



# Combining exploratory learning with structured practice educational technologies to foster both conceptual and procedural fractions knowledge

Manolis Mavrikis<sup>1</sup> · Nikol Rummel<sup>2</sup> · Michael Wiedmann<sup>2,3</sup> · Katharina Loibl<sup>2,3</sup> · Wayne Holmes<sup>1</sup>

Accepted: 9 March 2022  
© The Author(s) 2022

## Abstract

Educational technologies in mathematics typically focus on fostering either procedural knowledge by means of structured tasks or, less often, conceptual knowledge by means of exploratory tasks. However, both types of knowledge are needed for complete domain knowledge that persists over time and supports subsequent learning. We investigated in two quasi-experimental studies whether a combination of an exploratory learning environment, providing exploratory tasks, and an intelligent tutoring system, providing structured tasks, fosters procedural and conceptual knowledge more than the intelligent tutoring system alone. Participants were 121 students from the UK (aged 8–10 years old) and 151 students from Germany (aged 10–12 years old) who were studying equivalent fractions. Results confirmed that students learning with a combination of exploratory and structured tasks gained more conceptual knowledge and equal procedural knowledge compared to students learning with structured tasks only. This supports the use of different but complementary educational technologies, interleaving exploratory and structured tasks, to achieve a “combination effect” that fosters robust fractions knowledge.

**Keywords** Intelligent tutoring systems · Interactive learning environments · Applications in subject areas (mathematics) · Elementary education · Interdisciplinary projects

---

Manolis Mavrikis and Nikol Rummel have contributed equally to this work as first authors.

---

✉ Manolis Mavrikis  
m.mavrikis@ucl.ac.uk

<sup>1</sup> UCL Knowledge Lab, University College London, 29 Emerald St, London WC1N 3QS, UK

<sup>2</sup> Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

<sup>3</sup> Present Address: Pädagogische Hochschule Freiburg, Kunzenweg 21, 79117 Freiburg im Breisgau, Germany

## Introduction

Two commonly distinguished types of mathematical knowledge are conceptual and procedural knowledge (Anderson, 1987; Hiebert, 1986; Rittle-Johnson et al., 2001; Star & Stylianides, 2013). Both develop at the same time (Canobi et al., 2003; LeFevre et al., 2006), develop iteratively and evolve in a relationship of mutual dependence (Baroody et al., 2007; Rittle-Johnson & Koedinger, 2009; Rittle-Johnson et al., 2015), with increases in conceptual knowledge leading, in a virtuous circle, to parallel gains in procedural knowledge and vice versa. It is therefore somewhat surprising that prior work in the learning sciences and particularly educational technology has primarily focused on fostering either procedural knowledge or conceptual knowledge, rather than both. In contrast, the work reported here investigates whether a combination of tasks from different types of educational technology can be used to foster both conceptual and procedural fractions knowledge.

As we review in more detail in section ‘Background’, the educational technologies commonly known as intelligent tutoring systems (ITSs) typically decompose problems into sequences of steps and provide adaptive feedback. This is thought to primarily foster procedural knowledge, that is knowing how and when to apply a rule in order to solve a problem (Anderson, 1987; Mousavinasab et al., 2018; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001). Exploratory learning environments (ELEs), on the other hand, typically encourage the construction of knowledge and self-explanation through the manipulation of designed objects, tools and representations. This is thought to primarily foster conceptual knowledge, that is implicit or explicit understanding about underlying principles and structures of a domain (Rittle-Johnson & Alibali, 1999).

Both types of technologies, ITS and ELE, have important limitations. Early ITSs have been criticised for focusing excessively on automatizing procedures without ensuring an understanding of the underlying concepts, which may result in learners applying procedures inaccurately to problems based on shared surface elements (Jonassen & Reeves, 1996). ELEs, on the other hand, often fail to realize their promise because unguided or minimally-guided exploration places too-heavy cognitive demands on the learner (Kirschner et al., 2006).

Various attempts have been made to address these weaknesses. For example, some ITSs have been extended to include collaborative activities (e.g. Diziol et al., 2010), worked examples (e.g. Mathan & Koedinger, 2002), or reflective prompts (e.g. Rau et al., 2012) in order to promote sense-making and conceptual understanding. Conversely, for ELEs, Mavrikis et al. (2013) establish pedagogically-grounded requirements for providing intelligent support (i.e. guidance) while students undertake exploratory activities (Noss et al., 2012); while, more recently, Basu et al. (2017) investigated the use of adaptive scaffolding (i.e. an alternative approach to guidance) in learning-by-modelling tasks.

Each of these modifications aim to enhance one type of educational technology, an ITS or an ELE, so that the particular technology is better able to support the acquisition of both types of knowledge, procedural and conceptual. Against this background, in the studies reported in this paper, we asked whether the combined effect of the two types of educational technology, ITS and ELE, is greater than the sum of their individual effects. In other words, does combining the two technologies in one learning platform, leveraging both sets of individual strengths, have a synergistic outcome for both types of knowledge? We explored this question in the context of the interdisciplinary EU-funded project *iTalk-2Learn*, which developed an adaptive digital learning platform that enables the sequencing of content from two types of systems, an ITS and an ELE, based on the student’s

interactions including speech (see <http://www.italk2learn.eu>). To facilitate experimentation, *iTalk2Learn* focused on fractions, because this mathematical topic is known to be an important predictor for future mathematics performance (Siegler et al., 2012) and because its introduction in early years often poses challenges for students (Charalambous & Pitta-Pantazi, 2007).

In the following section, we discuss how these two types of educational technology, ITS and ELE, support either procedural knowledge acquisition or conceptual knowledge acquisition, respectively. We then present data from two quasi-experimental studies testing our research hypothesis using the *iTalk2Learn* platform, in Germany and the UK, and discuss the evidence that they provide for the effectiveness of combining exploratory learning with structured practice to foster robust learning.

## Background

### Intelligent tutoring systems

As the name suggests, intelligent tutoring systems (ITSs) are a type of educational technology designed for one-on-one, adaptive tutoring supported by feedback and hints (VanLehn, 2011). They typically involve a user-interface that presents students with instructional materials together with opportunities to answer structured questions, often breaking down each question into several steps to avoid student failure. A common classroom implementation of such systems involves them in blended learning instructional models with the aspiration of supporting personalised learning (Karam et al., 2016; Phillips et al., 2020). Of relevance to our research, one goal of such implementations is to support students practising procedural skills. Cognitive tutors are a particular type of ITSs (VanLehn, 2006) that support “guided learning by doing” with a mastery-based instructional approach (Kulik et al., 1990), and involve “model tracing” and “knowledge tracing”, two tutoring techniques that allow adaptive support of students’ learning. Model tracing assumes that a cognitive skill can be modelled as a set of independent if-then production rules, and supports students at the level of single problem-solving steps. Knowledge tracing tracks each individual student’s knowledge, in order to select which production rule and hence which task the student should experience next (Anderson et al., 1995; Koedinger, 2002; Koedinger et al., 1997).

One limitation of most ITSs is that they do not provide constructivist opportunities for the learner to self-construct knowledge, which are typically more effective than step-by-step learning for the development of conceptual knowledge (Doroudi et al., 2015). This has the risk of resulting in the student having little understanding of what is behind the procedures, how and why they work, and why one may want to learn them. There are, however, some ITSs that were specifically developed to foster sense-making, which will be discussed in section ‘Combining exploratory learning with structured practice tasks’. A second limitation of ITSs is that their adaptive support is usually based on pure performance indicators. The learners’ process or interaction data, on the other hand, is only rarely exploited to provide adaptive support within common ITSs (Mousavinasab et al., 2018). However, to identify individual learning needs and, thus, to allow for individually tailored learning, adaptive support should rely on a variety of different parameters. A third limitation of ITSs is that, drawing on the SAMR model of educational technologies (Hamilton et al., 2016), they typically only seek to “substitute” for standard teaching or to “augment” it. In other words, the ITSs usually operate only at the lower levels of the SAMR model,

reinforcing—while automating—step-by-step instructionist teaching practices. In particular, although ITSs are the archetypal “personalised learning” technologies, while they do personalise student learning pathways, they typically do not personalise student learning outcomes (they do not enable students to achieve their personal learning goals; Holmes et al., 2018).

### Exploratory learning environments

Exploratory learning environments (ELEs) are virtual environments that are designed to promote learning by discovery. They provide learners with opportunities to explore or experiment with a range of possibilities within a certain domain. There are different types of ELEs ranging from games to simulators, virtual labs, and open-ended learning environments, all of which usually target STEM subjects. In mathematics, in particular, ELEs enable the construction of some mathematical representation or abstract idea and are designed to empower learners to interact not only with the available objects, but also to explore their relationships and to investigate the underlying representations that enforce these relationships (Hoyles, 1993; Noss & Hoyles, 1996; Thompson, 1987). Using these tools, learners can explore mathematical objects from different but interlinked perspectives while the relationships that are key for mathematical understanding are highlighted, which helps the learner appreciate the various complexities. In this sense, ELEs can address the higher levels of the SAMR model (allowing for task redefinition and the creation of new tasks previously inconceivable) (Hamilton et al., 2016; Holmes et al., 2019).

Thus, ELEs may be a solution to a key limitation mentioned earlier of ITS, specifically their focus on procedural learning. However, the exploration at the heart of ELEs typically places high cognitive demands on the learner and, without guidance, may not be successful in fostering learning (Kirschner et al., 2006). To address this limitation, recent work has explored how to provide support in order to reduce the onerous cognitive demands experienced by some students in ELEs. In particular, intelligent components that provide feedback to support the student’s interaction with the learning environment, encourage goal-orientation, and exploit particular learning opportunities, have been incorporated in ELEs (e.g. Holmes et al., 2015; Mavrikis et al., 2013; Noss et al., 2012) and other open-ended environments (e.g. Basu et al., 2017; Bunt et al., 2004).

### Combining exploratory learning with structured practice tasks

With a few noteworthy exceptions (e.g. Rittle-Johnson et al., 2015; Star, 2005; Wang et al., 2013), the interdependence of procedural and conceptual knowledge has not received much attention. From a theoretical standpoint, the conceptual/procedural knowledge distinction is sometimes considered too coarse. For example, Star (2005) and de Jong & Ferguson-Hessler (1996) called for distinguishing between knowledge type and qualities, and Baroody et al. (2007) present a continuum of knowledge types and qualities, together with a further justification for their interdependencies. In short, in the field of mathematics education, the few studies that have examined the procedural/conceptual distinction have provided some useful evidence, but their application in educational technology contexts remains limited.

Similarly, research investigating the combination of exploratory learning and structured practice tasks is scarce. Either the combination of exploratory learning and structured practice tasks has not been compared to the separate approaches, or the outcomes were inconclusive (which might be due to the specific characteristics of the tasks or environments).

For example, Holmes (2013) designed a digital games-based learning environment for children who were low-attaining in mathematics. It offered opportunities for the children to self-construct solutions to authentic but covert numeracy problems (i.e. exploratory learning) followed immediately by structured practice to consolidate what they had learned. However, this work did not compare the outcomes of this combination with either exploratory learning or structured practice alone.

Corbett et al. (2013) report a study where students learned about genetics with an ITS. One condition provided a block of scaffolded reasoning problems aimed at eliciting sense-making, followed by a block where students solved problems to foster procedural knowledge. This combination led to better performance on transfer and preparation for future learning tasks than a condition where students only solved problems. However, the problem-solving condition led to better performance on problem-solving tasks than the combination condition.

Doroudi et al. (2017) report a study with *Fractions Tutor* comparing five conditions, out of which two are of particular interest in the context of this paper. One condition practised the application of procedures. A combination condition also solved problems aimed at sense-making and fluency-building. However, the authors did not find a significant performance difference between these conditions; possibly because problem selection was not adaptive but instead based on a spiral curriculum.

Finally, Rittle-Johnson and Koedinger (2009) investigated iterative lesson sequencing (lessons that alternate in focusing on concepts or procedures) within an ITS. They found that the iterative lesson sequence fostered procedural knowledge more effectively than a concepts-before-procedures sequence, and that there was no difference for the acquisition of conceptual knowledge. However, the lessons that focused on concepts were heavily structured and did not provide the affordances for discovery that ELEs typically offer.

In summary, while there is theoretical and some empirical support for combining exploratory learning with structured practice to promote both procedural and conceptual learning, the empirical evidence has been inconclusive. One reason may be that the hypothesis has previously been tested by combining exploratory learning with structured practice within a single educational technology. It has not yet been tested whether combining different but complementary educational technologies (ITS and ELE) that are designed specifically to support the acquisition of either procedural knowledge or conceptual knowledge can foster both types of knowledge. It was the goal of the work reported in this paper to investigate this possibility.

## Materials

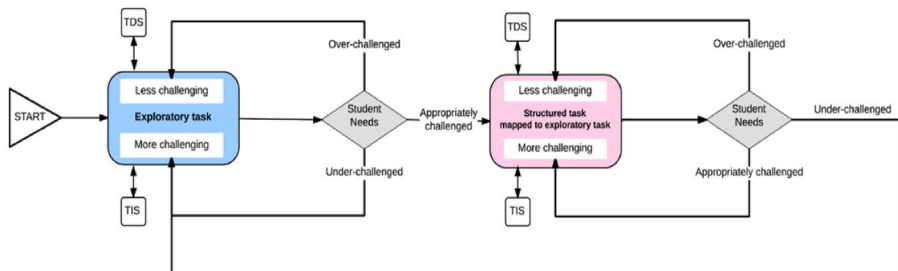
### iTalk2Learn platform

For the purposes of this research, the *iTalk2Learn* platform was configured in two parallel versions, English and German, both of which combined an ELE delivering exploratory tasks and an ITS delivering structured tasks. The English *iTalk2Learn* platform includes an ELE, *Fractions Lab* (Hansen et al., 2016), which was developed within the *iTalk2Learn* project, and a commercially available ITS, *Maths Whizz* ([www.whizz.com](http://www.whizz.com)). The German *iTalk2Learn* platform includes the same ELE, *Fractions Lab*, but translated into German, and an ITS called *Fractions-Tutor* (Rau et al., 2012, 2013), also translated into German. More details about these are provided below.

The pedagogy of the *iTalk2Learn* platform is based on an adaptive approach that uses a variety of inputs (e.g. screen/mouse action within the ELE, amount of feedback messages provided, and speech during reflective tasks) to sequence activities. In its default version *iTalk2Learn* combines the ELE and ITS activities. Building on previous research and theory in the field (Grawemeyer et al., 2017; Mazziotti et al., 2015), the pedagogical intervention model of *iTalk2Learn* specifies that students begin their session in the ELE, where they engage with an exploratory task (*c.f.* also a recent meta-analysis Sinha & Kapur, 2021 that favours engaging in problem solving followed by instruction akin to the model described here). While students undertake tasks, a *Student Needs Assessment* (SNA) component draws on the various inputs to determine whether the student is under-, over-, or appropriately challenged by the task and thus to identify the next task appropriate for them [based on an assessment of task difficulty by mathematics education experts and a set of rules, (Mazziotti et al., 2015)]. For example, if the student receives several supportive feedback messages, the system infers that the student is overly challenged, and they are given a less challenging exploratory task on the same concept. On the other hand, if the student is determined to be appropriately challenged, the *Student Needs Assessment* (SNA) component switches to the ITS where they are given a structured practice task (see Fig. 1). The first structured practice task that the student experiences in the ITS is mapped as closely as possible to the fine-grain goal of the completed task in the ELE (e.g. *fraction partitioning to find its equivalent*); while the next task in the ITS stays within the same fine-grain goal but increases the level of challenge (as determined by mathematics education experts). Students undertake a fixed sequence of ITS tasks until the *SNA* component determines that they are under-challenged, or until they have completed five structured practice tasks, whichever comes first, in which case they are returned to the ELE.

While engaging in the *iTalk2Learn* system, students receive *Task-Independent Support* (TIS), which is based on a Bayesian Network trained from past data. Positive affective states such as enjoyment are known to contribute towards constructive learning while negative ones such as frustration or boredom can inhibit learning (Kort et al., 2001). Accordingly, the *Task-Independent Support* aims to change a student's negative affective state into a positive affective state by adapting the feedback to the student's current affective state (which was inferred from the student's speech and interaction data such as whether or not feedback previously given had been followed). *Task-Independent Support* includes affect boosts (e.g. "Well done. You're working really hard!") and talk-aloud prompts (e.g. "Please explain what you are doing."). Grawemeyer et al. (2015, 2017) describe the *Task-Independent Support* in more detail.

The next section describes the learning environments in more detail, together with the tasks and the adaptive support provided by them. Tasks were chosen by mathematics



**Fig. 1** A visual representation of the pedagogical intervention model of *iTalk2Learn*

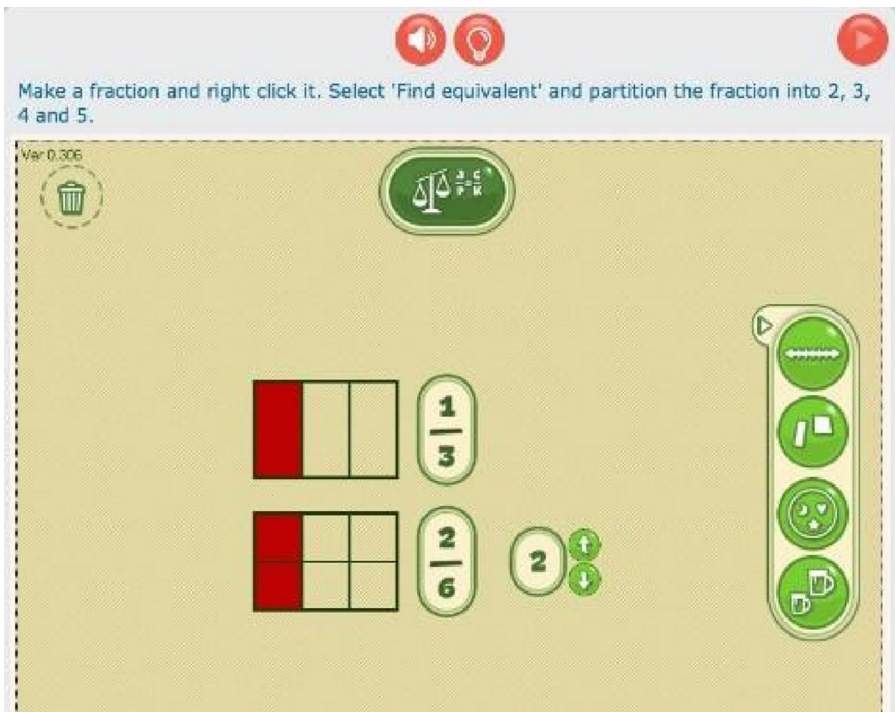


education experts, based on a mathematics education theory of fractions learning (Hansen et al., 2014) that takes into account misconceptions and errors that are typical for learners at the beginning of formal fractions instruction.

## Fractions lab

*Fractions Lab* is the ELE developed by the *iTalk2Learn* project, which was used both in Germany (in German) and the UK (in English). It provides exploratory tasks that aim to help the student develop conceptual knowledge of fractions. In the *Fractions Lab* interface (see Fig. 2), a learning task is displayed at the top of the screen. Students can choose from a range of graphical fraction representations (from the right-hand side menu)—number lines, rectangles, sets and liquid measures—which they manipulate in order to solve the given task. For example, they can change the fraction's numerator or denominator, and find an equivalent fraction. An example task is shown in Fig. 2, which served both to introduce the student to the available *Fractions Lab* functionality, and to introduce them to the idea of fractions equivalence with representations (Hansen et al., 2015).

To ameliorate the cognitive demands on the students associated with exploratory learning, *Fractions Lab* also provides students with *Task-Dependent Support* (TDS), in addition to the *Task-Independent Support* provided by the *iTalk2Learn* system, that varies by type and by method of delivery. The type of *Task-Dependent Support* is determined based on a rule-based system, operationalized according to two dimensions: the purpose of the feedback, depending on the task-specific needs of the student, and the level of feedback,



**Fig. 2** The ELE (exploratory learning environment) used both in the German study (in German) and in the UK study (in English, as shown): *Fractions Lab*

depending on the cognitive needs of the student (Holmes et al., 2015). Six feedback purposes were identified, each of which is triggered by a particular student response: *Polya* (understanding the problem, formulating goals and devising a plan, drawing on Polya, 1945), e.g. “Read the task again, and explain how you are going to tackle it.”; *instruction (next step)*, e.g. “You can use the arrow buttons to change the fraction.”; *instruction (problem solving)* (addressing misconceptions), e.g. “The denominator is the bottom part of the fraction.”; *instruction (opportunity for higher-level work)*, e.g. “You could now use the partition tool to make an equivalent fraction.”; *affirmation*, e.g. “The way that you worked that out was excellent. Well done.”; and *reflection*, e.g. “Please explain why you made the denominator 12.”

The second dimension, type of *Task-Dependent Support*, comprises four levels designed to address different levels of cognitive need (Holmes et al., 2015). As noted above, a particular student response in Fractions Lab triggers some feedback. Thereafter, if the same student response is repeated, the next level of feedback is triggered. The four levels of feedback are: *Socratic* (which emphasizes the benefits of open questioning to encourage students to think about and verbalize possible solutions), e.g. “Have you changed the numerator or denominator?”; *guidance* (to remind students of key domain-specific rules and the system’s affordances), e.g. “The denominator is the bottom part of the fraction.”; *didactic-conceptual* (a possible next step in terms of the fractions concept currently being explored), e.g. “Check that the denominator in your fraction is correct.”; and *didactic-procedural* (the next step that needs to be undertaken in order to move forward), e.g. “Check that the denominator, the bottom part of your fraction, is 12.” This rarely-delivered final procedural feedback operates as a backstop, ensuring that the student is not left floundering.

The method of *Task-Dependent Support* delivery is based on a Bayesian network, trained on previous data, that predicts whether the adaptation of the presentation of the feedback can improve a student’s affective state (Grawemeyer et al., 2015, 2017). In *Fractions Lab*, the feedback can be presented in either a low-interruptive way (by highlighting a light bulb at the top of the interface that indicates feedback is available that the student might or might not choose to access), or in a high-interruptive way (by providing a pop-up window that has to be dismissed before the student can proceed). The presentation of the feedback most likely to enhance the affective state of the student is inferred from the student’s current affective state and whether or not they followed the previous feedback.

## Maths Whizz

*Maths Whizz*, the ITS used in the UK study, is an English commercial system that provides mostly structured practice tasks (see Fig. 3). In *Maths Whizz*, each task is delivered in three stages. First, an instruction of how procedurally to complete the following tasks successfully; second, an interactive task with guided instruction and immediate feedback; and third a short test. The tasks use a range of graphical representations such as circles, rectangles, number lines, liquid measures, symbols and sets of objects within contexts that the students may be familiar with (e.g. a fairground). In addition to the *Task-Independent Support* provided by the *iTalk2Learn* system, *Maths Whizz* provides *Task-Dependent Support* when an incorrect answer is entered, in the form of a hint that encourages the student to elaborate and reflect on their problem-solving strategies before having another attempt (e.g. “Remember: you do not add the denominators. Add the numerators. Denominators stay the same”). Up to three hints are offered per question, at which point a student receives the correct answer. Correct answers are



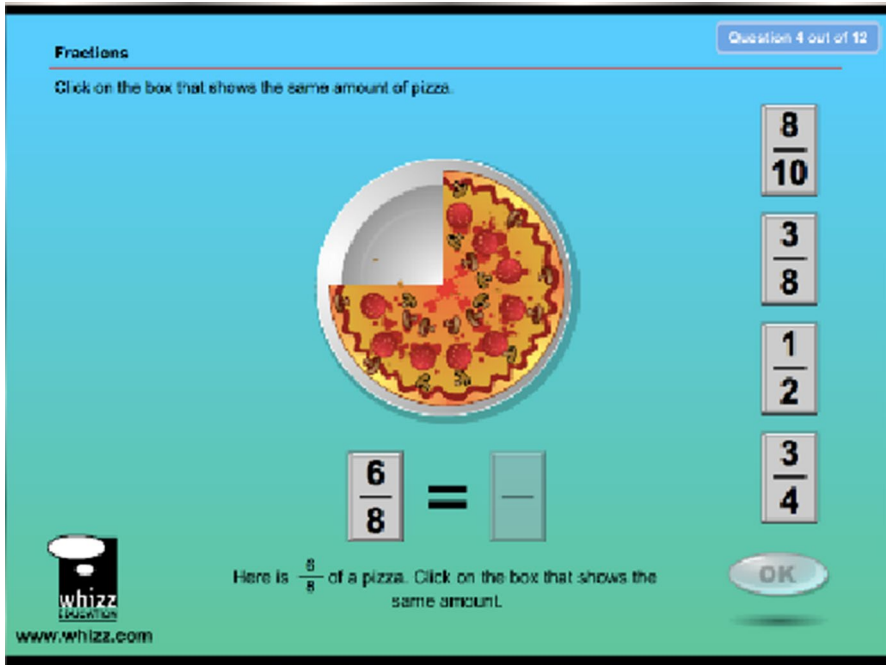


Fig. 3 The ITS (structured practice environment) used in the UK study: *Maths Whizz*

rewarded with a celebratory response. Following a set of tasks, a short test requires students to demonstrate their understanding without *Task-Dependent Support*, but with corrective feedback.

## Fractions Tutor

*Fractions Tutor*, the ITS used in the German study, is a web-based Cognitive Tutor for learning fractions (Rau et al., 2012, 2013) that enables students to solve fractions problems step-by-step, while receiving immediate feedback (on the steps) or asking for on-demand next-step hints. The version of *Fractions Tutor* used in this study had previously been translated into German.

Content is presented on the same page and revealed step-by-step while students solve the problem (for an example, see Fig. 4). The exercises use a range of graphical representations such as circles, number lines, and symbols. In addition to the *Task-Independent Support* provided by the *iTalk2Learn* system, *Fractions Tutor* functionalities allow students to ask for hints (i.e. to receive *Task-Dependent Support*) on up to three different levels: *clarification*, e.g. “Before you know what fraction of the whole cake you won, you need to divide the circle into equally sized pieces.”; *conceptual*, e.g. “The pieces are part of the same cake. Therefore, you keep the same denominator in the sum fraction.”; and *explicit instruction*, e.g. “Please divide the circle into four pieces.” (Rau et al., 2013).

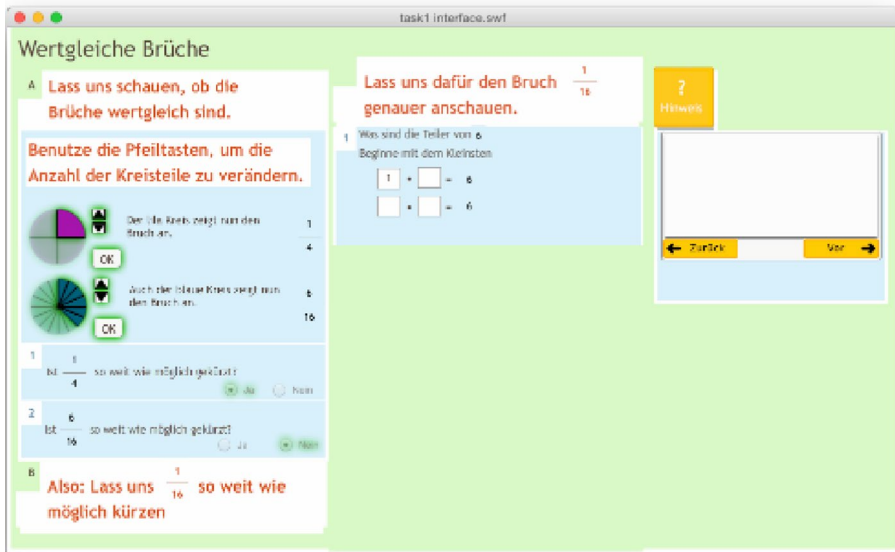


Fig. 4 The ITS (structured practice environment) used in the German study: *Fractions Tutor*

## Experimental design and participants

The two studies reported in this paper were undertaken in Germany and the UK, and focused on the learning of fractions. Both studies involved data from two experimental conditions:

- ITS & ELE: The full *iTalk2Learn* platform, incorporating both the structured practice (ITS) and exploratory learning (ELE) technologies.
- ITS only: The *iTalk2Learn* platform limited to the structured practice technology (ITS) only (i.e. no exploratory learning).

Participants in both countries were students who were just about to start, or were at the beginning of, formal fractions instruction. Fractions are taught earlier in the curriculum in the UK than in Germany, therefore the UK participants were slightly younger than the German participants. The parents or carers of participating school students provided informed consent for their child's involvement in the study; while, having been informed that they could withdraw from the study at any time without consequence and without having to give any reason, the students provided verbal assent for their involvement.

Participants in the study in Germany were fifth and sixth grade secondary school students aged between 10 and 12 years from four schools in suburban areas. Due to the readily observable differences in learning tasks between the conditions (i.e. that students would be able to observe what their near neighbours were doing), it was not feasible to run multiple conditions in the same classroom. Accordingly, the studies were run in a Pretest–posttest Non-equivalent Groups quasi-experimental design: students participated within their class, and classes within schools were randomly assigned to one of

the conditions. Class sizes varied, and, due to a technical failure, data was lost for one class of 33 students assigned to the *ITS only* condition, resulting in the following distribution across conditions:  $N_{ITS \& ELE} = 100$ , and  $N_{ITS \text{ only}} = 51$ .

Participants in the study in the UK were Year 4 and Year 5 primary school students aged between 8 and 10 years from three schools. The schools were from rural, suburban, and inner-city areas. Three groups per grade per school were randomly assigned to one of the conditions; while seven participating students did not complete the study and are excluded from the analysis. This resulted in the following distribution:  $N_{ITS \& ELE} = 61$  and  $N_{ITS \text{ only}} = 60$ .

### Dependent measures

Dependent measures were derived from an online fractions test, designed to differentiate between conceptual and procedural items (see Fig. 5), which was completed by the students before and after they interacted with the system. The test was administered to the UK students in English, while students in Germany received a German translation of the test.

Two isomorphic versions of the test instrument were designed (written in English and translated to German). Students were randomly allocated one version at the first time of measurement and the other version at the second time of measurement. Two subscales with three items each were constructed to measure procedural knowledge (see questions 22, 24, and 25 in Fig. 5) and conceptual knowledge (see questions 20, 21, and 23 in Fig. 5). The procedural knowledge items required simple computations using numerical representations of fractions, without the need to transition between different types of representations. They can be solved with a basic conceptual understanding of fractions (expanding fractions to share the same denominator). Conceptual items, on the other hand, can be solved without computations, but require an elaborated conceptual understanding of fractions: Students need to interpret non-numerical representations of fractions such as number lines or rectangles, transition between numerical and symbolic representations, and even compare different symbolic representations (e.g. question 23 in Fig. 5). The students received one point for each correctly-answered item and consequently obtained two aggregated scores, one

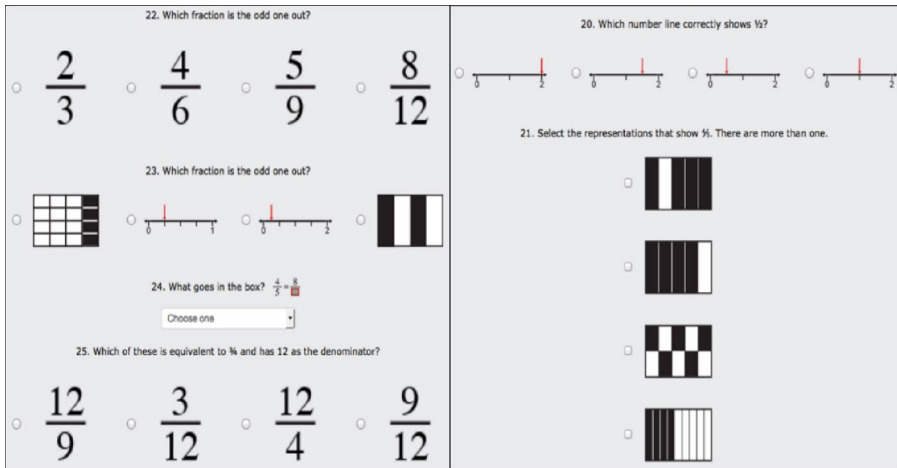


Fig. 5 Extract from the online fractions test (English version) undertaken by the participants

per subscale (i.e., scores are summed across three items per subscale and can vary between 0 and 3).

Internal consistency for the procedural scores at pre-test was  $\alpha_{UK} = .40$ ,  $\alpha_{Germany} = .07$ , and at post-test  $\alpha_{UK} = .53$ ,  $\alpha_{Germany} = .36$ . Internal consistency for the conceptual scores at pre-test was  $\alpha_{UK} = .40$ ,  $\alpha_{Germany} = -.03$ , and at post-test  $\alpha_{UK} = .36$ ,  $\alpha_{Germany} = -.06$ . Note that the  $\alpha$  values may be relatively low due to a combination of the small number of questions and the heterogeneity of the items (this issue is further considered in the discussion section below).

## Other instruments

The study also involved two questionnaires, each of which was administered in the appropriate language via a browser window: one on attitudes to learning, mathematics and fractions; the other a user-experience questionnaire. However, as neither questionnaire is included in the analysis reported in this paper, no further details will be given here.

## Procedure

Individual sessions were run with groups of up to 15 students in the UK, who interacted with the English version of the *iTalk2Learn* platform, and up to 30 students in Germany, who interacted with the German version of the *iTalk2Learn* platform. Half the groups in each country were allocated to the ITS only (structured practice only) condition, and the other half allocated to the ITS & ELE (structured practice and exploratory learning) condition. With the exception of the experimental condition and the language version of the platform, the sessions were the same for each group. In particular, learning, practising and testing time were held constant between groups.

The full session for all groups lasted approximately 90 min including breaks. During the first 10 min, the students were introduced to the study and to the *iTalk2Learn* platform with the ITS and ELE components being introduced depending on the experimental condition. To ensure that the introduction was as standardised as possible, it was scripted and delivered by the same researchers in both conditions. The students were then asked to complete the attitudes questionnaire (see section ‘Other instruments’ above) and then the online fractions test (see section ‘Dependent measures’ above), one after the other in a browser window. The students all completed the test within the given 10 min.

Students then worked with the *iTalk2Learn* platform for approximately 40 min. In the *ITS only* condition students received tasks based on a fixed sequence during this time (varying from 8 to 12 tasks in total). This sequence was similar to the ITS sequence in the *ITS & ELE* condition (but obviously with more procedural knowledge practice opportunities). In the *ELE & ITS* condition, tasks alternated between exploratory learning and structured practice tasks as described in section ‘Materials’ (again for a total of 40 min). Based on previous studies and further discussions with the teachers of the cohorts of this study, we had estimated that an average student would spend about half the time on the ELE tasks in this condition. Indeed, there was only a small variation across students (mean time on ELE tasks: 19.85 min; SD 0.004) with the rest of the time on ITS tasks (varying from 5 to 9 tasks in total).

During this main experimental period, the researchers adopted a strict intervention protocol that specified the allowable interactions and prompts. In particular, technical support was provided where needed, but no support was given for the fractions tasks. In the

last 30 min of the session, the students were asked to complete the final instruments. The online fractions test (see section ‘[Dependent measures](#)’ above) was presented followed by the user-experience questionnaire (see section ‘[Other instruments](#)’ above), one after the other in a browser window. Students were given twenty minutes in total for these two instruments, and all of them finished the test within time.

## Analyses

We investigated whether the combination of ITS and ELE can foster both procedural and conceptual knowledge by performing multivariate ANOVAs to account for the use of two dependent measures and possible alpha-error-inflation. To investigate differential effects of the conditions on procedural versus conceptual scores, we followed up with univariate ANOVAs. We did not treat country as an independent variable, but rather ran the analyses separately for the UK and Germany. All analyses were performed using SPSS v.27.

## Results

Table 1 presents scores on the online fractions knowledge test for the conceptual and procedural subscales. There was a medium correlation between these subscales on the post-test,  $r(151) = .25$  in Germany and  $r(121) = .26$  in the UK, both  $p < .01$ . In both countries, descriptively speaking, there were medium effects on both conceptual and procedural scores for the ITS & ELE condition, while in the ITS-only condition, effects on the conceptual scores were negative (see however the 95% confidence intervals which include zero) and effects on the procedural scores were low (Cohen, 1988).

Two-way, 2 (condition: *ITS & ELE* or *ITS only*)  $\times$  2 (time of measurement: pre-test or post-test) multivariate ANOVAs with repeated measures on the time variable and conceptual and procedural subscale scores as the two dependent measures were conducted for each country separately. We found significant effects for each factor and their interaction. Overall, analyses showed statistically significant learning gains from pre- to post-test, using Pillai’s trace for participants from both Germany (see Table 2).

**Table 1** Scores on online fractions knowledge test

Subscale	Country	Condition	Pre-test		Post-test		Effect size	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	95% CI
Conceptual	Germany	<i>ITS &amp; ELE</i>	0.79	0.74	1.21	0.69	0.59	[.30, .87]
		<i>ITS only</i>	0.73	0.63	0.53	0.64	−0.32	[−.71, .07]
	UK	<i>ITS &amp; ELE</i>	1.00	0.95	1.52	0.85	0.58	[.22, .94]
		<i>ITS only</i>	0.88	0.92	0.70	0.77	−0.21	[−.57, .15]
Procedural	Germany	<i>ITS &amp; ELE</i>	0.95	0.80	1.42	0.94	0.54	[.26, .82]
		<i>ITS only</i>	0.69	0.65	0.90	0.85	0.28	[−.11, .67]
	UK	<i>ITS &amp; ELE</i>	1.33	0.96	1.97	1.02	0.65	[.28, 1.01]
		<i>ITS only</i>	1.47	1.02	1.87	1.07	0.38	[.02, .74]

Scores are summed across three items per subscale and can vary between 0 and 3

**Table 2** Multivariate and univariate effects on procedural and conceptual scores

Multivariate effects	Germany			United Kingdom		
	<i>V</i>	<i>F</i> (2,148)	$\eta_p^2$	<i>V</i>	<i>F</i> (2,118)	$\eta_p^2$
Time of measurement	.117	9.834*	.117	.277	22.643*	.277
Condition	.132	11.274*	.132	.133	9.025*	.133
Time of measurement* condition	.109	9.068*	.109	.144	7.604*	.114
Univariate effects on procedural scores		<i>F</i> (1,149)			<i>F</i> (1,119)	
Time of measurement		18.552*	.111		16.337*	.265
Condition		10.618*	.067		.013	.000
Time of measurement* condition		2.552	.017		2.279	.019
Univariate effects on conceptual scores						
Time of measurement		2.206	.015		3.078	.025
Condition		16.465*	.100		13.999*	.105
Time of measurement* condition		16.697*	.101		13.245*	.100

\* $p < .001$ 

In summary, students in both countries and in both conditions showed learning gains, but importantly these learning gains were stronger for the *ITS & ELE* condition, with a medium effect size (Cohen, 1988), in both countries. This interaction is now investigated further for each subscale separately, using univariate ANOVAs.

Follow-up univariate analyses (Table 2) showed statistically significant learning gains on the procedural scores for participants from both Germany and the UK, but no overall learning gains on the conceptual scores for participants from either Germany or the UK. We could not detect a significant effect of the conditions on the procedural learning gain for participants from either Germany or the UK. However, on the conceptual scores (Table 2), analyses did show statistically significant effects of conditions on learning gains for participants from both Germany, and the UK.

In summary, while the results were similar in both countries, with students in both conditions showing significant learning gains on the procedural scores, only students in the *ITS & ELE* condition showed significant learning gains on the conceptual scores, a medium effect size (Cohen, 1988). The decrease in conceptual scores in the *ITS only* condition does not statistically differ from zero (the 95% confidence interval of the effect indicates that even a small increase is similarly likely).

## Discussion

Robust learning in mathematics depends on the acquisition of different types of mathematical knowledge (procedural and conceptual knowledge), each of which requires a different type of learning opportunity and support. Yet, learning systems developed for mathematics education are usually focused on only one type of knowledge, not on both: typically, they either focus on developing procedural knowledge by providing structured practice tasks (ITSs), or they focus on developing conceptual knowledge by providing exploratory tasks (ELEs). This limitation, providing only opportunities to learn one type of mathematical



knowledge, negatively impacts on robust learning and invites criticism of educational technology applied in mathematics education.

In the research reported in this paper, we investigated how this limitation may be overcome by combining the two types of educational technologies in one intervention: an ITS to support the acquisition of procedural knowledge, and an ELE to support the acquisition of conceptual knowledge. For the ITS, we used two state-of-the-art and well-established systems: *Fractions Tutor* (in Germany) and *Maths Whizz* (in the UK). For the ELE, we used *Fractions Lab*, which was developed especially for this project.

Our two studies (in Germany and the UK) both provided clear evidence that the combination of these two types of educational technologies, ELE (to foster primarily conceptual knowledge) and ITS (to foster primarily procedural knowledge) in one learning environment, promotes conceptual and procedural fractions knowledge more than ITS alone. In fact, despite the students in the combination condition (*ITS & ELE*) using the ITS for only around half the time that they used it in the *ITS only* condition, and therefore had less opportunities to repeat structured practice on certain topics, procedural learning was not compromised. Instead, in addition to gaining more conceptual knowledge of fractions, students who used both the ELE and ITS also gained more procedural knowledge of fractions.

Despite the contextual differences between the two studies, carried out in Germany and the UK (e.g. the different student ages and the different ITSs), the results were remarkably consistent across the two countries. This indicates, on the one hand, that our results are of sound external validity and, on the other hand, that the “combination effect” emerges unaffected by contextual factors.

Based on the outcomes of our study reported here, we can only speculate on the reasons for the “combination effect”. Possibly the “combination effect” is analogous to the “multiplier effect”, in which small changes in one factor can lead to disproportionate outcomes. In other words, perhaps the opportunities afforded to the student to *explore* fractions helped them better understand (‘multiplied’ their understanding of) the *procedures* that they were practising; while the *practice* helped them consolidate the *concepts*. This is in line with the iterative model of knowledge development, in which Rittle-Johnson et al. (2015) underline the reciprocal dependency of both types of knowledge, and thus warrants further research.

Although the clear procedural and conceptual learning gains observed in the *ITS & ELE* condition are promising, as in most studies conducted in naturalist contexts there are some limitations that need to be acknowledged. First, the contexts in which we deployed the two ITSs (*Maths Whizz* and *Fractions Tutor*) were quite different to those in which they are usually deployed, and the participating students had not worked with them before. Second, the intervention was of short duration, meaning that the participating students were given only a limited time to study very specific learning content. This was a consequence of conducting the studies in school classrooms for the purpose of increasing external validity, which placed constraints on the available intervention time.

Third, a more conceptual issue is the fact that it is challenging to measure the constructs of procedural versus conceptual knowledge independently (Jones et al, 2019; Schneider & Stern, 2010). This was particularly noticeable in the German sample, in which the conceptual scores were not internally consistent. This could be due to the low number of items, but also to the way items were constructed: solving items within one scale does not always require the same knowledge pieces; there is only a partial overlap. There may even be an overlap between scales: some basic conceptual knowledge was required to solve both the procedural and the conceptual items. This highlights the need to investigate the dimensional structure of procedural versus conceptual knowledge and to develop a valid and standardised measure. That said, reliability was large enough to detect an effect of

condition on the conceptual scores in Germany. So while there remains some ambiguity in what construct or constructs the scores are representing, the clear result patterns overall and their replication in two different countries do provide substantive first evidence of a noteworthy effect.

Fourth, we cannot be certain that the size of the combination effect was the same for both ITS. We did not investigate this because we cannot assume measurement invariance. Indeed, it is likely that cultural differences, and perhaps even the differences in student age, make direct comparisons between the ITS invalid. The evidence our study provided for a combination effect is therefore limited to a *replication of the effect* in two different contexts, not its *size*.

Future work, therefore, should look more into the components that make this ‘combination effect’ possible and follow-up questions that emerge from this study. For example, is the order of exploratory tasks followed by structured practice tasks essential for realising a worthwhile “combination effect”? A recent meta-analysis of problem solving followed by instruction seems to support this (Sinha & Kapur, 2021). But what would be the impact if the order were reversed? Similarly, further exploration is needed to explore emerging questions around the impact of time on task and optimal balancing of the sequencing suggested by the *Student Needs Assessment* (SNA) component, given students’ individual trajectories. In this study, we have kept the overall interaction time with the platform the same for experimental (internal validity) and practical (to fit with classroom timetabling) reasons. Thanks to the SNA and the feedback provision within each task, there was little variation overall in what the students covered in the given time within each condition. As such we did not treat time as an independent variable. Similarly, exploring the individual pathways was out of scope of this paper since the knowledge components covered by the different tasks were quite difficult to separate. However, in a longer intervention and with more topics to be covered the results could vary significantly.

Importantly, for a larger study it will be essential to develop ways to measure accurately conceptual understanding and procedural learning gains in this context (*c.f.* recent work that advocates comparative judgement as an instrument for measuring conceptual understanding in randomised controlled trials; Jones et al., 2019). Collecting data from an ELE only condition, perhaps with some other appropriate instruction to compensate for the inevitable lack of procedural knowledge, would also allow teasing apart whether the combination effect is not due only to practicing with the ELE but indeed from the combination of ELE and ITS.

## Conclusions

The study reported in this paper provides clear evidence that using two types of educational technologies (ITS and ELE) to combine in one intervention two types of educational tasks (structured practice and exploratory learning) in order to foster both procedural and conceptual knowledge, is effective and warrants further research. Furthermore, this “combination effect” stresses the need for fostering procedural and conceptual knowledge jointly, and supports the notion that, the two types of knowledge are reciprocally dependent (Baroody et al., 2007; Rittle-Johnson et al., 2015).

These findings also speak to broader debates in the field of educational technology, recast due to the growing attention being given to Artificial Intelligence (AI) and its advances. It should be noted that the ELE in this study was not what might be called a

“bare-bones” ELE (an ELE without any AI-driven support), but an ELE that incorporated by design a comprehensive system of automatic feedback i.e. feedback that responded automatically to student interactions. In any case, concerns about the role of AI, big companies and data in society apply to education as well (Williamson, 2019). ITS in particular are criticised as incorporating a retrograde pedagogy and operating only at the lower levels of the SAMR model, reinforcing the step-by-step instructional, behaviourist paradigm with limited student agency (Herold, 2017; Holmes et al., 2018, 2019)—a set of limitations that AI-supported ELEs at least partly address.

While our study has shown that the combination of exploratory learning supported by AI-driven feedback and structured practice can support classroom learning, the effectiveness or success of any classroom technology depends on the classroom pedagogy and how the technology is integrated (du Boulay, 2019), as well as on what is understood by “effectiveness” and “success” in educational contexts. As such, what remains to be discussed, and what should guide future work beyond validation or replication studies, is how these findings might have implications for classroom practices and, more broadly, the EdTech industry. Most commercially available educational technologies developed to support student learning are either ITSs by design or have mainly ITS features geared towards practicing procedural knowledge (Holmes et al., 2019). Notable exceptions for mathematics education (such as the exploratory environment Geogebra <https://www.geogebra.org>) require extensive support on behalf of the teacher. We postulate that the beneficial combination of structured tasks and exploratory learning might open up new possibilities for teaching and learning in class (i.e. it might transform existing practice) and, thus, might increase the likelihood of integration in the classroom. Teachers can take advantage of the combination effect in their classrooms by combining educational technologies or even engage their students with non-technology-based exploratory learning activities before using an ITS to consolidate what the students have learned. Such a possibility warrants further research.

In the meantime, our results also suggest that the EdTech industry might usefully either develop more standalone ELE technologies or incorporate ELE features in their existing products. Research has shown that teachers’ technology acceptance and adoption depend on how useful they perceive this technology in terms of both supporting students in their individual learning processes and achieving specific learning goals more effectively (e.g. Bray & Tangney, 2017; Hew & Brush, 2007; Holmes, 2013; McCulloch et al., 2018; Scherer et al., 2019). For example, in their interview study with early-career secondary mathematics teachers, McCulloch et al. (2018) found that teachers not only seek to provide additional opportunities for their students to practise mathematical procedures, but also aim to facilitate their students’ sensemaking of mathematical ideas, and, thus, they are open to using ELEs. However, if particular technologies are going to engage teachers’ interest and, hence, be used extensively in classrooms, perhaps in addition to surpassing teachers’ acceptance thresholds (being seen to benefit the students without impacting negatively on the teachers’ workloads), they also need to be seen to be somewhat exciting or at least intriguing. In other words, as suggested by Bray and Tagney’s guidelines (2017), the technology has to be potentially transformative of (rather than simply enhancing) existing classroom practices and student learning—as we have demonstrated to be possible with a judicious combination of AI-driven ELE and ITS.

**Acknowledgements** We would like to thank all our iTalk2Learn colleagues and partners for their contributions and support.

**Funding** The research reported here received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 318051—iTalk2Learn project.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Ethical approval** The research was approved by the Ethics Committee of University College London in UK and followed the ethical standards of the British Educational Research Association (BERA) and the German Psychological Society (DGPs).

**Informed consent** The parents or carers of participating school students provided informed consent for their child's involvement in the study; while, having been informed that they could withdraw from the study at any time without consequence and without having to give any reason, the students provided verbal assent for their involvement.

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

## References

- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review*, *94*(2), 192–210. <https://doi.org/10.1037/0033-295X.94.2.192>
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*(2), 167–207.
- Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). Research commentary: An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*(2), 115–131. <https://doi.org/10.2307/30034952>
- Basu, S., Biswas, G., & Kinnebrew, J. S. (2017). Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User Modeling and User-Adapted Interaction*, *27*(1), 5–53. <https://doi.org/10.1007/s11257-017-9187-0>
- Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research—A systematic review of recent trends. *Computers & Education*, *114*, 255–273. <https://doi.org/10.1016/j.compedu.2017.07.004>
- Bunt, A., Conati, C., & Muldner, K. (2004). Scaffolding self-explanation to improve learning in exploratory learning environments. In *International conference on intelligent tutoring systems* (pp. 656–667). Springer.
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology*, *39*(3), 521–534. <https://doi.org/10.1037/0012-1649.39.3.521>
- Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational Studies in Mathematics*, *64*(3), 293–316. <https://doi.org/10.1007/s10649-006-9036-2>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., & Baker, R. S. J. d. (2013). Differential impact of learning activities designed to support robust learning in the genetics cognitive tutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education* (Vol. 7926, pp. 319–328). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-39112-5\\_33](https://doi.org/10.1007/978-3-642-39112-5_33)
- de Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113. [https://doi.org/10.1207/s15326985ep3102\\_2](https://doi.org/10.1207/s15326985ep3102_2)
- Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89–102. <https://doi.org/10.1007/s10648-009-9116-9>

- Doroudi, S., Aleven, V., & Brunskill, E. (2017). Robust evaluation matrix: Towards a more principled offline exploration of instructional policies. In C. Urrea, J. Reich, & C. Thille (Eds.), *Proceedings of the fourth (2017) ACM conference on learning @ scale -L@S '17* (pp. 3–12). ACM Press. <https://doi.org/10.1145/3051457.3051463>
- Doroudi, S., Holstein, K., Aleven, V., & Brunskill, E. (2015). Towards understanding how to leverage sense-making, induction and refinement, and fluency to improve robust learning. In O. C. Santos, J. G. Boti-cario, C. Romero, M. Pecheniskiy, A. Merceron, P. Mitros, & M. Desmarais (Eds.), *Proceedings of the 8th international conference on educational data mining* (pp. 376–379).
- du Boulay, B. (2019). Escape from the Skinner Box: The case for contemporary intelligent learning environments. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.12860>
- Grawemeyer, B., Holmes, W., Gutiérrez-Santos, S., Hansen, A., Loibl, K., & Mavrikis, M. (2015). Light-bulb moment? Towards adaptive presentation of feedback based on students' affective state. In *Proceedings of the 20th international conference on intelligent user interfaces* (pp. 400–404). ACM. <https://doi.org/10.1145/2678025.2701377>
- Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction*, 27(1), 119–158. <https://doi.org/10.1007/s11257-017-9188-z>
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends*, 60(5), 433–441. <https://doi.org/10.1007/s11528-016-0091-y>
- Hansen, A., Mavrikis, M., & Geraniou, E. (2016). Supporting teachers' technological pedagogical content knowledge of fractions through co-designing a virtual manipulative. *Journal of Mathematics Teacher Education*, 19(2–3), 205–226. <https://doi.org/10.1007/s10857-016-9344-0>
- Hansen, A., Mavrikis, M., Holmes, W., & Geraniou, E. (2015). Designing interactive representations for learning fraction equivalence. In *Paper presented at the 12th international conference on technology in mathematics teaching* (pp. 395–402). Retrieved from [https://www.researchgate.net/profile/Alice-Hansen-4/publication/290324702\\_Designing\\_interactive\\_representations\\_for\\_learning\\_fraction\\_equivalence/links/569621d708ae425c6898b47e/Designinginteractive-representations-for-learning-fraction-equivalence.pdf](https://www.researchgate.net/profile/Alice-Hansen-4/publication/290324702_Designing_interactive_representations_for_learning_fraction_equivalence/links/569621d708ae425c6898b47e/Designinginteractive-representations-for-learning-fraction-equivalence.pdf)
- Hansen, A., Mavrikis, M., Holmes, W., Grawemeyer, B., Mazzioti, C., Mubeen, J., & Koshkarbayeva, A. (2014). *Report on learning tasks and cognitive models (iTalk2Learn deliverable 1.2)*. Retrieved from <http://www.italk2learn.com/deliverables-and-publications/deliverables/>
- Herold, B. (2017). The case(s) against personalized learning. *Education Week*, 37, 4–5.
- Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Education Technology Research & Development*, 55, 223–252. <https://doi.org/10.1007/s11423-006-9022-5>
- Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics*. Routledge. <https://www.routledge.com/Conceptual-and-Procedural-Knowledge-The-Case-of-Mathematics/Hiebert/p/book/9780898595567>
- Holmes, W. (2013). *Level up! A design-based investigation of a prototype digital game for children who are low-attaining in mathematics* (Unpublished doctoral dissertation). University of Oxford.
- Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). *Technology-enhanced personalised learning: Untangling the evidence*. Robert Bosch Stiftung GmbH. <http://www.studie-personalisierteres-lernen.de/en/>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education. Promise and implications for teaching and learning*. Center for Curriculum Redesign.
- Holmes, W., Mavrikis, M., Hansen, A., & Grawemeyer, B. (2015). Purpose and level of feedback in an exploratory learning environment for fractions. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Lecture notes in computer science. Artificial intelligence in education* (Vol. 9112, pp. 620–623). Springer International Publishing. [https://doi.org/10.1007/978-3-319-19773-9\\_76](https://doi.org/10.1007/978-3-319-19773-9_76)
- Hoyles, C. (1993). Microworlds/schoolworlds: The transformation of an innovation. In C. Keitel, & K. Ruthven (Eds.), *Learning from computers: Mathematics education and technology* (pp. 1–17). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-78542-9\\_1](https://doi.org/10.1007/978-3-642-78542-9_1)
- Jonassen, D. H., & Reeves, T. C. (1996). Learning with technology: Using computers as cognitive tools. In D. H. Jonassen (Ed.) *Handbook of research for educational communications and technology* (pp. 693–719). Association for Communications and Technology.
- Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal*, 45(3), 662–680. <https://doi.org/10.1002/berj.3519>

- Karam, R., Pane, J. F., Griffin, B. A., Robyn, A., Phillips, A., & Daugherty, L. (2016). Examining the implementation of technology-based blended algebra I curriculum at scale. *Education Technology, Research & Development*, 65(2), 399–425. <https://doi.org/10.1007/s11423-016-9498-6>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Koedinger, K. R. (2002). Toward evidence for instructional design principles: Examples from cognitive tutor math 6. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the annual meeting [of the] North American chapter of the international group for the psychology of mathematics education (24th, Athens, Georgia, October 26–29, 2002)* (Vol. 1–4, pp. 21–29). Retrieved from <https://files.eric.ed.gov/fulltext/ED471749.pdf>
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)*, 8, 30–42.
- Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE international conference on advanced learning technologies* (pp.43–46). <https://doi.org/10.1109/ICALT.2001.943850>
- Kulik, C.-L.C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265–299. <https://doi.org/10.3102/00346543060002265>
- LeFevre, J.-A., Smith-Chant, B. L., Fast, L., Skwarchuk, S.-L., Sargla, E., Arnup, J. S., Penner-Wilger, M., Bisanz, J., & Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology*, 93(4), 285–303. <https://doi.org/10.1016/j.jecp.2005.11.002>
- Mathan, S. A., & Koedinger, K. R. (2002). An empirical assessment of comprehension fostering features in an intelligent tutoring system. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Vol. 2363. Lecture notes in computer science, intelligent tutoring systems. 6th international conference, ITS 2002, Biarritz, France and San Sebastián, Spain, June 2–7, 2002: proceedings* (Vol. 2363, pp. 330–343). Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-47987-2\\_37](https://doi.org/10.1007/3-540-47987-2_37)
- Mavrikis, M., Gutierrez-Santos, S., Geraniou, E., & Noss, R. (2013). Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Personal and Ubiquitous Computing*, 17(8), 1605–1620. <https://doi.org/10.1007/s00779-012-0524-3>
- Mazziotti, C., Holmes, W., Wiedmann, M., Loibl, K., Rummel, N., Mavrikis, M., Hansen, A., & Grawemeyer, B. (2015). Robust student knowledge: Adapting to individual student needs as they explore the concepts and practice the procedures of fractions. In M. Mavrikis, et al. (Eds.), *Proceedings of the workshops at the 17th international conference on artificial intelligence in education* (Vol. 2, S. 32–40). Springer International Publishing.
- McCulloch, A. W., Hollebrands, K., Lee, H., Harrison, T., & Mutlu, A. (2018). Factors that influence secondary mathematics teachers' integration of technology in mathematics lessons. *Computers & Education*, 123, 26–40.
- Mousavinasab, E., Zarifसानای, N., Kalhori, S. R. N., Rakhshan, M., Keikha, L., & Saedi, M. G. (2018). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2018.1558257>
- Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers* (Vol. 17). Springer Science & Business Media.
- Noss, R., Pouloussillis, A., Geraniou, E., Gutiérrez-Santos, S., Hoyles, C., Kahn, K., Magoulas, G. D., & Mavrikis, M. (2012). The design of a system to support exploratory learning of algebraic generalisation. *Computers & Education*, 59(1), 63–81. <https://doi.org/10.1016/j.compedu.2011.09.021>
- Phillips, A., Pane, J. F., Reumann-Moore, R., & Shenbanjo, O. (2020). Implementing an adaptive intelligent tutoring system as an instructional supplement. *Education Technology, Research & Development*. <https://doi.org/10.1007/s11423-020-09745-w>
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press.
- Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction*, 23, 98–114. <https://doi.org/10.1016/j.learninstruc.2012.07.003>
- Rau, M. A., Aleven, V., Rummel, N., & Rohrbach, S. (2012). Sense making alone doesn't do it: Fluency matters too! ITS support for robust learning with multiple representations. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (Vol. 7315, pp. 174–184). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-30950-2\\_23](https://doi.org/10.1007/978-3-642-30950-2_23)



- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175–189. <https://doi.org/10.1037/0022-0663.91.1.175>
- Rittle-Johnson, B., & Koedinger, K. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. *British Journal of Educational Psychology*, *79*(3), 483–500.
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362. <https://doi.org/10.1037/0022-0663.93.2.346>
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. <https://doi.org/10.1016/j.compedu.2018.09.009>
- Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology*, *46*(1), 178–192. <https://doi.org/10.1037/a0016701>
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, *23*(7), 691–697. <https://doi.org/10.1177/0956797612440101>
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*. <https://doi.org/10.3102/00346543211019105>
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*, 404–411.
- Star, J. R., & Stylianides, G. J. (2013). Procedural and conceptual knowledge: Exploring the gap between knowledge type and knowledge quality. *Canadian Journal Science, Mathematics and Technology Education*, *13*, 169–181. <https://doi.org/10.1080/14926156.2013.784828>
- Thompson, P. W. (1987). Mathematical microworlds and intelligent computer-assisted instruction. In G. P. Kearsley (Ed.), *Artificial intelligence and instruction: Applications and methods* (pp. 83–109). Addison-Wesley Longman Publishing.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227–265.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Wang, M., Wu, B., Kinshuk, Chen, N.-S., & Spector, J. M. (2013). Connecting problem-solving and knowledge-construction processes in a visualization-based learning environment. *Computers & Education*, *68*, 293–306. <https://doi.org/10.1016/j.compedu.2013.05.004>
- Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology*, *50*, 1–16. <https://doi.org/10.1111/bjet.12849>

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

**Manolis Mavrikis** is Professor in Artificial Intelligence and Analytics in Education at UCL Knowledge Lab, and a Turing Fellow at The Alan Turing Institute. He holds an MSc with distinction in Informatics and Ph.D. in Artificial Intelligence in Education from the University of Edinburgh. His research interests and experience are in employing learning analytics to help teachers, schools, education ministries or researchers develop an awareness and understanding of the processes involved in learning, and on designing evidence-based intelligent technologies that provide direct feedback to learners, such as the work presented in this paper. Manolis is currently the Director of the Education and Technology master's at UCL and one of the editors of the British Journal of Educational Technology.

**Nikol Rummel** is a Full Professor and head of the Educational Psychology Lab in the Institute of Educational Research at Ruhr-University Bochum, Germany. She is also an Adjunct Professor in the Human-Computer Interaction Institute at Carnegie Mellon University, Pittsburgh, USA. Dr. Rummel was (2016–2017) president of the International Society of the Learning Sciences (ISLS). She is Associate Editor of the International Journal of Computer-Supported Collaborative Learning, and Editorial Board member of the Journal of the Learning Sciences, of the International Journal of Artificial Intelligence in Education, and of

Learning and Instruction. Her research published in numerous journal articles and book chapters focuses on developing and evaluating instructional support for learning in computer-supported settings, with an emphasis on CSCL and on adaptive learning support, in particular.

**Michael Wiedmann** (Dr. phil., University of Freiburg) is Research Manager for research and transfer at the University of Education Freiburg. In his role as a researcher in the Educational Psychology Lab in the Institute of Educational Research at Ruhr-Universität Bochum, Germany, he contributed to the summative evaluation in the iTalk2Learn project. He is interested in technology-enhanced learning, fostering excellent educational research and research utilization.

**Katharina Loibl** is Junior Professor in Interdisciplinary Research on Learning and Instruction at University of Education Freiburg. In her role as a researcher in the Educational Psychology Lab in the Institute of Educational Research at Ruhr-Universität Bochum, Germany, she contributed to the summative evaluation in the iTalk2Learn project and supported the data collection and analysis presented in this paper. She is interested in cognitive learning mechanisms of instructional designs, problem solving and inquiry learning, and diagnostic judgments of teachers.

**Wayne Holmes** (PhD, University of Oxford) is a learning sciences and innovation researcher who teaches at University College London, and is a consultant researcher on Artificial Intelligence (AI) and education for UNESCO and for IRCAI (the International Research Centre for Artificial Intelligence). Wayne takes a critical studies approach to the connections between AI and education, and their ethical and social implications. Recent publications include “Artificial Intelligence in Education. Promise and Implications for Teaching and Learning.” (2019), “Ethics of AI in Education: Towards a Community-Wide Framework.” (2021), and, for UNESCO, “AI and Education: Guidance for Policy-makers.” (2021).