implements and evaluates twelve professional development interventions across Austria, Estonia, Finland, Germany, and Israel. A mixed-methods research design underpins the study, combining pre-post quantitative measures with qualitative insights. The interventions vary in intensity and instructional design—ranging from theory-based knowledge instruction to collaborative design, situated learning, and personalized coaching. All training formats aim to bridge the gap between knowledge acquisition and practical application in classroom settings. The resulting data will be integrated through a meta-analysis to identify which components are most effective in fostering changes in teacher practice and student learning. While the full analysis is ongoing, initial results from the pilot phase suggest that collaborative and situated learning approaches show strong potential in enabling transfer to classroom practice. Additionally, preliminary data indicate that teachers' own self-regulated learning skills significantly moderate training effects, particularly in the development of situation-specific teaching strategies

and in fostering self-regulated learning among students. More intensive interventions appear to yield greater benefits in teacher confidence and instructional quality, despite requiring higher investment.
By connecting theoretical models with empirical findings from cross-national contexts, the EffecTive project provides evidence-informed insights for designing impactful and scalable PDC training. The poster illustrates how carefully structured professional development can lead to improved teaching quality, student autonomy, and more equitable access to high-quality digital education across Europe.

**P12: Learning Relational Categories: Effects of Combining Classification Tasks and Generation Tasks on Learning and Metacognitive Judgments**

Tim M. Steininger *(University of Freiburg)*, Thamar Voss *(University of Freiburg)*, Jörg Wittwer *(University of Freiburg)*

**Keywords:**
*category learning, relational categories, generating examples, classifying examples, metacognitive judgments, learning tasks*

**Abstract:**

Previous research provides evidence that learning relational categories differs from learning feature-based categories which has been predominantly examined in research on inductive category learning. Usually, categories are learned using examples. In particular, the learning of relational categories, which in contrast to feature-based categories cannot be recognized on the basis of superficial characteristics, requires a well-thought-out example design. For instance, tasks might ask learners to read provided examples with the accompanying category labels, to generate their own examples for the categories, or to classify provided examples into categories. In three experiments on learning relational category in teacher education, we revealed positive effects of classification tasks over reading tasks and generation tasks. In contrast, generation tasks were prone to have detrimental effects on learning relational categories in teacher education in terms of learning and metacognitive judgment accuracy. From the experiments it can be derived that to learn relational categories, a classification task compared with reading tasks or generation tasks improves learning by providing a better equilibrium of a constructive learning activity combined with an adequate level of instructional guidance. While these results provide valuable insights that can help designing effective learning environments, various learning task designs can be conceived. In a present study, we investigate if a combination of classification tasks and generation tasks can be equally or even more effective than using only classification tasks. Working on classification tasks before generating own examples might provide enough instructional guidance to enable learners to generate examples in a way that foster their learning of relational categories. Besides effects on learning, as indicated by the previous studies, we also examine effects on the accuracy of learners' metacognitive judgments. Currently, we are conducting the experiment. Hence, data is pending, but will be available when the AQUA-d Conference 2025 will take place. The results will provide additional information for research and for educators on effective learning tasks when teaching relational categories.

**P13: Accompanying Transfer Research in the Domains of Linguistic-Rhetorical Communication and Literary Writing Within the Initiative Leistung macht Schule**

Jenny Winterscheid *(Karlsruhe University of Education)*, Beate Laudenberg *(Karlsruhe University of Education)*

**Keywords:**
*Promoting Language Skills, Literary Writing, Rhetorical Communication, Transfer Research, Multiplication, Scientific Support*

**Abstract:**

As one of three research teams in the language content cluster (IC 4), the Laudenberg team at the Karlsruhe University of Education supports teachers and state representatives in LemaS-Transfer in passing on the concepts and materials developed in the first phase of the research and development project Leistung macht Schule (LemaS; 2018-2023) in their respective networks. In line with its focus, the Laudenberg team is researching the transfer of concepts and products to promote linguistic and rhetorical communication and literary writing (e.g. Mayer et al. 2024 or Laudenberg/Spiegel 2020). In addition to academic support for the multipliers through regular workshops and forums for exchanging experiences and best practice examples of promoting giftedness at both the mediation and teaching level, the conditions for successful transfer are being researched. To this end, the results of a longitudinal, guideline-based interview study from three content clusters (in addition to IC 4, IC 2 on interdisciplinary teaching/school development and IC 3 on STEM) are directly incorporated into the provision and adaptation of the materials, but also provide a deeper insight into transfer processes and findings regarding the sustainable implementation of transfer concepts (Lipowsky, 2014), which can promote domain-specific gifted education in the subject German.

**References:**

Laudenberg, B./Spiegel, C. (2020): Förderung sprachlicher Expertise am Beispiel des Schreibens und Kommunizierens. In: Lehren und Lernen 46. Jg., H. 8/9, 22-26.

Lipowsky, F. (2014): Theoretische Perspektiven und empirische Befunde zur Wirksamkeit von Lehrerfort- und weiterbildung. In: Terhart, E./Bennewitz, H./Rothland, M. (Hrsg.): Handbuch der Forschung zum Lehrerberuf. Münster/New York: Waxmann, S. 511-541.

Mayer, J./Laudenberg, B./Mempel, C./Winterscheid, J./Zuber, J. (2024): Sprachlich-literarisches Lernen begabungs- und begabtenförderlich gestalten (Symposium). In: Rogl, S./Resch, C./Bögl, E./Gürtler, B./Hinterplattner, S./Klug, J. (Hrsg.): Begabung verändert – förderliche Lernwelten erforschen, gestalten, implementieren. Münster/New York: Waxmann, 394-419. (= Bd. 17; Begabungsförderung: Individuelle Förderung und Inklusive Bildung hrsg. v. C. Fischer)

**P14: The Interplay of Perceived Mental Effort and Judgment of Learning in Self-Regulated Learning within a Digital Tutoring System**

Samuel Wissel *(University of Mannheim)*, Marc Philipp Janson *(Karlsruhe University of Education)*, Stefan Münzer *(University of Mannheim)*

**Keywords:**
*Self-regulated Learning, Digital Tutoring System, Metacognition, Momentary Assessment*

**Abstract:**

Monitoring one's learning process is a crucial component in self-regulated learning (SRL). Judgements of learning (JOL; Rhodes, 2016) might be based on indirect cues (Koriat, 1997), such as perceived mental effort (PME; Paas et al., 1994). Previous research has confirmed the assumption that higher PME is related to lower JOL and lower learning outcomes, whereas JOL are positively related to learning outcomes (David et al., 2024). We pre-registered a mediation model in which JOL serve as mediator between PME and learning outcome. A negative association between PME and learning outcomes, a positive relation between JOL and learning outcomes, and a negative relation between PME and JOL were assumed. The same model was stated for learning times. The present study aims to extend the literature by providing evidence from a real-world setting with a digital tutoring system (DTS) used by university students.
95 teacher students utilized a DTS for exam preparation. The DTS presented questions repeatedly and prompted students to provide JOL and PME ratings for every seventh question. PME and JOL ratings were linked to the next occurrence of that question. Learning time was the time to solve the question. An average student provided 94.14[SD=136.26] PMEs and JOL. In the multi-level-analyses, we controlled for correctness of responses to the questions for which PME and JOL were provided. For learning success, we found a negative association between PME and learning outcome and a positive between JOL and learning outcome, however, PME was positively associated with JOL (ps < .01). Furthermore, a positive effect of PME via JOL (as mediator) was found (indirect effect: ß = .025, SE = .002, p < .001), potentially showing a suppression effect. For learning times, corresponding effects were found (indirect effect: ß = -.015, SE = .002, p < .001). However, JOL and learning times were negatively and PME and learning times were positively related (ps < .01).
The complex association between PME and outcomes indicate effort to be a cue for perceived learning and hindrances at the same time. Surprisingly, PME was positively related to JOL. An explanation could be that in this real-world learning context, students followed goal-driven rather than data-driven strategies (Baars et al., 2020). The negative relation between JOL and learning time might be explained by the diminishing criterion model (Ackerman, 2014) which predicts decreasing time investment with increasing confidence.

## S5: Strategies for Enhancing Performance with Perceptually Rich Displays

**Presenter:** Killyam Forge *(CLLE, CNRS & University of Toulouse Jean Jaurès)*, Manuela Glaser *(Leibniz-Institut für Wissensmedien, Tübingen)*, Robin Wagner *(Leibniz-Institut für Wissensmedien, Tübingen),* Patricia Engel-Hermann *(Karlsruhe University of Education)*
**Chair:** Alexander Skulmowski *(Karlsruhe University of Education)*
**Discutant:** Jean-Michel Boucheix (*University of Dijon*)

October 9, 2025, 10:30-12:30
3.101

Realistic visualizations play a major role for learning in schools as well as in a wide range of learning scenarios that require to visualize content in an authentic and informative way, such as museum exhibitions. To harness the power of such perceptually rich visualizations, strategies to overcome some of their drawbacks are a relevant topic in this area of research. For instance, how should qualitative differences of the information in visualizations be made transparent to the learner in order to enhance understanding? Several strategies, ranging from marking the relevance and reliability of elements in perceptually rich visualizations to utilizing virtual reality as a form of pre-training for learning scenarios will be presented and discussed in this session.

**Interactivity and Cueing in the Context of Realistic Visualizations**

Patricia Engel-Hermann *(Karlsruhe University of Education), &* Alexander Skulmowski *(Karlsruhe University of Education)*

**Abstract:**

Detailed visualizations are utilized in many school subjects. In biology, for instance, detailed visualizations are often used as a realistic representation of an object precluded from being easily accessed in the real world, such as cell structures or anatomical parts of the human body. However, detailed visualizations require learners to cope with the perceptual richness induced by the visualization. Cognitive load theory suggests to enhance learning performance through an optimized utilization of learners cognitive capacities, which could be attempted by reducing the visual complexity of the material or through various design strategies which have developed into „principles" in the context of multimedia learning.
The present study evaluates the transferability of two of these strategies for learning with a perceptually rich visualization, namely cueing and interactivity. In a laboratory experiment with university students, participants were provided either with the possibility to slightly rotate a perceptually rich visualization of the parotid gland, with a version of the visualization featuring visual cues in the form of outlines of its components, a combination of these design factors or a static visualization without cues. Cueing and interactivity were hypothesized to support learning performance through a reduction of unnecessary task demands. Moreover, we expected cueing to particularly positively affect learning with the interactive visualization. Hypotheses were formulated on the basis of cognitive load theory and subjective measures of cognitive load were collected to evaluate the findings accordingly.
An interaction effect of interactivity and visual cues on retention performance has been found. Constrained interactivity (i.e., possibility to slightly rotate the visualization) raised retention performance in the absence of visual cues. Interestingly, ratings of germane cognitive load (i.e., the beneficial allocation of mental resources towards the task) were higher for the interactive version featuring visual cues.

The findings highlight the complex interplay between instructional strategies and perceptually rich visualizations and raise questions regarding the assessment of cognitive load in research on learning with perceptually rich visualizations.

**Investigating Self-Controlled Signaling in Instructional Animations for Learners with Different Prior Knowledge Levels: An Eye-Tracking Study**

Killyam Forge, *(CLLE, CNRS & University of Toulouse Jean Jaurès)*, Lemarié, J, *(CLLE, CNRS & University of Toulouse Jean Jaurès, France,* Gaillard de Saint Germain, P., *(CLLE, CNRS & University of Toulouse Jean Jaurès),* Paubel, P. V., *(CLLE, CNRS & University of Toulouse Jean Jaurès),* Boucheix, J. M., *(LEAD, CNRS & University of Dijon)*

**Keywords**:
Psychology, Multimedia Learning, Comprehension of Graphic, Signaling principle, Eye-Tracking

**Abstract:**

Instructional animations are visual media designed to explicitly represent movement, trajectories, and transformations in space and time (Ploetzner et al., 2020). However, the transient

nature of information limits their advantages for learning by imposing a significant external cognitive load on learners with low prior knowledge (LPK) that may exceed their processing capacity (Kriz & Hegarty, 2007), making it difficult for them to select the most relevant information (Castro-Alonso et al., 2019). While integrating signals that highlight relevant information in animations enhances LPK learners' performance, scientific literature shows that learners with high prior knowledge (HPK) do not benefit from signaling or may even be hindered by it, as it compels them to process redundant information with their mental schemas (Richter et al., 2016; Alpizar et al., 2020). To reduce this reversal effect, one idea is to allow learners to control the presence of signals in the animation based on their attentional guidance needs. The goal of this research is to examine how prior knowledge and learners' control of signals jointly affect learning performance and cognitive load when studying an animation about the piano mechanism (Boucheix & Lowe, 2010).

Ninety-seven students participated in this eye-tracking study: sixty-seven psychology students (LPK group) and thirty mechanical engineering students (HPK group). Participants learned how the piano mechanism works from an animation taken from Boucheix and Lowe (2010) which was presented under two inter-group conditions: (a) a version in which participants can choose to display or hide the signals before each viewing (controlled signaling condition), and (b) a version in which signals are embedded in the dynamic content without the possibility of being removed by the learner (imposed signaling condition).

Preliminary results indicate that HPK learners had significantly higher comprehension scores and developed a more complete mental model of the mechanism than LPK learners, but only when they controlled the signaling. In contrast, when signaling is imposed by the system, HPK learners performed comparably to LPK learners. The presentation will explore these findings through an analysis of eye-tracking paths and pupil diameter, which serve as objective indicators of learners' cognitive strategies and the cognitive load they experience during learning (Zu et al., 2019). These insights will contribute to a better understanding of the role of learner-controlled signaling in dynamic visualizations and their potential implications for optimizing their use in educational contexts.

**References:**

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. Educational Technology Research and Development, 68, 2095-2119. https://doi.org/10.1007/s11423-020-09748-

Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. Learning and Instruction, 20(2), 123-135. https://doi.org/10.1016/j.learninstruc.2009.02.015

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. Visuospatial processing for education in health and natural sciences, 111-143. https://doi.org/10.1007/978-3-030-20969-8_5

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. International Journal of Human-Computer Studies, 65(11), 911-930. https://doi.org/10.1016/j.ijhcs.2007.06.005

Ploetzner, R., Berney, S., & Bétrancourt, M. (2020). A review of learning demands in instructional animations: The educational effectiveness of animations unfolds if the features of change need to be learned. Journal of Computer Assisted Learning, 36(6), 838-860. https://doi.org/10.1111/jcal.12476

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. Educational Research Review, 17, 19-36. https://doi.org/10.1016/j.edurev.2015.12.003

Zu, T., Hutson, J., Loschky, L., & Rebello, N. S. (2019). Using Eye Movements to Measure Intrinsic, Extraneous, and Germane Load in a Multimedia Learning Environment. Journal of Educational Psychology,112(7), 1338. https://doi.org/10.1037/edu0000441

**Influence of Perceived Trustworthiness of Information Sources on the Cognitive Processing of Uncertain Information in Archaeological Reconstructions**

Manuela Glaser *(Leibniz-Institut für Wissensmedien, Tübingen)*

In history lessons and also during class excursions to history museums or memorial sites, pupils encounter reconstructions of historical places and events that naturally contain uncertain information. It is necessary to make this recognizable for the learners and to provide them with the sources on which the reconstructions are based. Research on the cognitive processing of uncertain information (e.g., Glaser et al. 2022) has shown that information, their uncertainties, and the justifications of these uncertainties are better remembered from certain than from uncertain information. These certainty biases may be influenced by the trustworthiness of the information source and the thereof resulting perceived plausibility of the information. This is investigated using a video of an archaeological reconstruction of the Athena sanctuary in Pergamon. The following hypotheses are postulated:

H1: Memory of the appearance of archaeological elements, their uncertainties, and the justifications of these uncertainties should be better for certain than for uncertain archaeological elements (certainty biases).

H2: Memory of the appearance of archaeological elements, their uncertainties, and the justifications of these uncertainties should be better with an information source of high compared to low trustworthiness.

H3: The certainty biases (H1) should be more pronounced for an information source with high compared to low trustworthiness.

H4: The influence of the source's trustworthiness on memory (H2) should be mediated by perceived plausibility of the appearance of the architectural elements.

A sample of $n$ = 98 participants recruited via Prolific and without prior knowledge of archaeology, history and architecture are examined. They are presented with a multimedia learning environment on the ancient city of Pergamon. Half of them are informed in advance that the learning environment was created based on information from a renowned archaeologist and the other half on the basis of an archaeologist who has been accused of falsifying data. At the end of the learning environment, the participants watch a video that shows a reconstruction of the Athena sanctuary and describes twelve architectural elements, their uncertainty (6 uncertain, 6 certain), and the justifications of these uncertainties. After a subsequent intermediate task, the memory of the appearance of the architectural elements (cued recall test), perceived plausibility of the architectural elements (Likert scales), the memory of the uncertainties of the appearances of the architectural elements (single choice test), and the justifications for these uncertainties (cued recall test) are measured. The data collection and analysis should be completed and results available by the conference.

**From Reality to Virtuality or Vice Versa: Can Virtual 360° Museum Experiences Complement or Replace the Real Visit?**

Robin Wagner, *(Leibniz-Institut für Wissensmedien),* Luisa Scherer, *(Leibniz-Institut für Wissensmedien, Eberhard Karls Universität Tübingen),* Georg Pardi, *(Leibniz-Institut für Wissensmedien),* Peter Gerjets, *(Leibniz-Institut für Wissensmedien),* Birgit Brucker, *(Leibniz-Institut für Wissensmedien)*

**Abstract:**

Immersive 360° recordings of informal learning environments, such as museum exhibitions, are becoming increasingly realistic and allow people to experience this content in locations other than in the actual museum setting. This prompts the question of whether these recordings can serve as a substitute or as a valuable supplement to real in-person visits to the respective exhibition. The present experimental study aims to examine the impact of a virtual 360° museum experience in the Palaeontological Collection in Tübingen (i.e., a 360°-photo of an exhibition room with an accompanying audio guide on all objects in the room) presented through head-mounted displays (HMDs) compared to the actual museum visit. Each participant experienced the same exhibition twice in the two different modalities (within-subject: real visit vs. virtual visit). The sequence of both visits was varied (between-subjects: real-virtual vs. virtual-real) to assess the effect of presentation order. This approach addresses the question whether virtual reality applications are more effective as preparatory tools or for post-visit reflection following real field trips to informal learning environments. The study will test hypotheses regarding the influence of the visit modality and of their presentation order on visitors' overall experience, learning, and content perception. In particular, we expect a beneficial effect of engaging with a museum exhibition in a physical setting over experiencing the same exhibition through a 360° recording viewed with an HMD on variables, such as motivation, perceived impressiveness and learning outcomes. Furthermore, we expect improvements regarding these measured variables for participants who have previously visited the real exhibition room in the first round compared to participants encountering the environment for the first time in the virtual format. The study is currently under experimentation and the results will be presented at the conference. The aim of this research is to evaluate how closely a 360° virtual museum exhibition replicates the experience of a real visit and to identify optimal presentation orders of virtual and in-person visits, thereby providing informed guidance for the integration of virtual tours in educational contexts.

## S6: Profiling Writing Tasks in Terms of Literary and Media Didactics (German)

**Presenters:** Afra Sturm (*University of Applied Sciences and Arts Northwestern Switzerland)*, Johannes Wild *(University of Regensburg)*, Daniela Matz *(University of Tübingen)*, Maximilian Stoller *(Karlsruhe University of Education)*

October 9, 2025, 10:30-12:30
3.107

The symposium addresses the challenges and potential of digital writing in various contexts, bringing together current scientific perspectives from writing didactics, literary didactics, and media didactics. A key focus is on analyzing the dimensions of digital writing, particularly regarding narrative text production and its assessment. Literary writing is examined from a literary didactic perspective using the example of tall tales, highlighting their didactic potential. Additionally, the

method of digital storytelling is presented as a multimodal narrative approach and linked to empirical research findings on the promotion of narrative text production competence. The symposium provides an interdisciplinary platform for scholarly exchange on innovative approaches in writing didactics and offers space for the discussion of key developments in this field.

**The dimensions of digital writing**

Afra Sturm (*University of Applied Sciences and Arts Northwestern Switzerland)*

**Abstract:**

Drawing on a model of writing as a social practice (Graham, 2021), this presentation elaborates on various dimensions of digital writing. Firstly, it outlines how digital writing can affect the text production process. Secondly, it explores the opportunities and challenges that digital writing poses for writers and for the teaching of writing.
Another focus is on the interaction between writers and readers, or more broadly, members of a writing community. Digital writing offers the particular opportunity to include members beyond the classroom and to assign different roles or degrees of responsibility to them. The latter must also be clarified if a chatbot is to be used as a writing partner or tutor (Steinhoff & Lehnen, 2025). In a school context, this raises the question of how to design digital writing arrangements that promote learning.
The dimensions identified in this way are exemplified using the new.mymoment.ch writing platform, particularly with regard to writing arrangements that promote learning.

**References:**

Graham, S. (2021). Executive control and the writers(s)-within-community model. In T. Limpo & T. Olive (Hrsg.), *Executive functions and writing* (S. 38–76). Oxford: Oxford University Press.
Steinhoff, T. & Lehnen, K. (2025). Schreiben mit Künstlicher Intelligenz: Das GPT-Modell (Ghost, Partner, Tutor). *Leseräume*, (Preprint), 1–20.

**Measuring narrative (digital) text production skills. Validating RANT**

Johannes Wild *(University of Regensburg)*

**Abstract:**

The quantitative assessment of written narrative competence represents a methodological challenge in writing research (cf. Becker-Mrotzek, 2014). In particular, the validation of the measurement instruments developed for this purpose often receives too little attention in research projects (van Steendam et al., 2012; Wild, 2023). As part of the RESTLESS project funded by the Mercator Institute, an investigation was carried out into the longitudinal development of pupils' writing performance. The development and validation of a corresponding instrument was essential for conducting this study.
As part of the project, written narrative was conceptualized as a multidimensional ability (cf. Wild, Schilcher & Pissarek, 2018; Wild, 2020) and the construct was evaluated in the form of an instrument for recording the quality of narrative texts in the fourth and fifth grades (RANT). The present validation study was dedicated to investigating the fit of the operationalized narrative aspects using the student texts collected as part of the project (cf. Wild, 2020). A confirmatory factor analysis supports the theoretically assumed dimensional structure of the narrative construct in the RANT

model (Wild, 2023). The results of the descriptive analyses and the part-whole corrected item-total correlations show that narrative-specific facets are to be rated as more demanding, but have a higher discriminatory power than generic characteristics (Wild, 2023).

**References:**

Becker-Mrotzek, M. (2014). Schreibleistungen bewerten und beurteilen. In H. Feilke & T. Pohl (Hrsg.), Schriftlicher Sprachgebrauch. Texte verfassen (501–513). Schneider.

van Steendam, E., Tillema, M., Rijlaarsdam, G. & van den Bergh, H. (2012). Introduction. In E. van Steendam, M. Tillema, G. Rijlaarsdam & H. van den Bergh (Hrsg.), Measuring Writing. Recent insights into Theory, Methodology and Practices (ix–xxi). Brill. https://doi.org/10.1163/9789004248489

Wild, J. (2020). Schriftliche Erzählfähigkeiten diagnostizieren und fördern. Eine empirische Studie zum Erfassen von Textqualität in der Primar- und Sekundarstufe [Regensburger Diss.]. Waxmann.

Wild, J. (2023). Schriftliches Erzählen diagnostizieren. Ergebnisse der empirischen Validierung eines Instruments zum Erfassen der Qualität narrativer, schriftlicher Texte im Rahmen des RESTLESS-Projekts. In N. Kruse, V. Lemke, T. Steinhoff & A. Sturm (Hrsg.), Schreibforschung und Schreibunterricht (S. 271-290). Waxmann.

**Writing Stories, Evaluating Texts: Task Formats, Discourse Practices, and Assessment Criteria in Literary Writing**

Daniela Matz *(University of Tübingen)*

**Abstract:**

The writing of literary texts by students in German classes is the core concern of a didactic concept of literary writing. Appropriate writing prompts give learners the opportunity to "not only test the effect of the aesthetic use of language in the respective genre, but also to examine it in class" (Abraham et al., 2022, Übersetzung D.M.). Literary writing has its conceptual origins in a cultural practice of literature (Abraham et al., 2023). For everyday classroom practice, which is often subject to structural and organizational constraints, the question arises as to how the concept can be successfully transformed into German classes.

The presentation first illustrates the didactic approach by drawing on exemplary tasks involving two narrative genres, tall tales and short stories. It is shown what potential for literature-related learning emerges through a production-oriented approach. While tall tales invite learners to playfully engage with adventurous narrative worlds in various media implementations and to reflect on the relationship between truth and lies (Glasenapp, 2011), short stories encourage the portrayal of the protagonists' everyday problems in the condensed form of a single situation (Köppert, 2012). A central challenge lies in assessing both drafts and final texts in ways that are sensitive to genre-specific features. While a literary didactic perspective focuses on literature-related learning, a writing didactic lens must also address the design of writing processes. Building on existing proposals (Abraham & Matz, 2024), this presentation draws on exploratory data to compare dialogic strategies for evaluating texts and applying assessment criteria - both by teachers in classroom settings and by authors in writing workshops in the cultural field. The analysis contrasts how criteria are introduced, negotiated, and made concrete in these contexts, with the aim of distilling principles of dialogic understanding around literary assessment.

**References:**

Abraham, U., Bräuer, C., & von Brand, T. (2022). Literarisches Schreiben. *Praxis Deutsch*, *293*, 4–13.

Abraham, U., Bräuer, C., & von Brand, T. (Hrsg.). (2023). *Literarisches Schreiben als kulturelle Praxis. Deutschdidaktische, schreibpädagogische und medienkulturelle Impulse für den Unterricht*. Klett Kallmeyer.

Abraham, U., & Matz, D. (2024). Kreatives und literarisches Schreiben bewerten. In I. Petersen, R. Reble, & J. Kilian (Hrsg.), *Texte schreiben in allen Unterrichtsfächern. Textbeurteilung als Grundlage für Schreibförderung und Leistungsbewertung* (S. 199–219). Waxmann.

Glasenapp, G., von. (2011). Von Wahrheiten, Un-Wahrheiten, Lügen und Fiktionen. Formen und Funktionen der Lügendichtung. *kjl&m*, *63*(1/2011), 15–24.

Köppert, C. (2012). Kurzgeschichte. In H.-J. Kliewer & I. Pohl (Hrsg.), *Lexikon Deutschdidaktik* (Bd. 1, S. 353–355). Schneider.

**Digital storytelling in multimedia writing arrangements**

Maximilian Stoller *(Karlsruhe University of Education)*

**Abstract:**

This talk presents findings from an intervention study conducted within the framework of the AQUA-d research program. The study explores the potential of digital writing environments to foster narrative text production competence in lower secondary education. The central question addresses whether, and in what ways, multimedia storytelling—understood as Digital Storytelling (Hauck-Thum, 2018; Preis, 2020) can qualitatively enhance students' writing processes.

The theoretical grounding spans writing pedagogy, literary education, and media didactics, drawing additionally on recent perspectives from multimodality research (Ochsner et al., 2018).

Three intervention groups were compared: a baseline group without specific instructional support; a strategy-based group using the Regensburg model “Burg Adlerstein” (Goldenstein et al., 2020); and a group that created multimodal eBooks using the Book Creator app (Uhl, 2016). All students first completed a digital learning pathway aimed at consolidating genre knowledge related to the tall tale („Lügengeschichte”).

To assess narrative text quality, two complementary instruments were applied: the RANT scale (Wild, 2020) and an adapted version of the NAEP holistic rubric (National Center for Education Statistics, 2008), modified specifically for tall tales. This allowed for differentiated analysis of both narrative structure and perceptual elaboration.

Additional control and moderator variables were collected using the FLVT reading comprehension test (Souvignier et al., 2008), the SMiSK questionnaire on writing motivation (Hennes et al., 2023), the Family Affluence Scale (Currie et al., 2024), and an Alphabet Task to assess keyboard-based writing fluency (Linnemann et al., 2022). Other background variables such as literacy experience, media socialization, home language, age, and gender were also considered in the analysis.

The presentation highlights key findings on the effects of multimodal and analytical writing support on the quality of students’ revised narrative texts. It also reflects on methodological and conceptual implications for researching digital writing processes. The talk seeks to contribute to an empirically grounded and interdisciplinary discourse on the future of literary writing instruction in digital contexts.

**References:**

Currie, C., Alemán Díaz, A. Y., Bosáková, L., & De Looze, M. (2024). The international Family Affluence Scale (FAS): Charting 25 years of indicator development, evidence produced, and policy impact on adolescent health inequalities. *SSM - Population Health, 25*, 101599. https://doi.org/10.1016/j.ssmph.2023.101599

Goldenstein, M., Knott, C., Schilcher, A., Sontag, C., Stöger, H., & Wild, J. (2020). *Burg Adlerstein – Schreibtraining. Arbeitsheft*. Westermann Schulbuchverlag.

Hauck-Thum, U. (2018). Fachspezifische Möglichkeiten und Potentiale von Adaptable Books im Lese- und Literaturunterricht der Grundschule. *Mitteilungen des Deutschen Germanistenverbandes, 65*(3), 294–305. https://doi.org/10.14220/mdge.2018.65.3.294

Hennes, A.-K., Michels, V., Becker-Mrotzek, M., Koch, A. M., Möller, J., Woerfel, T., & Fleckenstein, J. (2023). Der SMisK – Ein Fragebogen zur Erfassung der Schreibmotivation im schulischen Kontext. *Didaktik Deutsch, 55*. https://doi.org/10.21248/dideu.682

Linnemann, M., Steinlen, A., & Wieland, L. (2022). Messung schriftsprachlicher Kompetenzen im digitalen Raum: Potenziale des Alphabet-Tasks. *Diagnostik im Deutschunterricht, 12*(1), 45–63.

National Center for Education Statistics. (2008). *The Nation's Report Card: Writing 2007 (NCES 2008–468)*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/nationsreportcard/pubs/main2007/2008468.asp

Ochsner, B., Bateman, J., Thies, S., Curtis, R., & Sachs-Hombach, K. (2018). Medienwissenschaftliche Multimodalitätsforschung [Application/pdf]. https://doi.org/10.17192/EP2018.1.7708

Preis, M. (2020). Multimodale Lektüren: Konfigurationen der Sinne im Literaturunterricht. *MiDU – Medien im Deutschunterricht*, 1–19. https://doi.org/10.18716/OJS/MIDU/2020.2.6

Souvignier, E., Trenk-Hinterberger, I., Adam-Schwebe, S., & Gold, A. (2008). *Frankfurter Leseverständnistest für 5. und 6. Klassen (FLVT 5-6)*. Hogrefe.

Uhl, B. (2016). Storytelling-Apps: Potenziale und Einschränkungen von digitalen Storytelling-Apps zur Förderung des mündlichen Erzählens. *Deutschdidaktik für die Primarstufe, 4*, 49–56.

Wild, J. (2020). *Schriftliche Erzählfähigkeiten diagnostizieren und fördern: Eine empirische Studie zum Erfassen von Textqualität in der Primar- und Sekundarstufe*. Waxmann.

## S7: Formative Assessment - Noticing, Interpreting and Constructing Implications Based on Learning Progress Data

**Presenter:** Daniel Sommerhoff *(Leibniz Institute for Science and Mathematics Education, Mathematics Education Kiel)*, Anouschka van Leeuwen *(Utrecht University)*, Erika Lunowa *(Karlsruhe University of Education)*
**Chair:** Erika Lunowa *(Karlsruhe University of Education)*
**Discussion:** Sarah Bez *(Karlsruhe University of Education)*

October 9, 2025, 10:30-12:30
3.109

**Abstract:**

Advancing digitalization makes it easier to collect, evaluate and document data. These data can be used in an iterative process to inform decisions in schools, called *data-based decision making*. These decisions often aim to improve the quality of schools and teaching. At the micro level, formative assessment is considered as a form of data-based decisions that promote learning. Formative

assessment is a cyclical feedback process characterized by three essential steps: (1) setting a learning goal, (2) determining students' current understanding, and (3) implementing instructional steps that are helpful to achieve this learning goal. However, an important factor for the effectiveness appears to be the alignment between recurring assessments and the following instructional steps. Therefore, it is necessary that teachers notice important information in data (e.g., the assessment results), interpret them appropriately and, based on that, construct implications for instruction. However, teachers often struggle to interpret learning progress data.

Against this background, the symposium addresses the question how do teachers notice learning progress data and how they can be supported in interpreting and constructing implications for instruction.
**Contribution 1** presents different theories, each providing an explanation for the important question of how, when and why teachers use digitally provided student data to inform instructional decisions. Building on this, promising future fields of research and guidelines for dashboard design and implementation will be outlined.
**Contribution 2** provides insights into the effectiveness of learning progress monitoring on students' academic performance based on a meta-analysis. The results suggest that support relating to data appears to be important with regard to instructional adaptation.
**Contribution 3** focuses on the potential of visualizations. It presents the results of two experiments that investigated whether graphical representations of learning progress data influence teachers' interpretations.

Overall, this symposium offers various complementary approaches to the question how teachers can be supported in interpreting learning progress data. This appears to be crucial for an effective implementation of formative assessment in practice.

**Teachers' Decision Making with Learning Analytics: a Discussion of Different Theoretical Perspectives**
Anouschka van Leeuwen *(Utrecht University)*

**Abstract:**

Teaching is an inherently complex and dynamic practice that requires constant monitoring, reflection, and adjustment. At the heart of this complexity lies the critical process of teacher decision making. Teachers continuously observe and interpret students' behaviors, progress, and engagement in order to adapt their instructional strategies and better support learning. This ongoing process of noticing, interpreting, and responding is central to effective teaching practice, yet it remains challenging, especially in increasingly large, diverse and technology-enhanced classrooms.
In recent years, Learning Analytics (LA) has been proposed as a promising tool to support teachers in this endeavor. By providing visualizations, dashboards, and feedback derived from student data, LA systems aim to enhance teachers' capacity to make informed instructional decisions. However, simply providing LA is not enough. Whether, how, and why teachers use LA in their decision making is currently an active area of research across educational contexts.
In this talk, I aim to provide an overview and comparison of various theoretical lenses that researchers have employed to understand teacher interactions with LA. I explore various types of lenses that provide complementary insights into this complex process. The frameworks I will discuss include those focusing on technology integration in existing teacher practices, frameworks describing motivational aspects of teacher use of LA, frameworks that focus on power dynamics in

terms of relationships and structures within education, and, finally, frameworks that focus on teachers' reasoning skills and evidence use.
By comparing and contrasting these frameworks, I demonstrate that each perspective captures a specific element of the multifaceted nature of teacher decision making with LA. The talk concludes with a reflection on how combining these lenses can lead to new avenues for research and guidelines for design and implementation of effective LA tools.

**The Relevance of Teacher Support: Unlocking the Potential of Learning Progress Monitoring**

Daniel Sommerhoff & Amelie Fuchs *(Leibniz Institute for Science and Mathematics Education, Mathematics Education Kiel)*

**Abstract:**

To maximise learning achievements, individualised instruction and evidence-based teaching are essential. However, this requires teachers to continuously and accurately assess their students' learning status and individual progress—ideally without spending excessive time on diagnosis or requiring extensive qualifications from the teachers.
In this regard, learning progress monitoring (LPM) is a promising approach to capture students' learning progress and allow individualised instruction. LPM uses recurrent short tests to measure individual learning progress and report the results to teachers (and students). While some studies suggest that LPM can improve student performance, previous findings have been heterogeneous, indicating that LPM may be particularly effective under certain conditions.
Therefore, we conducted a meta-analysis to answer the following questions:
(i) To what extent does LPM impact the development of students' academic performance?
(ii) To what extent does the effectiveness of LPM depend on aspects of the setting, implementation of the LPM, and handling of the LPM results?
In our systematic literature search, we used terms such as LPM, curriculum-based measurement, and learning progress assessment to identify relevant studies. Only studies that investigated the effectiveness of LPM on academic performance using a (quasi-)experimental design with no additional variations in the setting beyond LPM and that were published in English in a peer-reviewed journal were included. We synthesised 87 effect sizes from 25 studies, involving 7,379 students.
The results revealed a small positive overall effect of LPM on students' performance ($g = 0.30$) compared to business-as-usual instruction. Based on moderator analyses focusing on the setting, implementation, and handling of LPM results, our data show that LPM is effective for students, both with and without educational needs. Furthermore, LPM tests should be implemented at least weekly, and additional support for teachers should be provided. In particular, offering teachers data-specific guidance on how to adapt their instruction led to large effects of LPM ($g = 0.73$).
These findings highlight that LPM is effective, but support for teachers is crucial for unlocking its full potential. However, it remains unclear which characteristics of teacher support are responsible for its enhanced effectiveness and how they relate to the stages of the LPM process. Thus, further research is needed to investigate the mechanisms underlying the effectiveness of LPM in more detail, especially regarding the most effective types of teacher support. This requires both basic research and studies that examine scalability and cost-benefit considerations, as, for example, one-on-one teacher support will not be viable.

**The Influence of Graphical Representations on Teachers' Interpretation of Learning Progress Data**

Erika Lunowa *(Karlsruhe University of Education)*, Sarah Bez *(Karlsruhe University of Education)*, Tom Rosman *(Leibniz Institute for Psychology)*, Martin J. Tomasik *(University of Zurich)*, & Samuel Merk *(Karlsruhe University of Education)*

**Abstract:**

Formative assessment can be characterized by the features assessing the current understanding of students related to the learning goal and determining next instructional steps. For a successful implementation, teachers need to determine the learner's learning level. To do this, they need to evaluate the results by applying the individual reference norm, the criterial reference norm, or the social reference norm. For formative assessment, the use of the individual or criterial reference norms seems to be favourable, as these reference norms are inherent to the concept of formative assessment.

However, in the few studies that exist, teachers generally seem to struggle when interpreting formative assessment results. Moreover, they tend to use the social reference norm, which seems to be less optimal for learners. In connection with the general tendency of a teacher to prefer a certain reference norm (the so-called reference norm orientation), it is assumed that there are both trait-like and situation-influenced components. Against this background, supporting teachers in interpreting learning progress data seems to be important. Visualizations, which are increasingly used to present data to teachers, could be used to present data in a clear way, e.g., by highlighting important aspects and could serve a tool to encourage teachers in applying appropriate reference norms.

Therefore, we investigated in a randomized (mixed within-between-person) study with $N = 223$ pre-service teachers the following questions:

1. Which reference norms do teachers apply when interpreting learning progress data?
2. Can teachers' use of reference norms while interpreting learning progress data be influenced by the graphical representation of learning progress data?
3. Does teachers' (trait-like) reference norm orientation predict the reference norm applied in a specific assessment situation?

The results suggest that preservice teachers generally adopted the individual and/or criterial reference norm in a (comparatively) neutral presentation. A presentation emphasizing a specific reference norm seems to have a medium to strong effect on the reference norm adopted in the interpretation ($OR = 4.25$ [individual reference norm]; 4.75 [criterial reference norm]; 11.17 [social reference norm]). However, pre-service teacher's reference norm orientation could hardly predict the situationally adopted reference norm. Overall, the results suggest that applying the individual and criterial reference norms can be supported by graphical representations that emphasize these reference norms.

To strengthen the internal, external and construct validity, this study will be replicated with in-service teachers. The results will be presented as well.

## S8: Metacognition and Desirable Difficulties: Enhancing Self-Regulated Learning and Judgment Accuracy in Digital Learning Environments

**Presenter:** Louise David *(Maastricht University)*, Martin Fifka *(Karlsruhe University of Education)*, Noortje Janssen *(Radboud University)*, Theresa Walesch *(Karlsruhe University of Education)*
**Discussant:** Anique de Bruin *(Maastricht University)*
**Chair:** Anja Prinz-Weiß *(Karlsruhe University of Education)*

October 9, 2025, 13:30-15:30
3.101

This symposium explores different digital contexts in which judgments and judgment accuracy play a significant role to support self-regulated learning (SRL). Students in digital learning environments must self-regulate their learning to a greater extent than in traditional learning environments, due to reduced teacher guidance and increased latitude in learning decisions. However, students across educational levels and domains frequently struggle with SRL. Improving SRL is crucial for school students and university students. This indicates the need for effective interventions to support students' SRL skills.

An important aspect of effective SRL are accurate judgments of one's own learning, as well as the quality of learning materials. Specifically, more accurate judgments lead to more effective regulation processes (e.g., investment of study time), which in turn improve performance. However, previous studies have indicated that learners' judgments are often inaccurate, leading to suboptimal study decisions. This raises the question of how to best support students in improving their judgment accuracy. Retrieval practice is one of the most effective, yet underused learning strategies. Challenging retrieval tasks (e.g., parameterized tasks (4) or problem solving tasks (3)) are so-called "desirable difficulties", which are learning conditions that enhance learning, but are perceived as effortful by learners. Specifically, retrieval tasks support long-term retention but are often perceived as difficult – something learners might misinterpret as a lack of learning progress. Thus, desirable difficulties may decrease learners' judgment accuracy and motivation.

The effectiveness of retrieval practice can be enhanced by feedback, yet it is unclear which type of feedback is optimal. Visual feedback, in particular, may support accurate self-monitoring, self-regulation, and motivation (2). Furthermore, parameterized tasks are a type of retrieval task that have interchangeable parameters (e.g., numbers). Hence, they have the potential to provide learners with more information to judge their own learning (4). Emerging technologies such as Large Language Models (LLMs) offer new opportunities to implement retrieval practice. For example, they enable learners to generate highly individualized practice questions. However, learners might misjudge the quality of LLM-generated questions (1). Furthermore, new research shows that primary school students have low judgment accuracy in the emerging field of creative problem solving. LLMs are discussed as a promising tool to support these students. (3).

Through examining interventions and discussing emerging technologies, this symposium aims to provide actionable insights to support learners in making more accurate judgments, fostering their motivation, and ultimately making better decisions when self-regulating their learning in naturalistic settings.

**(1) Enhancing University Students' Self-Regulated Learning Through LLM-Generated Practice Questions**

Louise David *(Maastricht University)*, Felicitas Biwer *(Maastricht University)*, Dietbert Neumann *(Maastricht University),* & Anique de Bruin *(Maastricht University)*

**Abstract:**

Practice testing is a highly effective learning strategy (Carpenter et al., 2022), yet university students often underutilize it during their self-regulated learning partially due to the lack of accessible

practice questions (Carpenter et al., 2023). Creating self-testing materials is often perceived as time-consuming and effortful by students discouraging their engagement (Biwer et al., 2020; Rea et al., 2022). One way to overcome this barrier is to provide testing resources, but this increases educator workload and might limit students' ownership of their learning. Large Language Models (LLMs) could offer a promising alternative by enabling students to generate their own self-testing materials (Lee, 2023). However, little is known about how students evaluate and use LLM-generated resources (Kasneci et al., 2023; Gabajiwala et al., 2022). Therefore, in the current study, we explored whether LLMs could serve as a feasible tool to increase university students' engagement in practice testing. We explored whether first-year biomedical science students (N = 149) critically evaluate the quality of practice questions generated by LLMs and to what extent they use these questions when given access through a course-integrated question bank. As part of a mandatory course assignment, students generated practice questions using an LLM which were then rated by a domain expert and peer. Questions of sufficient quality were integrated into a digital question bank, and we tracked students' voluntary engagement throughout the course. Results indicate that LLM-generated questions received a mean quality rating of 4.05 out of 5 (SD = 1.04) from experts and 4.74 out of 5 (SD = 0.58) from students, suggesting a high quality perceived by both experts and students. However, experts would only include 50% (SD = 0.5) of these questions as an exam question due to low complexity. Contrastingly, students would have expected 87% (SD = 0.33) of the questions to be included in an exam. Usage data of the question bank indicated most students used the question bank only for one (n = 47) or two (n = 32) sessions in the week of the course exam. These findings suggest that while students are willing to use accessible practice questions, they tend to do so primarily shortly before exams and may overestimate the quality of LLM-generated items. This highlights both the potential of LLMs to support self-regulated practice testing but also a need for scaffolding to help students evaluate question quality effectively.

**References:**

Biwer, F., Egbrink, M. G. A. O., Aalten, P., & De Bruin, A. B. H. (2020). Fostering effective learning strategies in higher education—A mixed-methods study. *Journal of Applied Research in Memory and Cognition*, 9(2), 186–203. https://doi.org/10.1016/j.jarmac.2020.03.004

Carpenter, S.K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review* 35, 96. https://doi.org/10.1007/s10648-023-09811-8

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, 1–16. https://doi.org/10.1038/s44159-022-00089-1

Gabajiwala, E., Mehta, P., Singh, R., & Koshy, R. (2022). Quiz Maker: Automatic Quiz Generation from Text Using NLP. In P. K. Singh, S. T. Wierzchoń, J. K. Chhabra, & S. Tanwar (Eds.), Futuristic Trends in Networks and Computing Technologies (Vol. 936, pp. 523–533). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5037-7_37

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lee, H. (2023). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*, 2270. https://doi.org/10.1002/ase.2270

Rea, S. D., Wang, L., Muenks, K., & Yan, V. X. (2022). Students Can (Mostly) Recognize Effective Learning, So Why Do They Not Do It? *Journal of Intelligence*, 10(4), 127. https://doi.org/10.3390/jintelligence10040127

**(2) Visual Feedback on Learners' Self-Assessments: Supporting Motivation, Regulation, and Comprehension in School Contexts**

Martin Fifka *(Karlsruhe University of Education)*, Nicolas Hübner *(University of Bonn)*, Anja Prinz-Weiß *(Karlsruhe University of Education)*

**Abstract:**

Accurate Judgments of Learning (JOLs) are a crucial component of self-regulated learning. They allow learners to allocate their study time more effectively and focus on material that requires further review, ultimately improving learning outcomes (e.g., Rawson et al., 2011; Robey et al., 2017). Despite their potential benefits, JOLs are rarely implemented systematically in educational settings (e.g., De Bruin & Van Gog, 2012). A notable exception in some German schools are so-called can-judgments, a structured form of self-assessment in which students rate their ability to perform specific learning tasks, using statements such as "I can name the different kinds of polygons" or "I can draw points into a Cartesian coordinate system" (e.g., Hoffmann, 2016; Schneider et al., 2012). However, even this type of JOL is often used without further scaffolding such as feedback.

Prior research has indicated that providing learners with visual feedback on their learning behavior can enhance their motivation and self-regulated learning (e.g., Wäschle et al., 2014). Building on this work, we investigated whether visual feedback on students' can-judgments might improve their motivation, self-regulated learning, and comprehension when learning from text.

In this study, ninth-grade students first read an expository physics text. They then answered comprehension questions as a form of retrieval practice. For each question, they provided a can-judgment, rating how well they believed they could answer similar questions on the same topic. Depending on the experimental condition, participants then received different types of visual feedback while restudying the text: (a) feedback indicating the accuracy of their judgments (i.e., highlighting areas where they were accurate, overconfident, or underconfident), (b) feedback reflecting their self-assessed understanding without accuracy information on over- or underestimations, or (c) no feedback (control group). After the restudy phase, participants rated their motivation during this second reading. Finally, all participants completed a comprehension test consisting of parallel items from the retrieval practice phase.

We hypothesized that visual feedback, particularly feedback that includes information on judgment accuracy, would lead to higher motivation, longer rereading times (as an indicator of self-regulation), and greater improvements in comprehension compared to conditions without or with less informative feedback. Data from 303 secondary school students are currently being analyzed. In addition to the primary outcome measures, we also explore how learners' interest in physics, reading self-concept, and self-assessment ability in the subject domain relate to the dependent variables.

This study aims to provide new insights into the role of metacognitive feedback in supporting motivation and learning in school contexts, with potential implications for integrating JOL-based strategies into regular classroom situations.

**References:**

De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Hoffmann, K. (2016). Eigene Lernwege. In GGG Verband für Schulen des gemeinsamen Lernens & Debus Pädagogik Verlag (Eds.), Leistungen ermitteln – Lernen fördern (pp. 36–39).

Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology: Applied, 17*(3), 288–302. https://doi.org/10.1037/a0024749

Robey, A., Dougherty, M., & Buttaccio, D. (2017). Making retrospective confidence judgments improves learners' ability to decide what not to study. *Psychological Science, 28*, 095679761771880. https://doi.org/10.1177/0956797617718800

Schneider, S., Graf, G., & Würth, T. (2012). *Individuelle Förderung: Unterrichtssequenzen.* https://lehrerfortbildung-bw.de/u_matnatech/bio/bs/2bfs/2bfs1/vorwort/

Wäschle, K., Lachner, A., Stucke, B., Rey, S., Frömmel, C., & Nückles, M. (2014). Effects of visual feedback on medical students' procrastination within web-based planning and reflection protocols. *Computers in Human Behavior, 41*, 120–136. https://doi.org/10.1016/j.chb.2014.09.022

**(3) The Association Between Self-efficacy and Judgment Bias in Creative Problem Solving Tasks**

Noortje Janssen *(Radboud University)*, Inge Heida, Anne Cornelissen and Mare van Hooijdonk *(Radboud University)*

**Abstract:**

**Introduction**

Creativity, the ability to generate both original and useful ideas, is an essential skill for addressing contemporary societal challenges (Plucker et al., 2004). Elementary schools increasingly engage students in creative problem solving. However, students are often biased in their judgments of creativity (Urban & Urban, 2021), leading to inefficient problem solving. Little is known about the underlying factors that determine students' judgment bias. This study investigates whether students' creative self-efficacy—their self-perception of their creative potential (CSE; Karwowski et al., 2018), is associated with their judgment bias.

**Methods**

Elementary school students (n = 130) filled out the CSE questionnaire (Karwowski et al., 2018) and completed three creative problem-solving tasks: the alternative uses task (Guilford, 1967), the product improvement task (Van Broekhoven et al., 2022), and a science task (Van Hooijdonk et al., 2023). After each task, students assessed the originality and usefulness of their solution on four-point Likert scales. Experts rated each solution on the same scales. Judgment bias was calculated as: judgments − expert-ratings.

**Results**

Students were generally overconfident, which was higher for originality ($M = 0.84$, $SD = 0.82$) than for usefulness ($M = 0.29$, $SD = 1.02$). Regression analyses showed that CSE was positively related to students' bias regarding both originality ($F(1,127) = 6.97$, $p = .009$, $R^2 = 0.05$) and usefulness ($F(1,127) = 10.35$, $p = .002$, $R^2 = 0.08$). However, exploratory task-specific analyses showed that the science task

elicited greater bias in students' ratings of originality and usefulness than the others. Furthermore, Tau correlations between students' CSE and judgment bias in the science task were not statistically significant.

**Conclusion and future research**

Elementary school students tend to overestimate the creativity of their solutions, with creative self-efficacy (CSE) contributing a small but significant effect. This finding aligns with prior research on academic tasks (Vössing & Stamov-Roßnagel, 2016). The science task showed higher bias unrelated to CSE, suggesting that both CSE and the nature of the task should be considered to reduce judgment bias.

AI presents an intriguing opportunity to enhance students' judgment accuracy. Recent language models have been able to reliably score students' creative solutions (Marrone et al., 2022). Such automatic scoring provides an external benchmark, which has been shown to support judgment accuracy in academic problem-solving (Janssen & Lazonder, 2024), making it a promising tool for enhancing students' judgment of creative ideas.

**References:**

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior, 1*(1), 3-14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Janssen, N., & Lazonder, A. W. (2024). Meta-analysis of Interventions for Monitoring Accuracy in Problem Solving. *Educational Psychology Review*, *36*(3), 96. https://doi.org/10.1007/s10648-024-09936-4

Karwowski, M., Lebuda, I., & Wiśniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *The International Journal of Creativity & Problem Solving*, *28*(1), 45-57.

Marrone, R., Cropley, D. H., & Wang, Z. (2023). Automatic Assessment of Mathematical Creativity using Natural Language Processing. *Creativity Research Journal*, *35*(4), 661-676. https://doi.org/10.1080/10400419.2022.2131209

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83-96. https://doi.org/10.1207/s15326985ep3902_1

Urban, M., & Urban, K. (2021). Unskilled but aware of it? Cluster analysis of creative metacognition from preschool age to early adulthood. *The Journal of Creative Behavior*, *55*(4), 937-945. https://doi.org/10.1002/jocb.499

Van Broekhoven, K., Belfi, B., & Borghans, L. (2022). Instructing children to construct ideas into products alters children's creative idea selection in a randomized field experiment. *Plos One, 17*(8), https://doi.org/10.1371/journal.pone.0271621

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2023). Creative problem solving in primary school students. *Learning and Instruction, 88*, 101823. https://doi.org/10.1016/j.learninstruc.2023.101823

Vössing, J., & Stamov-Roßnagel, C. (2016). Boosting metacomprehension accuracy in computer-supported learning: The role of judgment task and judgment scope. *Computers in Human Behavior*, *54*, 73-82. https://doi.org/10.1016/j.chb.2015.07.066

**(4) The Influence of Parametrized Tasks on Students' Accuracy of Judgements of Learning – A Secondary Analysis**

Theresa Walesch *(Karlsruhe University of Education)*, Carolin Baumann *(Karlsruhe University of Education)*, Samuel Merk *(Karlsruhe University of Education)* & Anja Prinz-Weiß *(Karlsruhe University of Education)*

**Abstract:**

**Background**

In digital learning environments, students must self-regulate their learning (e.g., Carpenter et al., 2020). However, students' judgments are often inaccurate (e.g., Prinz et al., 2020), which can lead to suboptimal study decisions (e.g., Choi et al., 2023).

Retrieval practice is among the most effective learning strategies (e.g., Moreira et al., 2019). An innovative way to implement retrieval practice in digital learning environments is through parametrized tasks – dynamic, algorithmically generated tasks that vary relevant parameters (e.g. numbers, words) (Michael, 2021). Providing students with increased task variation might prevent them from just recognizing the correct answer and provide them with more information (i.e., cues) to judge their learning.

**Present Study**

In the context of a secondary analysis, we investigated the role of different types of retrieval tasks (parametrized vs. non-parametrized tasks) for students' judgment accuracy. We hypothesized that students would make more accurate judgments after completing parametrized tasks compared to non-parametrized tasks.

**Method**

The data was collected during four consecutive weeks in a statistics lecture at a German university. In the first week of the lecture, students were randomly assigned to a retrieval task type (parametrized vs. non-parametrized), with the task type being altered the following weeks (i.e., within-person design). Each week, after working on the tasks, the students predicted their success in the final exam (judgment). At the end of the semester, the students took the final exam. Judgment accuracy was operationalized as bias, that is, the signed difference between a judgment and performance in the final exam.

**Findings**

A Bayesian multilevel model with a random intercept was fitted to account for repeated measures. The intercept for parametrized tasks was $\beta = 0.29$ (95% CI [0.27, 0.32]). The estimated effect of the condition (slope) was $\beta = 0.01$ (95% CI [−0.02, 0.04]), suggesting that a difference in accuracy between the two task types is likely minimal. An exploratory analysis on the distribution of judgment accuracy indicated underestimation for both task types.

**Discussion**

This study extends previous research by focusing on an innovative type of retrieval task, namely parametrized tasks, and its relationship with students' judgment accuracy. Contrary to expectations, students did not make more accurate judgments after working on parametrized tasks and tended to underestimate their performance across both task types. This tendency can lead to ineffective regulation and impair performance. Hence, this study suggests that students need support with retrieval tasks to more accurately judge their learning.

**References:**

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition, 9*(2), 137–151. https://doi.org/10.1016/j.jarmac.2019.12.009

Choi, H., Jovanovic, J., Poquet, O., Brooks, C., Joksimović, S., & Williams, J. J. (2023). The benefit of reflection prompts for encouraging learning with hints in an online programming course. The Internet and Higher Education, 58, 100903. https://doi.org/10.1016/j.iheduc.2023.100903

Michael, B. (2021). E-Assessment: Automatische Generierung parametrisierter Aufgaben für mathematische Assessments in E-Learning-Systemen [Doctoral disserthation, Technische Universität Ilmenau]. DBT Open Repository. https://doi.org/10.22032/DBT.49387

Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Jaeger, A. (2019). Retrieval practice in classroom settings: A review of applied research. Frontiers in Education, 4, 5. https://doi.org/10.3389/feduc.2019.00005

Prinz, A., Golke, S., & Wittwer, J. (2020). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. Educational Research Review, 31, 100358. https://doi.org/10.1016/j.edurev.2020.100358

## S9: Digital learning in mathematics education

**Presenters:** Caroline Marx *(Karlsruhe University of Education),* Daniel Sommerhoff *(Leibniz Institute for Science and Mathematics Education Kiel)*, Manuel Ninaus *(University of Graz),* Korbinian Moeller *(Loughborough University)*
**Chair:** Christiane Benz (*Karlsruhe University of Education)*

October 9, 2025, 13:30-15:30
3.101

Drawing on national and international research, this symposium explores how Digital Learning Resources (DLRs) can be effectively designed and implemented to promote mathematical learning processes. Thus, it offers four complementary perspectives on the design, evaluation, and practical use of technology-enhanced learning environments in mathematics instruction.
The first contribution focuses on the educational foundations of digital task design in early mathematics education. It presents a DLR grounded in learning theories on part-whole understanding, integrating structured visualisations, different levels of action (enactive, iconic, symbolic), systematic decompositions of quantities, and motivational feedback. An initial study with $n$= 111 primary school students highlights both, the theoretical underpinnings, and practical applications of DLRs in early mathematics education.
The second presentation emphasizes the need to design digital assessment tasks theoretically grounded to effectively support assessing and improving mathematical skills. In that vein, research indicates that the use of Difficulty Generating Item Characteristics (DGICs) to design digital assessment tasks for mental addition and subtraction in the number range up to 100 significantly predicted item difficulty. Thus, integrating subject-specific DGICs when designing digital assessment tasks supports teachers' evidence-based teaching and individual student support.
The third talk examines the use of game-based learning elements in DLRs and their potential to enhance both, cognitive and non-cognitive learning processes. In that context, studies on fraction learning and number line estimation tasks indicate that implementing emotional design elements

in DLRs can significantly enhance students' conceptual knowledge, their situational interest and self-efficacy. Moreover, there are promising effects underlining that game elements reduce negative emotions towards mathematics, especially among learners with dyscalculia. This underscores that implementing game-based elements in DLRs supports students' motivation and self-efficacy while maintaining rigorous instructional effectiveness.
The fourth presentation focusses on the use of intelligent tutor systems (ITS) to support mathematical learning processes. In that vein, recent evidence from large-scale data on ITS use in mathematics education will be discussed and critically evaluated. Taken together, research suggests that ITS facilitate learning mathematics by automatic adaption to user needs even in situations of distance learning, but still need teachers to fully exploit their potential.
Together, these contributions highlight how DLRs—when aligned with educational theory and supported by data—can foster individualized, engaging, and effective mathematics learning. The symposium contributes to a nuanced understanding of how DLRs can be developed, validated, and embedded in educational practice, offering valuable implications for researchers, practitioners, and teacher educators alike.

**Facilitating early part-whole understanding using a digital learning resource**

Caroline Marx *(Karlsruhe University of Education)*

**Abstract:**

Understanding part-whole relations is widely considered one of the most fundamental milestones in early mathematical development. Thus, mathematics education is highly interested in evaluating how to support students' development of part-whole understanding using analogue as well as digital learning materials. However, recent evidence suggests that digital learning materials largely fail to offer systematic approaches supporting students' learning processes in that context. Therefore, we are developing a Digital Learning Resource (DLR) providing a systematic approach to support understanding part-whole relations and enhancing technology supported mathematics instruction. Hence, based on the Design Thinking approach, we followed some iterative circles to develop a first prototype of our DLR. Accordingly, in a multidisciplinary team consisting of mathematicians, teachers, cognitive psychologists as well as a software engineer we completed a first micro cycle of empathizing, defining and ideating as well as prototyping. During these iterative stages, it was key to identify how existing learning theories on the development of children's part-whole understanding might support the development of our first prototype. Nonetheless, a literature review revealed that no comprehensive framework exists providing systematic explanations and suggestions how to support these learning processes. Consequently, we systematically collated principles considered important from a mathematics education as well as software developer's perspective when designing a DLR. Thus, in the first prototype, the following principles were implemented: *a*) the use of a ten-strip as a *structured visual representation, b*) number decompositions across all *levels of action* (*enactive, iconic* and *symbolic*), *c) systematic decomposition with*in the number range from three to ten (i.e., implementing all four different number decompositions for the quantity three: 1-2, 2-1, 0-3, 3-0), *d)* the option to individually select numbers and representation levels, and e*)* immediate m*otivational feedback* provided by a squirrel after each task.
An initial testing phase of that prototype embedded within a pre-/posttest pilot study with *n*=111 primary school students revealed what worked well and what needs to be improved. Based on qualitative user feedback as well as on statistical analyses, we can now elaborate the potential of our DLR regarding the support of children's part-whole learning processes. Accordingly, initial results and findings will be presented and discussed. As such, we will critically reflect upon the

principles we implemented as our educational foundation for that prototype, as well as practical applications when using the DLR in class in order to enhance digitally supported learning processes in mathematics instruction.

**Digital assessment in mathematics education: Why 65+3 is like 53+6 but not like 36+5.**
Daniel Sommerhoff *(Leibnitz Institute for Science and Mathematics Education Kiel)*

**Abstract:**

Digital learning resources hold the promise (i) to support educators in addressing student heterogeneity and adapting their teaching, as well as (ii) to facilitate students in acquiring basic mathematical skills. However, following the "garbage in, garbage out" principle, the effectiveness of digital learning resources depends on the quality of the embedded tasks and materials, which need to be carefully designed. This is especially true for all forms of digital assessment.
To highlight and address this issue, this talk focuses on designing assessment tasks for mental addition and subtraction in the number range up to 100 in a digital learning progress monitoring environment. This environment aims to assess students' individual learning progress and serves as a basis for teachers' evidence-based teaching and individual student support. Items were systematically constructed using three basic Difficulty Generating Item Characteristics (DGICs) derived from prior research: (1) the type of arithmetic operation (addition or subtraction), (2) whether a decade boundary is crossed (e.g., 38 + 7 crosses a ten), and (3) the number of digits in the second operand (single-digit vs. two-digit). These DGICs act as cognitive demand factors that influence task complexity and can be systematically combined to create items of varying difficulty.
To validate the predictive power of the DGICs on item difficulty, $N = 591$ students in grade 2 ($n = 205$) and grade 3 ($n = 386$) were studied. The Rasch Model (RM) and the linear logistic test model (LLTM) were used to estimate the difficulty of 41 items and to assess the impact of DGICs on item difficulty. The results indicate that all three DGICs significantly predicted item difficulty in the LLTM approach. In particular, each DGIC holds predictive power, even when controlling for the other DGICs. Furthermore, the DGICs explain about 20% of the variance in the item parameters of the RM.
Overall, the results emphasize that addition items are not just addition items and that a subject-specific focus is necessary when developing digital learning resources aimed at assessing or improving mathematical skills. At the same time, the data emphasize that corresponding items can be created (at least to some extent) using a rule-based approach that can be further refined by incorporating additional, more specific DGICs.

**Affordances of game-based learning to improve mathematics education**
Manuel Ninaus *(University of Graz)*

**Abstract:**

Game-based learning has emerged as a promising approach to enhance mathematics education, particularly by addressing both cognitive and non-cognitive aspects of learning. This talk will present findings from two studies investigating how game elements and emotional design within game-based environments can positively impact learners.
The first study explores the role of emotional design in scaffolding within a fraction learning game for school-aged children. Besides the fact that the fraction learning game significantly improved conceptual fraction knowledge, we were particularly interested in non-cognitive outcomes. By comparing neutrally designed scaffolds to scaffolds utilizing emotional design principles within the learning game, the study demonstrated that emotional design can significantly enhance situational

interest and self-efficacy, thereby increasing the educational value of game-based learning environments.
The second study focuses on children with dyscalculia, a population particularly vulnerable to math-related anxiety and low self-efficacy. Comparing game-based, non-game, and paper-pencil formats of a number line estimation task, children expressed a strong preference for the game-based task and perceived themselves as performing better in it, suggesting that game elements can foster motivation and engagement. Importantly, the game elements utilized in the game-based number line estimation task did not distract the children, as performance was comparable across conditions. As such, the study suggests that game elements might play a vital role in mitigating negative emotions towards mathematics.
Together, these studies highlight the affordances of game-based learning in addressing both the cognitive and non-cognitive dimensions of mathematics education. By incorporating emotional design and game elements, educators and designers can create digital learning environments that enhance motivation and self-efficacy while maintaining rigorous instructional effectiveness. Implications for pedagogical practices and design considerations in game-based learning will be discussed.

**Learning mathematics online – evidence from intelligent tutor systems**
Korbinian Moeller (*Loughborough University)*

**Abstract:**

Intelligent tutor systems (ITS) are a key approach to learning mathematics which will probably gain relevance in future educational settings. ITS provide more and more sophisticated implementation of latest theoretical approaches to learning mathematics integrated with recent advances in educational technology and have recently seen a considerable increase in use during the CoViD-19 pandemic. As typically all user interactions with such ITS are logged, they also provide rich process data (including information on what was learned when, for how long and how successfully, requiring which help options, etc.) closely reflecting naturalistic learning trajectories. In this presentation, I will evaluate recent research on an ITS for learning mathematics widely used by thousands of students in Austria, Germany, and the Netherlands in four consecutive steps to evaluate the validity of data taken from such ITS. In a first step, the validity of the data is discussed by evaluating whether critical findings from basic research on longitudinal learning trajectories in mathematics can be replicated leveraging large-scale data from an ITS. In a second step, I will elaborate on how school-closures and, thus, the increased reliance on online learning affected performance on the learning environment. Third, I will then discuss how specific characteristics of using the digital environment (e.g., tasks assigned by teachers vs. self-assignment) critically affect learning mathematics. In a final step, I discuss how closed-loop research on the effects of nudging users to close identified knowledge gaps particularly may benefit learners as well as ITS designers alike. Results clearly indicate that i) typical learning trajectories can be replicated within the ITS, ii) performance increased during school closures, but iii) depended critically on teacher involvement, and iv) system generated hints to close identified knowledge gaps. This seems to suggest that ITS facilitate learning mathematics by automatic adaption to user needs even in situations of distance learning, but still need teachers to fully exploit their potential.