Journal of Science Education and Technology
https://doi.org/10.1007/s10956-024-10197-7

Problem-Solving Strategies in Stoichiometry Across Two Intelligent Tutoring Systems: A Cross-National Study

Conrad Borchers[1] · Hendrik Fleischer[2] · David J. Yaron[1] · Bruce M. McLaren[1] · Katharina Scheiter[3] · Vincent Aleven[1] · Sascha Schanze[2]

Accepted: 26 December 2024

Abstract

Intelligent tutoring system (ITS) provides learners with step-by-step problem-solving support through scaffolding. Most ITSs have been developed in the USA and incorporate American instructional strategies. How do non-American students perceive and use ITS with different native problem-solving strategies? The present study compares Stoich Tutor, an ITS with a high level of scaffolding, with ORCCA, an ITS with dynamic scaffolds that can support a range of problem-solving strategies. We conducted a think-aloud study with university students in the USA ($N = 10$) and Germany ($N = 11$), where students worked with either Stoich Tutor and ORCCA before solving stoichiometry problems on paper. Two human coders derived a coding scheme to investigate the strategies American and German students employ during problem solving on paper without instructional support. We derive a taxonomy of three stoichiometry problem-solving strategies. Next to the American factor labeling method, this taxonomy includes a strategy based on equation transformations and a previously undocumented strategy using abstract symbols to isolate a target variable and then pluck in given values and compute the solution. German students exclusively used the latter strategy, which was not explicitly supported by any of the two tutoring systems. Further, students who did not use the factor-label method for paper-based problem solving, most of whom were German, initially had difficulty setting appropriate goals and working with fractions in the Stoich Tutor. While German students preferred ORCCA based on short interviews, they more often successfully solved problems in Stoich Tutor. Therefore, Stoich Tutor, although misaligned with German instruction, could still support German students' learning. Still, revisions to ITS based on local instructional cultures could make them potentially more effective and aligned with curricular goals.

Keywords Chemistry education · Intelligent tutoring systems · Scaffolding · Problem-solving strategy · Stoichiometry · Think-aloud protocols

Introduction

Acquiring domain-specific problem-solving strategies constitutes an important learning goal in STEM (Astuti et al., 2021; Priemer et al., 2020). Intelligent tutoring system (ITS) is a class of adaptive learning technologies that enhance strategy acquisition multi-step problem solving (Chi & VanLehn, 2010). By providing step-level instruction and supporting problem solving through as-needed hints, ITS can improve STEM learning (Aleven et al., 2017; Kulik & Fletcher, 2016). ITS constrains the inputs a student can perform and often provides an interface resembling a template for approaching problems. These interfaces support the acquisition of domain-relevant strategies (Shakya et al., 2022; Tenison & MacLellan, 2014).

Most ITSs have been developed in in the USA (Aleven et al., 2009; Azevedo et al., 2009; King et al., 2022; McLaren et al., 2011a; Pardos et al., 2023), thus reflecting American instructional and problem-solving strategies. One exception is the German ActiveMath ITS (Melis & Siekmann, 2004). An example of an American cultural signature in ITS design is the factor label method typical for US education (Schmidt, 1997) and pivotal in the design of Stoich Tutor (McLaren et al., 2011a, b). To the best of our knowledge,

✉ Conrad Borchers
cborcher@cs.cmu.edu

1 Carnegie Mellon University, Pittsburgh, PA, USA

2 Leibniz Universität Hannover, Hannover, Germany

3 University of Potsdam, Potsdam, Germany

it has not been investigated how scaffolding in ITS for a predominantly American context might elicit differences in problem-solving strategies for European students. There is a research opportunity to study how students educated outside of the US perceive and use ITS, who might approach problem-solving with different native strategies. Such an investigation could shed light on what re-design to existing ITS is needed to accommodate differences in instructional cultures. To measure native strategies, we analyze German and American students' problem-solving with pen and paper. By focusing on paper-based problem-solving, we can identify the types of strategies that students employ without any instructional support, representing curricular standards and defaults across national contexts. Grouping students depending on the expression of these strategies, we then analyze these student groups' use of two American ITS.

The present study compares tutors with different degrees of scaffolding in stoichiometry, a domain where past work has noted diversity in problem-solving strategies and student difficulties in strategy acquisition (BouJaoude & Barakat, 2003; Mandina & Ochonogor, 2018; Schmidt & Jignéus, 2003; Tóth & Sebestyén, 2009). We compare Stoich Tutor (McLaren et al., 2011a, b), an ITS rich in scaffolding for the factor labeling method typical for US education and not part of most European instruction (Schmidt, 1997), and ORCCA (King et al., 2022), an ITS enabling open-ended problem-solving alongside dynamic scaffolds that accommodates diverse strategies beyond American contexts, such as the mole method and the proportionality method (Schmidt, 1997; Tóth & Sebestyén, 2009). We conducted a think-aloud study involving university students from the USA ($N=10$) and Germany ($N=11$), with students starting with either Stoich or ORCCA depending on their assigned condition. We investigate three research questions (RQs):

- RQ1: How do problem-solving strategies during paper-based problem solving without any instructional support compare between German and American students?
- RQ2: Depending on the problem-solving strategies expressed during paper-based problem solving, how do these students use an open-ended vs. structured problem solving tutoring system developed in a US context?
- RQ3: How do perceptions of an open-ended vs. structured problem solving tutoring system developed in a US context compare between German and American students?

Background and Related Work

Strategy can be defined as a sequence of problem solving operations that are selected adaptively and flexibly to solve a problem (Siegler, 1991; Star & Rittle-Johnson, 2008). Successful strategy implementation requires domain knowledge, including valid operations, symbolizing, and vocabulary. Opportunities to learn from problem solving can foster conceptual understanding in STEM (Heyworth, 1999). To this end, prior work has developed instructional support during problem solving, or *scaffolding*, that provides learners with additional structure and constraints during problem solving, allowing them to acquire desirable problem-solving strategies (Clark & Mahboobin, 2017; Yuriev et al., 2017). As they constrain, support, and shape the types of strategies students employ, meta-analyses deem computerized scaffolding a suitable instructional element to improve STEM learning (Belland et al., 2017).

Scaffolding and Strategy Support

Scaffolds must meet learners' current needs to be effective. Mismatched scaffolds can slow down learning if they remove opportunities to learn. Too much scaffolding can prohibit the learner from performing problem solving steps that pose productive opportunities to learn from errors (for example, when axes during 2D plotting are prelabeled; Baker et al., 2002). Conversely, offering too little scaffolding, for example, during open-ended problem solving, can also take away learning opportunities from novices by being too challenging to produce meaningful solutions or self-explain (Borchers et al., 2023). Balancing the right amount of assistance and scaffolding is a fundamental design challenge in ITS and instruction, generally (Koedinger & Aleven, 2007).

Another, although less well-researched, way in which strategy support can be suboptimal is if native and routinely used problem-solving strategies of a student are not supported by a scaffold. Technology-based problem solving environments constrain the number of viable problem-solving strategies by the range of inputs and interface elements and, secondarily, the types of problem-solving strategies a system providing adaptive instruction can recognize. Learners being used to differing predominant cultures of instructions can hence constrain instructional effectiveness in such systems. Cultural differences in instruction have been documented based on curricular analyses in algebra (Leung et al., 2014) and stoichiometry (Molnár & Molnár-Hamvas, 2011; Schmidt, 1997) as well as survey-based comparisons between American and Turkish students in approaches to physics problem solving (Balta et al., 2016), among others. However, none of these studies has investigated how problem-solving strategy differences interact with technology-based scaffolds or problem solving interfaces developed in a predominantly US-American context. Specifically, it is an open question whether providing a highly scaffolded

ITS that supports a foreign (hence unfamiliar) problem-solving strategy is more desirable than an open-ended ITS that supports multiple strategies, including familiar ones, though with a lower level of scaffolding.

Problem-Solving Strategies in Stoichiometry

Stoichiometry is a subdomain of chemistry dealing with the quantitative relationship between the reactants and products in chemical reactions (Niaz & Montes, 2012). In summarizing known strategy differences between American and European students, we review prior research that primarily used textbooks as canonical representations of what teachers are likely to teach in class, and strategies that students in a specific region are likely to have adopted. Schmidt (1997) identified two strategies based on a review of textbooks in the USA, UK, and Germany. The first strategy (mole method) is documented in international literature for expert audiences but rarely in textbooks for students. The mole method involves relating substance masses and molar masses to one another to then work out the relationship between the given and target variables. The second strategy (proportionality method) is taught in the US and Germany and is based on using proportionality (i.e., fraction relationships) to derive a target variable. However, the proportionality method takes on different forms in American and German textbooks. The factor label method, used and taught in the USA, involves (a) identifying a series of ratios, each representing a unit conversion, (b) finding the answer by multiplying them, and (c) canceling units. Open educational resources in the USA, representative of US education, offer some examples of that method. For example, to calculate the number of atoms from a given mass of a substance, the factor labeling method instructs students to construct a set of multiplications by the reciprocal of the molar mass from left (given value and unit) to right (target unit). The units and substances are then canceled to derive the final value (Flowers & Theopold, 2019).

Documentation of how the proportionality method is taught in German textbooks or applied by German students is lacking. However, related work in physics education, which involves similar forms of reasoning with units and quantities, indicates that textbooks instruct students to transform abstract quantity relationships to then substitute givens to derive a target variable (Dröse & Prediger, 2018). Curricular standards and example exercises provided by the various ministries of education of the German federal system, also document this strategy (e.g., https://cms.sachsen.schule/msw/unterricht/physik/loesen-von-physikalischen-sachaufgaben.html). Taken together, the key differences in known problem-solving strategies based on instruction between the US and Germany are the expression of the proportionality method: The predominantly taught problem-solving strategy in the US is the factor label method (1997), while German students, at least in physics, are taught to transform quantity relationships through abstract representations (Dröse & Prediger, 2018). Given this discrepancy, German students' perceptions and use of tutoring systems strongly structuring problem solving around the factor label method may differ between that of American students. To the best of our knowledge, no further investigation has documented differences in problem-solving strategies between American and German students. There is a lack of (a) integrated previously identified strategies observed in student responses to chemistry problems to instructional strategies in textbooks and (b) an investigation of how students' strategies relate to scaffolding for strategies, including in adaptive learning technologies.

Intelligent Tutoring Systems for Stoichiometry

Two ITSs for stoichiometry have been developed to date and are subject of the present study: Stoichiometry Tutor (Stoich Tutor) and the Open-Response Chemistry Cognitive Assistant (ORCCA). The Stoich Tutor has been shown to improve the stoichiometry performance of high school students (McLaren et al., 2016).

The Stoich Tutor instructs students to solve stoichiometry problems by reasoning from a given value to a target value through a structured, fraction-based approach. That is, the tutor essentially uses the factor labeling method discussed earlier (Flowers & Theopold, 2019). The ORCCA Tutor is a rule-based intelligent tutoring system (ITS) for chemistry, meaning that it compares problem solving rules with problem solving sequences demonstrated by the student, allowing for flexible problem-solving strategies through a formula entry interface and space for notes (King et al., 2022). ORCCA's instructional effectiveness has not been formally investigated at the time of this study.

Both ITSs tend to support different strategies for solving problems in stoichiometry. We highlight three differences between both ITSs. First, the Stoich Tutor instructs students to think about the target variable and given values and to then reason from the given values to the target value. This is apparent through the Stoich Tutor's interface, constrained to force construction of a series of ratios that are multiplied in sequence. Stoich Tutor further supports initial reflection about the target and given values through the sequence in which it provides hints. In contrast, ORCCA's hint sequence always suggests the next transformations of the given values toward the problem solution rather than initially specifying a target value. Further, ORCCA's scaffolding through hints is dynamic and provides students with a generic hint regarding the best next step in the problem solving sequence (e.g., "your next step should be determining the volume of the solution in L"), allowing students to enter upfront transformations or compound steps.

Second, the Stoich Tutor instructs students to solve problems using fractions and related unit conversions, which closely aligns with the factor labeling method typical for the American instructional context (Flowers & Theopold, 2019; Schmidt, 1997). These unit conversions are explicit, must be entered, and must be canceled. It is worth noting that units are an important instructional goal in stoichiometry to foster conceptual knowledge (Hafsah et al., 2014). In contrast, using units in ORCCA is optional, and students generally do not receive feedback on the accuracy of their unit use when transforming equations except when they are not viable for a given problem.

Methods

Data Collection

Sample

Twenty-one participants took part in this study, including ten students in the USA and eleven students in Germany. Data collection took place between February and June of 2023. The average age of the participants was $M = 21.57$ years ($SD = 3.21$). All German students participated in-person and were enrolled in a large public university. Five US students were recruited at a large private research university (and participated in person). The other five US students were recruited from a large public university (and participated via Zoom). All students were recruited through course-related information channels by instructors in courses known to include students still learning stoichiometry. Further, students were asked to circulate recruitment materials to peers and encourage them to do the same. Participants received $15 Amazon gift cards (US sample) or 15 Euros in cash (German sample) as compensation.

Materials

The participants worked with two different tutors: Stoichiometry Tutor (Stoich Tutor) and the Open-Response Chemistry Cognitive Assistant (ORCCA). The current version of the Stoich Tutor has been used to include previous adaptations made during design studies (e.g., polite language; McLaren et al., 2011a, b). A total of 14 stoichiometry problems were taken from earlier studies with the Stoich Tutor with higher scaffolding (McLaren et al., 2016) and adapted to the ORCCA ITS with lower scaffolding. Both tutoring systems are available online upon request.[1]

The translation of both ITSs from English to German was performed by taking each individual piece of text (e.g., a problem statement, hint, or error feedback) from both systems, automatically translating them with the DeepL API (DeepL, n.d.) and manually adjusting translations by (a) defining a dictionary of common translations that are specific to the domain of stoichiometry and chemistry to be parsed into the DeepL API (e.g. *substance* is translated to the technical term *Stoff* instead of *Substanz*, which is the commonplace translation of substance) and (b) having two experts in chemistry fluent in English and German (a PhD student and Professor in that instructional science) adjust any lacking translations.

All participants' interactions with the ITSs were logged as transactions (e.g., problem solving step attempts, hint requests, and selection of units) to the PSLC DataShop (Koedinger et al., 2010), following best practices for such log data. Think-aloud utterances were recorded via a MacBook laptop, including screen captures of participants' problem solving process in the ITSs. Both interfaces are exemplified in Figs. 1 and 2.

Procedures

The study lasted between 45 and 60 min per student. First, German and US students completed a survey about their demographic information and prior proficiency (e.g., completed courses and grades in chemistry). To avoid bias in student problem solving based on their exposure to specific ITS and problems, we counterbalanced three factors, leading to eight problem sampling paths: (a) whether the students were working with the ORCCA or the Stoich Tutor first, (b) which content unit (moles and gram conversions vs. stoichiometric conversion) was practiced first, and (c) the order of items within units (Fig. 3). The first two sample paths (1 and 2) represent solving problems involving moles and gram conversion, which is signified through the letter A, or unit A. The only difference is that path 1 starts with Stoich Tutor, while path 2 begins with ORCCA. Similarly, the next two paths (3 and 4) involve solving stoichiometric conversion problems, which is signified through the letter B, or unit B. Paths 5–8 are equivalent to paths 1–4 except for the order of problems being reversed (as signified by problem numbers following letters A and B). Paths 1 and 2 were traversed by four students and path 3 by three students. Paths 4–8 were each processed by 2 students each. Although each of the eight paths included only between two and four students, their design aimed to

[1] https://stoichtutor.cs.cmu.edu/ and https://orcca.ctat.cs.cmu.edu/

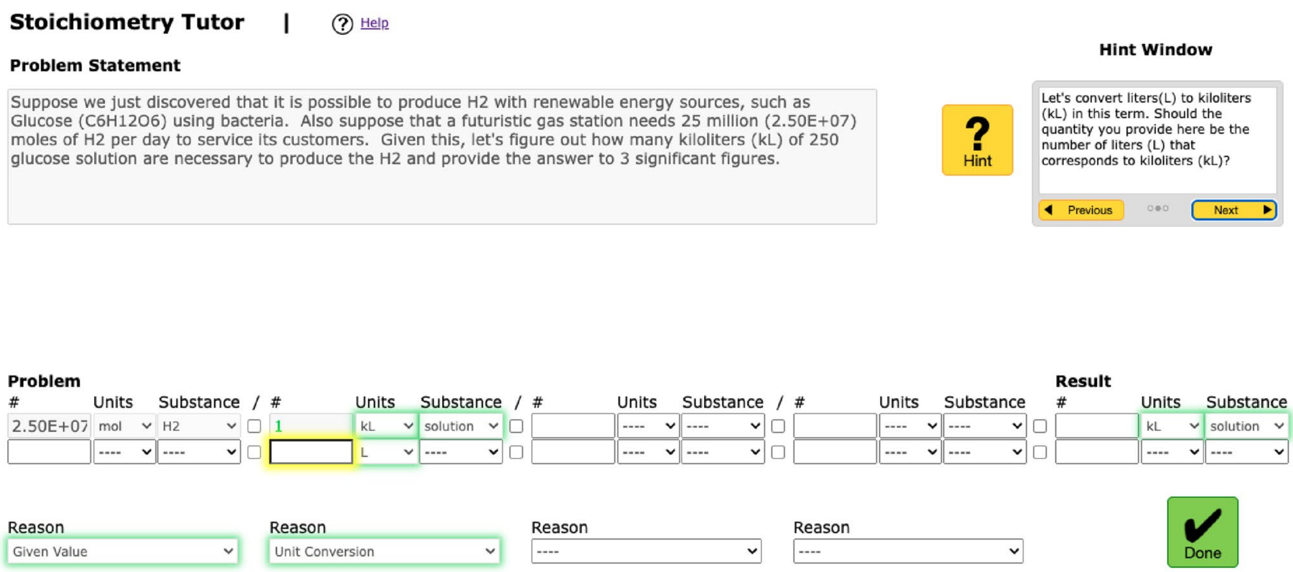

Fig. 1 Example problems representing students' experience working through stoichiometry problems with the Stoich Tutor

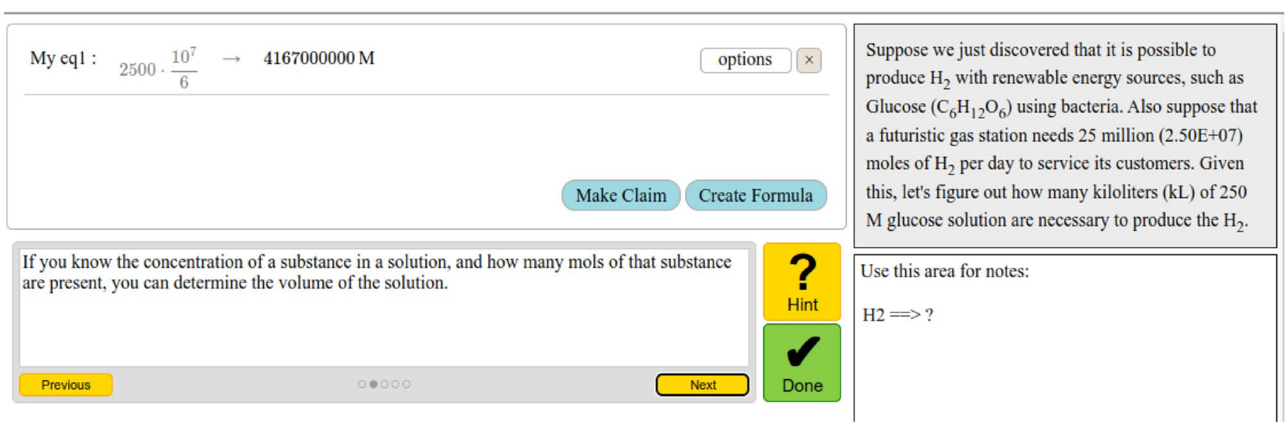

Fig. 2 Example problems representing students' experience working through stoichiometry problems with the ORCCA Tutor

minimize bias due to factors that could influence the qualitative data collected during think-aloud sessions. Specifically, exposure to one ITS might influence perceptions and usage of the other ITS, and variations in problem order and content units could affect students' familiarity and fluency when working on subsequent problems.

The students watched a short ITS video tutorial and were introduced to the think-aloud method by an experimental conductor. The students worked on four problems, two completed with the tutor and another two on paper. Students worked with the tutor first to prioritize data collection of tutor interactions given the limited session time. If there were still 15 min or more left (within sessions of up to 60 min), the participants were offered the opportunity to try out the second tutor, working on up to two problems. These problems were taken from the respective other content unit. A total of seven students worked with both ITSs. Of these, three are German and four are US participants (Fig. 3).

Finally, a semi-structured interview was conducted, which lasted 5–10 min. Participants were asked to discuss positive and negative aspects of each ITS and their fit to their individual's problem-solving strategy. If participants worked with both ITSs, they were asked to explicitly contrast them.

Data Processing

Coding of Problem-Solving Strategies During Paper-Based Problem Solving (RQ1) RQ1 considers what stoichiometry problem-solving strategies German and American college students use during paper-based problem solving without any instructional support. To answer this question, we derived a coding scheme for strategies on paper responses,

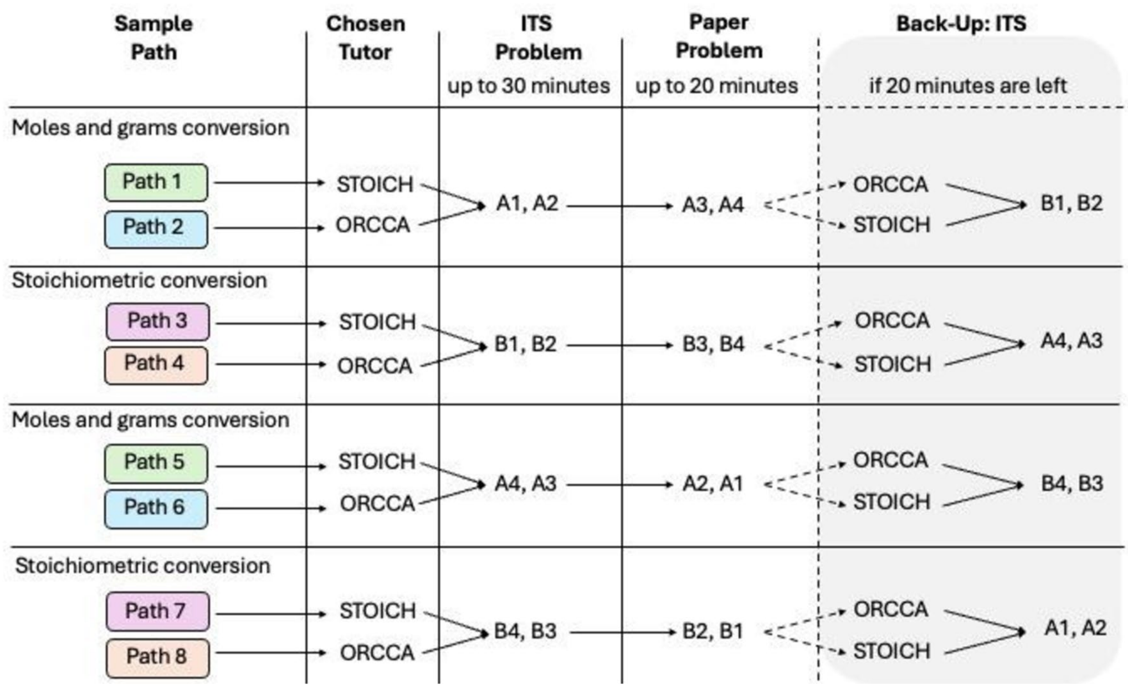

Fig. 3 Schematic representation of the study process with eight problem pathways depending on the tutors

we followed an expert-based approach, consulting instructional design literature and official curricular guidelines and open educational content in Germany and the USA. We consulted literature on problem-solving strategies in stoichiometry (Schmidt, 1997), but also in physics, which often deals with similar problems, requiring the transformation, conversion, and computation of problems with multiple units and relationships (Dröse & Prediger, 2018) and in mathematics, including its notations (Posamentier & Krulik, 2008). We further consulted with developers of both ITSs to consider the instructional design and differences between both ITSs.

Two coders familiar with both tutoring systems discussed and derived a set of coding rules for three freeform problem-solving strategies on paper based on 4–6 example cases (Appendix 1). Both coders then independently coded each of the 34 paper artifacts from 18 students in our study sample according to these rules. Both coders only disagreed on one artifact, yielding an excellent inter-rater agreement as expressed in Cohen's Kappa of $\kappa = 0.955$. One of the two coders then coded four additional paper problem solving responses collected later on.

Audio Transcription of Think-Aloud Utterances and Short Interviews The American think-aloud and interview recordings were transcribed using Whisper, an open-source transcription model for voice (Radford et. al, 2023). While Whisper can transcribe German, we deemed it unsuitable for creating transcripts for this study due to lower face accuracy and audio quality in the German recordings. Hence, the German recordings were segmented and transcribed manually using the FOLKER program following transcription rules of the "Gesprächsanalytisches Transkriptionssystem 2" (GAT2) (Selting et al., 2011). Segmentation was based on (a) speaker changes between user and interviewer, (b) pause lengths between words and phrases, and (c) changes in content, such as between problems.

Thematic Analysis of Student Use and Perceptions of the Tutoring Systems (RQ2 and RQ3) To analyze student use differences of the tutoring systems depending on their paper-based problem-solving strategy (RQ2), we qualitatively analyzed how students initially approached problem solving in each ITS, with two research team members coding think-aloud utterances based on reviews of screen capturing recordings through the same methodology as student interview responses. We sampled utterances from the first 5 min of students working with a tutoring system for the first time, given that inspecting all utterances in our sample was not feasible and that these utterances are likely to reveal discrepancies between internalized strategies and tutoring system use while still learning how to use each ITS. One research assistant fluent in English and German first familiarized themselves with all transcripts, excluded utterances irrelevant to participants' problem solving with two ITSs, and sorted all utterances by ITS and sample. Then, the coder grouped individual utterances into themes, which were then consolidated and discussed with two Ph.D. students on the research team to reduce bias.

To better understand differences in German and American students' perceptions of open-ended vs. structured tutoring systems developed in a US context (RQ3), transcripts of the short interviews regarding the users' impressions of each ITS were analyzed through an open coding scheme to derive themes (Neuendorf, 2018). The aim was to analyze impressions regarding the perceived fit of the ITS to users' problem-solving strategies.

Data Analysis Methods

For RQ1, we investigate differences in the employed problem-solving strategies of German and American students on paper. We report characteristics of each strategy through example cases. We also report how commonly these strategies were employed by students in both national samples, with each student being assigned a unique strategy. Assignments of freeform strategies to students were based on the first paper-based problem response, which differed from the classification of their second problem for seven (33%) students. This decision was based on the observation that five of these students started with the more structured strategy and then exhibited a more free-form strategy at their second problem, which could be due to increased fluency on similar problems, aborted problem solving due to time constraints, or overconfidence and boredom during repeated problem solutions. In these cases, students' first problem response might be more representative of their natural problem-solving strategy absent instructional support.

For RQ2, we relate a student's strategy choice when working on paper to their problem solving when working with an open-ended vs. structured tutoring system. In particular, we take into account students' first problem completed with each tutoring system to analyze their initial approach to using each system and report qualitative themes found in their think aloud utterances and screen recording. Matches between each student's freeform strategy and the two ITSs are reported based on considerations of the form of scaffolding and instruction of each system. To confirm our assumption that free-form strategies on paper were independent and not influenced by the platform students practiced with before problem-solving on paper (at least in terms of the type of strategy expressed), we conducted a χ^2-test of independence between the ITS students worked with before paper-based problem solving and their strategy expressed on paper. The independence assumption was not rejected, $\chi^2(2)=1.83$, $p=0.400$.

Third, themes arising in short interviews related to the perception of German and American students regarding both tutoring systems (RQ3) were grouped and reported by national sample.

Finally, to descriptively describe student-ITS interactions, we generated standard summary statistics with each student-problem pair as the unit of analysis (Koedinger et al., 2010). We computed the percentage of correct attempts in the system (representing problem solving step attempts, for example, setting the right unit of a denominator in the Stoich Tutor). This computation was performed on first step attempts only, which contains most of the diagnostic information about student knowledge, as students receive tutor feedback on each attempt. We also computed how many problems were completed (excluding problems started in the back-up ITS, which would bias the statistic due to the limited time given for these problems).

Results

Descriptive Statistics

Students completed 28 out of 64 of the started problems successfully (43.8%). This ratio was comparable for the Stoich Tutor (14 out of 30; 46.7%) and ORCCA (14 out of 34; 41.2%) but not across samples, where more US students completed problems (19 out of 27; 70.4%) compared to German students (9 out of 37; 24.3%). Descriptive differences of students' problem solving in both ITSs are in Table 1.

Accuracy on first attempts was overall comparable between conditions. Students completed around twice as many problem solving steps in the Stoich Tutor compared to

Table 1 Quantitative differences of students' solving problems between samples in the Stoich Tutor and ORCCA, with (completed) indicating successfully completed problems only. Median and *IQR* were used due to a right-skewed time distribution

	Germany	USA
% Correct first attempt	70.3% Stoich/ORCCA: 66.3/74.6%	64.4% Stoich/ORCCA: 65.9/62.7%
Mean N problems (completed)	0.82 Stoich/ORCCA: 1.3/0.6	1.46 Stoich/ORCCA: 1.3/1.8
Mean N steps (completed)	16.8 Stoich/ORCCA: 26.0/9.4	21.7 Stoich/ORCCA: 28.1/14.6
Median (*IQR*) min (completed)	6.93 (6.10) Stoich/ORCCA: 10.8 (3.52)/3.5 (1.93)	4.63 (6.62) Stoich/ORCCA: 4.82 (8.39)/4.63 (3.78)

ORCCA. This means that the number of steps mainly reflects the differences in the way the two tutors are designed, for example, the Stoich Tutor requiring unit steps. German students completed fewer problems on average than American students ($M = 0.82$ compared to $M = 1.46$), especially in ORCCA ($M = 0.60$). Accordingly, German students also took around 2 min longer to finish problems, requiring especially more time in the Stoich tutor (*Median* = 10.8 min) than in ORCCA (*Median* = 3.5 min). However, given the low completion rate of German students solving problems in ORCCA, these differences might also be partially explained by a selection effect.

RQ1: Differences in Freeform Strategies on Paper in German and American Students

To answer RQ1, we identified three distinct strategies that US and German students in our sample employed while solving problems on paper and compared how often they occurred in each sample. The first strategy involves students attempting to construct a multiplication chain of terms when given a value and a target value, similar to the factor label method (Flowers & Theopold, 2019). This strategy was the majority strategy of American students (80%) but also expressed by two German students (18%). The second strategy involves transforming a set of formulas into the target value after establishing an initial equation, without explicitly stating a general relationship. This strategy was employed by 10% of American and 18% of German students. The third strategy entails students utilizing a given value and a general relationship, if necessary, rearranging the equation according to the target variable to determine the target value by plugging in specific values. This strategy was the majority strategy of German students (64%) and exclusively expressed by German students. Table 2 illustrates examples of artifacts corresponding to each strategy.

Strategy 1 is characterized by a structured problem solving approach in which students chain fractions through multiplication to reason from a given value to a target value. The prototypical example highlights two additional characteristics: students usually explicitly specify units of individual quantities and then use these units to cancel them out until they arrive at the target quantity unit. Canceling units is performed by striking through unit labels across individual fractions that are chained together. The final result is then derived by multiplying the resulting fraction chain. Labels (i.e., units and substance descriptors) serve as the foundation to check if that chain is appropriate to solve the problem at hand (Fig. 1).

Strategy 2 involves a flexible approach of solving stoichiometry problems in a series of equation transformations (Fig. 2). The prototypical example highlights two additional characteristics: The flexibility of not labeling units and carrying out implicit calculations during steps. What is typical about Strategy 2 is that students would often not explicitly label individual quantities in transformations with units. Furthermore, individual calculations would not be fully specified. For example, a unit conversion would be implicitly carried out by transforming a value in the notation of $A \rightarrow B$ with B representing the transformed A.

Strategy 3 is characterized by transforming general relationships of abstract quantities via known or given equivalences (Fig. 4). Students start out by writing out a general relationship relevant to the problem, often accompanied by the descriptor “es gilt” (German for: “it applies that”). They then transform that equivalence to derive a formula that isolates and results in the target quantity and then filling in the relevant numbers to compute the final result. Students would usually explicitly label individual quantities in transformations when filling in numbers, as these units are implied or represented by the abstract formulas they solve the problem with. Furthermore, students would tend to stack individual transformations of formulas vertically, transforming them via computations to both sides, which is akin to Strategy 2 (formula transformation).

Individual responses of student problem solving on paper highlighted boundary cases containing elements of two strategies but were coded based on what strategy they resemble more. A case coded as 3 (general relationship) is close to being coded as 2 in the sense that it primarily contains transformations of equations (typical for both strategies) but only minimal amounts of abstract representation of quantities that are then filled in (typical for Strategy 3) with given values and remain untransformed (atypical for Strategy 3), as shown in Fig. 5.

One boundary case between Strategy 1 (given-target) and 2 (formula transformation) contained elements typical for Strategy 1: spelling out units, canceling units, and representing computations as fraction multiplication. However, the student chose to carry out formula-like transformations of intermediary results, which is typical for Strategy 2. As the characteristics of Strategy 1 were more numerous, Strategy 1 was assigned (Fig. 6).

RQ2: Practice in Two Tutoring Systems Based on Freeform Strategy

The two ITSs provide varying degrees of support for the strategies identified in RQ1. For each strategy, we define whether each of the tutoring systems constitutes a “strategy match” (Table 3) to investigate the relationship between students’ expressed strategies on paper and ITS-based problem solving (RQ2). Specifically, we compare the use of each ITS of students with and without a strategy match. We define strategy match as the ability of the tutoring system to *allow*

Table 2 Example responses of the three problem solving strategies based on the same stoichiometry problem. We note that one student did not get to the paper exercises in the session time and was not assigned a strategy

1 Given-Target (N = 10 students, 8 American and 2 German)	Can we calculate the number of moles of salt (NaCl) that are in a 350 ml solution that is 1.12 molar (mol/L)? Our result should have 3 significant figures. $\cancel{1.12\ \frac{mol}{L}}$ $350\ mL\ NaCl \left(\frac{1\ L}{1000\ mL}\right)\left(\frac{1.12\ mol}{1\ L}\right) = 0.392\ mol\ NaCl$
2 Formula Transformation (N = 3 students, 1 American and 2 German)	Can we calculate the number of moles of salt (NaCl) that are in a 350 ml solution that is 1.12 molar (mol/L)? Our result should have 3 significant figures. 350 ml of salt $\xrightarrow{L} \frac{350}{1000} = 0.35\ L$ $1.12\ \frac{mol}{L}$ $\to 1.12 \times 0.35 = \boxed{0.392}$
3 General Relationship (N = 7 students, all German)	**Bitte löse das folgende stöchiometrische Problem und kreise deine endgültige Antwort ein, wenn du Zwischenschritte aufschreibst. Diese Seite kann für Notizen verwendet werden.** Können wir die Anzahl an Mol des Salzes NaCl in einer 350 ml Lösung berechnen, die 1,12 mol/l hat? Unser Ergebnis sollte auf drei Stellen genau sein. $c = 1{,}12\ mol/l$ $V = 0{,}35\ l$ $\cancel{n = \frac{c}{V} = \frac{1{,}12}{0{,}35}}$ $c = \frac{n}{V}$ $n = c \cdot V$ $= 1{,}12 \cdot 0{,}35$ $= 0{,}392\ mol$

(meaning that expressing relevant input is possible) and *prefer* (meaning that the tutoring system offers support to execute the strategy or recommends it) a strategy (Aleven et al., 2009). Strategy 1 requires an interface that allows for unit cancellation based on a set of fractions chained via multiplication, which is explicitly supported (and required) by Stoichiometry Tutor. Further, ORCCA currently does not provide hints to initially reflect on the target variables, adding additional rationale to assign a mismatch between ORCCA and Strategy 1. Strategy 2 requires a formula transformation interface, where the specification of units and intermediary computations are optional, a distinct characteristic of ORCCA. Such flexibility in problem solving step transformation and compounding steps in one operation is also typical for Strategy 3, making it supported by ORCCA. Further, flexible specification of units in strategy is another argument to consider ORCCA a match to Strategy 3 (Fig. 7).

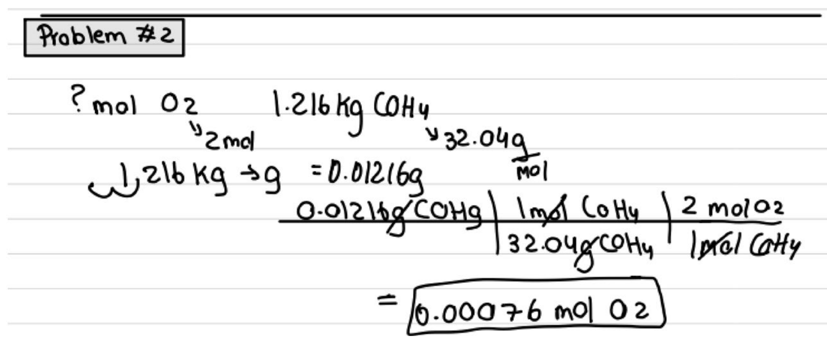

Fig. 4 Prototypical example of Strategy 1 (given-target)

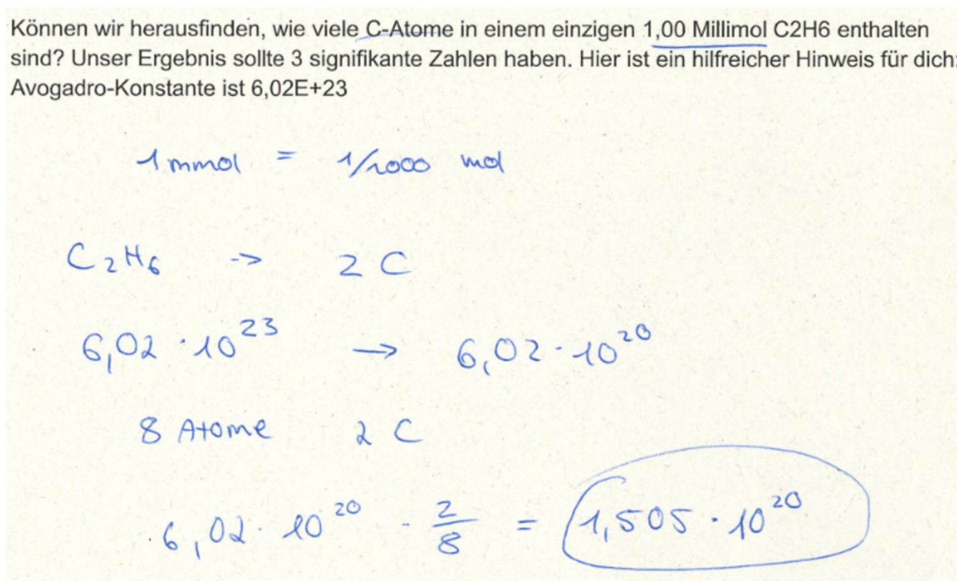

Fig. 5 Prototypical example of Strategy 2 (formula transformation). Note: In this case, the student came up with an incorrect solution by multiplying by 2/8 instead of two

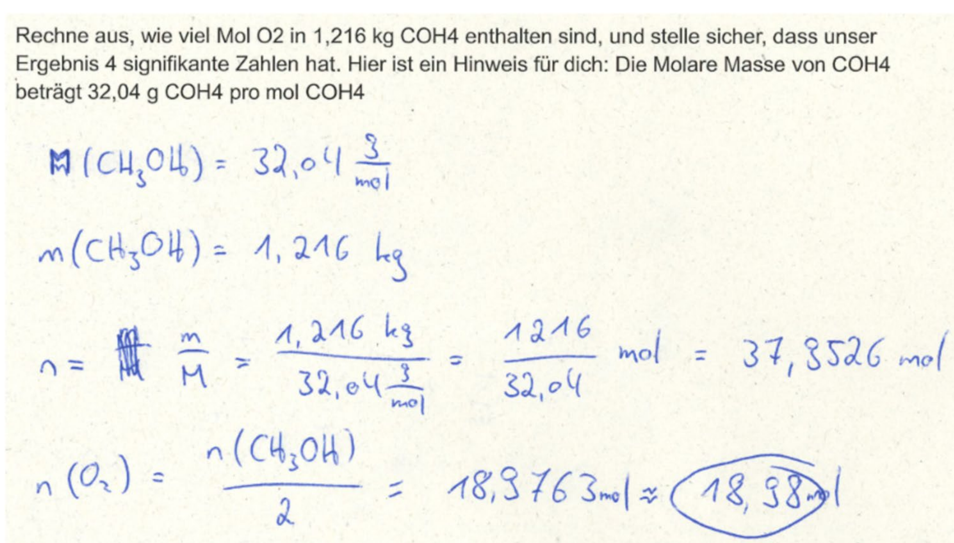

Fig. 6 Prototypical example of Strategy 3 (general relationship)

Table 3 Overview on how each tutoring system supports the three derived strategies

Strategy	Stoichiometry Tutor	ORCCA
1: Given + target	Match	No match
2: Transformation	No match	Match
3: Substitution	No match	Match

Stoich Tutor was more accommodating of the majority strategy of American students (Strategy 1; assigned to 8 out of 9 students) and ORCCA more of the majority strategy of German students (Strategy 2 and 3; assigned to 9 out of 11 students). Hence, we summarize interview themes related to both national samples next (Fig. 8).

Students' Initial Approaches to Tutoring Systems Qualitative investigation of students' initial approaches to problems in the Stoich Tutor and ORCCA based on log data and video recordings exhibited distinct characteristics based on whether a given student's strategy constituted a match to the ITS (Table 4). In the following, we discuss each case in depth.

Students aligning with the Stoich Tutor (a) initially reflected about the target value, (b) employed a unit-based reasoning approach, (c) and processed instruction to identify given and target. Regarding (a), students with a converging strategy would typically begin with expressing a target value as a problem solving goal, such as user1_US, whose first utterance was: "I am going to work on my first unit conversion because we want to go to mols of COH_4." Similarly, thinking about units (b), user5_Ger reflected on their problem solving goal by stating: "That means I have to convert how many moles I end up with," indicating that the student thinks about the problem in terms of unit conversion. As evidence for (c), user4_US, like other students, would often explicitly state givens and target values during their initial scanning of the problem: "I see that the gas station needs 25 million moles of H_2. And I know that I should figure out how

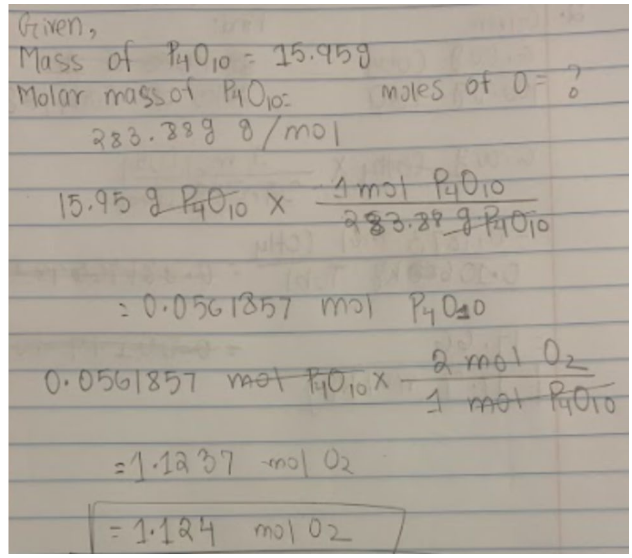

Fig. 8 Boundary case between Strategy 1 (given-target) and 2 (formula transformation) that was eventually coded as 1

many kiloliters of glucose solution are necessary to produce this H_2. So I'm trying to figure out how to relate the moles to the kiloliters […]."

Conversely, students with a strategy conflicting with Stoich Tutor tended to (a) be confused related to the fine-grained problem solving interface, (b) skip steps, specifically setting the target substance, and (c) verbalize fractions rather than inputting them into the interface and canceling them. Specifically, for (a), we found that students would often request and use a hint to get started on the Stoich Tutor: "Give the number of moles of H_2 per mole of $C_6H_{12}O_6$. Oh, that's a good idea that the hint suggests" (user10_Ger). Some students, like user3_Ger, would even explicitly acknowledge that they need hints to complete the problem in the interface: "I will just do everything with hints here." If these students started to work, they would not think about the target substance but rather start out with givens and their units, such as user3_Ger: "So I'm searching for the unit kL of the glucose solution, i.e., $C_6H_{12}O_6$." Regarding (c), we found that

Fig. 7 Boundary case between Strategy 3 (general relationship) and 2 (formula transformation) that was eventually coded as 3

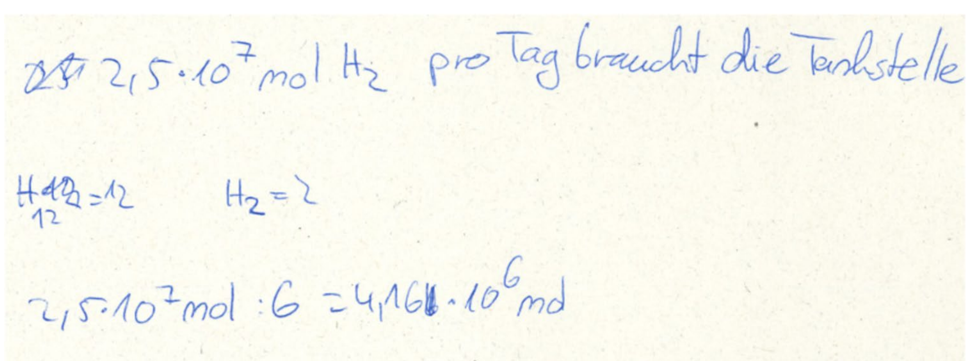

Table 4 Characteristics of students' approaches to problems in the Stoich Tutor and ORCCA

	Supported	Not supported
Stoich	(a) Thinking about the target variable first, then the given value, and reasoning their way toward the target (b) Units-first reasoning and getting the target unit right (c) Initial processing of instruction and problem statement filter out targets and givens	(a) Confusion related to fine-grained problem solving interface (b) Skipping steps, specifically setting the target substance (c) Verbalizing fractions rather than entering them into the interface and canceling them
ORCCA	(a) Compounding problem solving steps (b) Students start setting up a formula first, then thinking about the next value to calculate (c) Neglecting units and substances	(a) Trying to reason backward from the target variable via note-taking, internal monologue, or "all-in-one" equations (b) Recording the most important values and quantities in the notes and only then plan calculations, sometimes saving intermediary results

students would often say what the next fraction is, sometimes converting them in their head, but not initially input the corresponding fraction into the Stoich Tutor interface, with a typical verbalization reading: "Okay, I want to enter here that I have one liter of the solution, um, and this liter contains 1.12 mol."

Students with a strategy aligning with ORCCA (a) compounded problem solving steps, (b) started setting up a formula first and then thinking about the next value to calculate, and (c) tended to neglect units and substances. When compounding steps (a), students would often verbalize transformations without writing them into the interface, for example, user8_Ger: "The Avogadro constant is 6.02, which means that 1 mol of the substance would have 6.02 times 10 to the power of 23 carbon atoms. And now we can divide that by 1000, because we want to find millimoles. And then take another 2, because C_2H_6 has 2 carbon atoms." While this student demonstrated a high fluency in the task, other students (e.g., user2_US and user4_US) would show similar compounding by reasoning on ORCCA's notepad. However, students also exhibited similar reasoning when setting up formulae in ORCCA. For example, user10_Ger compounded a stoichiometric composition with a molecular weight transformation into the formula interface while self-explaining: "That means I first calculate how many oxygen molecules—um, uh, how many molecules they have." In these cases, students created formulae step by step without explicitly setting up given and target values, substances and units. As evidence for (c), students in these scenarios would neither mention nor specify units in the ORCCA formula interface while still getting the problem correct with few errors along the way.

Finally, students working with a diverging strategy in ORCCA tended to (a) try to reason backward from the target variable via note-taking, internal monologue, or "all-in-one" equations and (b) record the most important values and quantities in the notes and only then plan calculations, sometimes saving intermediary results. (a) For example, similar to students aligning with the Stoich Tutor, user8_US would process the problem instructions focusing on the target variables and givens: "Grams of hematite, gotta end with grams of iron […] got 55.847 g/mol of iron." Notably, this student then checked whether their hypothesis about needing to do a unit conversion was correct by checking a tutor hint. Other times, students would write their hypotheses onto ORCCA's notepad to then enter a single, long equation (or "all-in-one question") to solve the equation, user4_US: "Let's make some notes that we don't forget anything like the 6 before." After four more such utterances, user4_US acknowledged: "So this time we can write a formula." Similarly, these students would resort to using ORCCA's notepad to write out fractions similar to the Stoich Tutor, rather than working with the formula interface, to save intermediary calculations and results. For example, user6_Ger, using the notepad, said: "[…] um accordingly I would first calculate how many moles of COH_4 I have. And I have 1, 2 uh- 16 kg, that's 1216 g. And I know that 32.04 g uh corresponds to one mole." These steps align with reasoning from the givens to the target and search for the individual problem solving steps in consideration of the units and substances.

RQ3: German and American Students' Perceptions of Two American Tutoring Systems Based on Short Interviews

To answer RQ3, we contrast the perceptions of German and American students on Stoich Tutor and ORCCA, two ITSs developed in the USA. Two themes emerged from the analysis of short interviews: First, German students tended to prefer ORCCA over the Stoich Tutor as they perceived the Stoich Tutor to impose unnecessary steps. Second, all students perceived ITS feedback and as-needed instruction through hints as helpful, with German students acknowledging that they were required to successfully work with the Stoich Tutor, meaning that they learned how to use the interface through its scaffolding. The key reason for German students preferring ORCCA over the Stoich Tutor was because they perceived ORCCA as more aligned with their problem-solving strategy. Most users in the German sample (3, 5, 6, 7, 9, and 11) stated that the problem solving process in the Stoich Tutor is too predetermined and is therefore perceived as a constraint on their own thought process, as it requires steps that users would usually not take. For example, user9_DE mentioned: "I had the feeling that I had to follow the solutions that the tutor [Stoich Tutor] had considered beforehand and that I had to do a lot more tasks than were necessary to solve the problem normally." user5_DE mentioned: "Uhm, so the tutor [ORCCA] would give you less and you could work more, so I kind of worked the way I would always do and didn't necessarily write down reasons as required in the first tutor or didn't write everything down in detail." In comparison, there was no clear preference for any of the two ITSs in the US sample. Still, some American students considered the Stoich Tutor as less suitable for more experienced learners due to the heavily predefined problem solving process (user9_US): "Yeah. It's good if it's the first time learning it. But because I have some experience with stoichiometry, and I learned in a certain way this is making it a bit more difficult, especially because it's on computer."

Distinct to German users, two users (10 and 11) stated that they would not have been able to solve problems with Stoich Tutor without hints due to its instructional strategy. For example, user10_DE mentioned: "And here with the second one [Stoich Tutor], it was definitely a necessity that there were these instructions, because I don't know, when I saw the instruction video, I actually thought: that looks really cool." Similarly, user9_DE stated: "Uhm, I think it's good that you are pointed to canceling units."

The perceptions of German and American students related to both ITS systems also qualitatively emerged during their initial problem-solving steps when starting to work with each system while thinking aloud. More in-depth reporting on these differences is in Appendix 1.

Discussion

The present study investigated university students' problem-solving strategies in stoichiometry across two US-American ITS with different levels of scaffolding. Twenty-one university students across the USA and Germany worked with at least one of each system, with their freeform problem-solving strategies recorded based on responses to analogous problems on paper.

The first contribution of this study is to introduce a taxonomy of three distinct problem-solving strategies in stoichiometry used by American and German students, with two of these strategies constituting the majority strategy of a US-American and German university student population (RQ1). To the best of our knowledge, the strategy specific to the German population has not been explicitly described in chemistry education and involves equation transformations with abstract symbols to isolate a target variable and then finally pluck in given values and compute the solution. Similar approaches to problem solving have been observed in German physics instruction (Dröse & Prediger, 2018) and are good practice in German education, including expected formalization (Kurzweil, 2023). This strategy was observed only when students completed problem-solving steps on paper without any instructional support. Further, this strategy is not accommodated by any of the two American ITS we investigated, suggesting that it might be useful to revise it for use in Germany so it accommodates such abstract symbolization. The strategy, involving transformations of equations with symbols representing variables to pluck in values, could be supported through ORCCA's formula interface. In American students, we observed that most students used a strategy that aligns with the factor label method typical for US education documented in prior work (Flowers & Theopold, 2019), which involves chaining fractions for multiplication. Still, we observed a third strategy in both student populations, which is characterized by symbolizing implicit transformations, which involves shorthand transformations based on equivalences and relationships (e.g., 2:1). Notably, this strategy has not been empirically documented in student responses or textbooks, but most closely aligns the proposal of a "logical method" in Schmidt (1997) that the author deemed most desirable for stoichiometry problems due to its straightforward mathematical problem representations.

Our second contribution relates to differences in problem-solving strategies in ITS depending on strategies expressed on paper (RQ2). Our study makes an important contribution to the field by being one of the first to investigate how scaffolding in tutoring systems developed in North America are used by students inside and outside of the US who tend to use and are taught different problem-solving strategies. Our findings reveal that students approach ITS, such as the Stoich Tutor and ORCCA, with problem-solving strategies depending on their problem-solving strategy on paper, which likely represents the main method they were taught. For example, students unfamiliar with the factor-label method (who were primarily German) often struggled with the Stoich Tutor interface, frequently skipping key steps, such as identifying the target substance. Additionally, these students tended to verbalize fractions instead of entering them directly into the system and canceling them. In contrast, students using the factor-label method (who were primarily American), tended to adopt unit-based reasoning and set clear goals in Stoich Tutor. While prior research suggests that scaffolded interfaces help students learn new strategies over time (Clark & Mahboobin, 2017; Yuriev et al., 2017), it remains an open question whether students' learning would benefit from highly scaffolded environments using unfamiliar strategies. Our findings underscore the importance of ensuring that tutoring systems are aligned with the rest of instruction and a student's curriculum. As that is not always possible in practice, one possible solution to ease students' initial steps with an unfamiliar tutoring system is to implement familiarization strategies in new learning environments to ease the transition for students using a different strategy, such as encouraging peer discussions or providing additional resources (Shadiev et al., 2022). Alternatively, a more flexible ITS that supports a wider range of problem-solving strategies could accommodate student instructional needs. In the case of ORCCA, students employing a formula transformation strategy were able to transform equations and progress through steps as expected. However, these students often ignored substances in their calculations—a practice allowed in ORCCA, which they also exhibited during paper-based tasks. The lack of unit feedback in ORCCA, however, may hinder students from identifying errors flagged by the system, limiting their learning opportunities (Dahsah & Coll, 2007; Van der Westhuizen, 2015).

Our third contribution relates to the finding that German students generally perceived an open-ended ITS with little, dynamic scaffolds as more aligned with their problem-solving strategy than the highly scaffolded Stoich Tutor aligning with a foreign instructional strategy (RQ2). Key differences include (a) the requirement of the Stoich Tutor to complete steps some students deemed unnatural for their problem-solving strategy and (b) German students being more used to a formula transformation strategy rather than chaining and canceling multiplications, as prompted by the Stoich Tutor. The ORCCA tutor is designed to support a broader range of problem types and to support more open ended problem solving. As a result, it lacks the more detailed support that the Stoich tutor provides for the factor label method commonly used in US education (Flowers & Theopold, 2019; Schmidt, 1997). However, the finding that ORCCA supported a strategy commonly found in German students speaks to a strength of an open-ended ITS: through its flexibility in matching problem solving states to viable next problem solving steps and keeping the requirements to express steps (in terms of units and other labels) minimal, it can accommodate a wide variety of problem-solving strategies and sequences. Still, this flexibility comes at a cost: by treating difficult problem solving steps for novices as optional, it is limited in providing additional scaffolding and support for such steps (e.g., by explicitly requiring units, a step stoichiometry novices often struggle with; Dahsah & Coll, 2007; Van der Westhuizen, 2015).

Fourth, although initially struggling with Stoich Tutor's interface, German students were more successful in completing problems in the Stoich Tutor compared to ORCCA, although the Stoich Tutor misaligned with their instructional strategy. Why was that the case? First, German students expressed usability issues in ORCCA that could be improved (e.g., accommodating the German way of expressing decimals with commas instead of dots). However, the same issue also exists in the Stoich Tutor. Yet, the Stoich Tutor's instruction through hints supports navigating and working through the interface (e.g., by scaffolding and hinting at specific unit steps). In contrast, ORCCA's hints remain abstract and prompt for specific transformations to a formula. The former might have been more useful in helping students complete problems in an ITS. Indeed, two German students noted that they would not have been able to solve problems with Stoich Tutor without hints due to its instructional strategy. In other words, the highly scaffolded Stoich Tutor interface made it easier for German students to engage in deliberate practice. Hence, ITS scaffolding for foreign strategies might still be useful for student learning, as ITS provides learning opportunities through feedback on each step (VanLehn, 2006). A hypothesis for future work is that German students can adapt to the predominantly American factor labeling method present in the Stoich Tutor and benefit from it, including from learning an additional strategy to adaptively select when problem solving (Siegler, 1991; Star & Rittle-Johnson, 2008).

Limitations and Future Work

First, we acknowledge that the study design cannot speak to how German and American students might adapt to both Stoich Tutor and ORCCA over time, as is common in real-world applications, such as rollout in remedial or entry-level college courses. It is an open question how students benefit from and adapt to both ITSs over time, including fading both ITSs in sequence, which might have even greater benefits (McNeill et al., 2006). Similarly, our present study design and sample does not allow for a systematic study of learning gain differences between Stoich and ORCCA across German and American students, including across different levels of prior knowledge. One potential direction to explore in future work is whether the highly scaffolded Stoich Tutor environment is conducive to learning, given that its rigorous application of the factor-label method might elicit algorithmic problem-solving in students. Such strategies, involving remembering and involving specific problem-solving steps without conceptual reasoning, have been documented in prior literature (Schmidt & Jignéus, 2003). While they might help students solve problems successfully (as observed in the present study), they may not be enough to elicit learning of difficult stoichiometry concepts, such as units (Hafsah et al., 2014).

Second, German and American samples might be systematically different because the American sample included more students that did not study chemistry as their primary major but rather took relevant chemistry courses (a little over half), which was not the case for the German sample. More research is desirable to pinpoint specific difficulties of different populations working with the two ITSs.

Finally, we acknowledge that the present study's sample might have been small to make robust empirical inferences about quantitative differences in problem solving between German and American students (e.g., likelihood of solving problems successfully in ORCCA). This limitation also applies to detecting potential relationships between students' use of a specific ITS and their strategy during paper-based problem solving. Our data did not show evidence of such associations, although statistical power to detect such associations reliably is low, given our sample size. Yet this lack of association is not surprising, given the short duration of the use of the ITS (shorter than students would typically use it in an actual educational setting). In more serious instructional contexts, students may adopt strategies learned through ITS during free-form problem-solving on paper.

Summary and Conclusions

This study contributes a novel taxonomy of stoichiometry problem-solving strategies in German and American students, documenting a strategy distinct to German students involving transforming abstract relationships to then pluck in values, which has previously been noted in physics. Further, the taxonomy includes a strategy with shorthands based on mathematical relationships and conceptual equivalences that prior work has not documented in textbooks or student responses. While German students perceived an ITS with dynamic scaffolding as more desirable and aligned with their problem solving, they still were more likely to successfully solve problems in an ITS scaffolded for a foreign strategy. Still, students who were not using the factor-label method during paper-based problem solving (who were primarily German) initially struggled with correctly setting goals and using fractions in the Stoich Tutor. Therefore, revisions to ITS based on local instructional cultures could make them potentially more effective and aligned with curricular goals. For German students, providing an interface for transforming equations with abstract symbolization in stoichiometry is the foremost improvement based on this study's findings.

Appendix 1 Freeform Problem-Solving Strategy Coding Rules

The coding of freeform problem-solving strategies followed a two-stage process where two coders initially discussed and noted rules for coding and then coded all remaining solutions to establish inter-rater reliability. The initial establishment of coding rules and the resulting categories emerged through a joint discussion of 4–6 example cases. The coders reviewed this randomly sampled subset of artifacts and noted down the following rules for coding, which were following an order or priority where solutions were first coded by notable formalisms and then based on the way in which operations and transformations are represented if they were present. The assigned Strategies 1 through 3 are represented through (1), (2), and (3).

- Step 1: Can the paper response be coded through a notable formalism?

  ○ "Es gilt" ("it applies that") (3)

  This formalism is used to represent relationships, for example, of a chemical formula. Chemical reagents or quantities are notated with letters.

  ○ How are unit conversions represented?

  Case #1: Canceling (1)

  - Unit conversions are represented in fractions and usually units are crossed out to complete a conversion.

  Case #2: Informal notation on the side of an Eq. (2)

  - For example "*2" or "multiply by 2" is written next to the term to indicate a multiplication by 2. A second, transformed equation is then juxtaposed, usually below the original one.

  ○ Definition of variables through letters (3)

  Formulae or givens are expressed in letters, then transformed, and eventually substituted with numbers.

  ○ Kurzschreibweise/implizite Umformung (2) (shorthands)

  Arrows, dashes or other stylistic symbols are used, which represent a transformation, usually without specifying all transformations explicitly.

  - Example #1: Multiplication by two is expressed as "→*2"
  - Example #2: A unit conversion is expressed as "→g/mg"

- Step 2: If the paper response cannot be coded through a notable formalism, then code it according to the following ways in which chains of operations and transformations are represented.

  ○ Case #1: Left to right/tight layout (1)
  ○ Case #2: Individual equations and terms vertically aligned (2)
  ○ Case #3: Both Case #1 and Case #2 are mixed or there are multiple chains of thought (3)

Author Contribution Conrad Borchers: conceptualization, methodology, investigation, formal analysis, writing—original draft, writing—review and editing, visualization, project administration, funding acquisition, and data curation. Hendrik Fleischer: conceptualization, investigation, formal analysis, writing—original draft, writing—review and editing, and data curation. David J. Yaron: software and writing—review and editing. Bruce M. McLaren: software and writing—review and editing. Katharina Scheiter: conceptualization, writing—review and editing, and supervision. Vincent Aleven: conceptualization, writing—review and editing, and supervision. Sascha Schanze: conceptualization, writing—review and editing, funding acquisition, and supervision.

Funding Open Access funding provided by Carnegie Mellon University.

Declarations

The authors have no known conflict of interest to disclose. The authors have no relevant financial or non-financial interests to disclose. Research activities conducted in the USA were approved by Carnegie Mellon University's Institutional Review Board and included informed consent. Research activities conducted in Germany were regulated under a data protection declaration, serving as informed consent. Carnegie Mellon University's Graduate Student Assembly/Provost GuSH Grant funding was used to support this research under the project name "Scaffolding effects on problem-solving in intelligent tutoring systems." Research materials are available upon request by accessing https://stoichtutor.cs.cmu.edu/ and https://orcca.ctat.cs.cmu.edu/. All data analysis code is publicly available at: https://github.com/conradborchers/think-aloud-chem. The log data stemming from student-tutor interactions during research activities in the USA are available at https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=5371. Three conference papers using the US sample have been published (see citations below) with no ideas, analyses, or other manuscript content re-used in the present manuscript:

Borchers, C., Zhang, J., Baker, R. S., & Aleven, V. (2024). Using Think-Aloud Data to Understand Relations between Self-Regulation Cycle Characteristics and Student Performance in Intelligent Tutoring Systems. In Proceedings of the 14th Learning Analytics and Knowledge Conference (pp. 529–539). https://doi.org/10.1145/3636555.3636911

Zhang, J., Borchers, C., Aleven, V., & Baker, R. S. (2024). Using Large Language Models to Detect Self-Regulated Learning in Think-Aloud Protocols. Proceedings of the 17th International Conference on Educational Data Mining (EDM). https://doi.org/10.5281/zenodo.12729790

Zhang, J., Borchers, C., & Barany, A. (2024). Studying the Interplay of Self-regulated Learning Cycles and Scaffolding Through Ordered Network Analysis Across Three Tutoring Systems. In: Kim, Y.J., Swiecki, Z. (eds) Advances in Quantitative Ethnography. ICQE 2024. Communications in Computer and Information Science, vol 2278. Springer, Cham. https://doi.org/10.1007/978-3-031-76335-9_17

References

Aleven, V., Mclaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 522–560). Routledge.

Astuti, N. H., Rusilowati, A., & Subali, B. (2021). STEM-based learning analysis to improve students' problem solving abilities in science subject: A literature review. *Journal of Innovative Science Education, 10*(1), 79–86.

Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C., & Fike, A. (2009, October). MetaTutor: A MetaCognitive tool for enhancing self-regulated learning. In 2009 AAAI Fall symposium series.

Baker, R.S., Corbett, A.T., & Koedinger, K.R. (2002) The resilience of overgeneralization of knowledge about data representations. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, USA.

Balta, N., Mason, A. J., & Singh, C. (2016). Surveying Turkish high school and university students' attitudes and approaches to physics problem solving. *Physical Review Physics Education Research, 12*(1), 010129.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344.

Borchers, C., Carvalho, P. F., Xia, M., Liu, P., Koedinger, K. R., & Aleven, V. (2023, August). What makes problem-solving practice effective? Comparing paper and AI tutoring. In European conference on technology enhanced learning (pp. 44–59). Cham: Springer Nature Switzerland.

BouJaoude, S., & Barakat, H. (2003). Students' problem-solving strategies in stoichiometry and their relationships to conceptual understanding and learning approaches. The Electronic Journal for Research in Science & Mathematics Education.

Chi, M., & VanLehn, K. (2010). Meta-cognitive strategy instruction in intelligent tutoring systems: How, when, and why. *Journal of Educational Technology & Society, 13*(1), 25–39.

Clark, R. M., & Mahboobin, A. (2017). Scaffolding to support problem-solving performance in a bioengineering lab—a case study. *IEEE Transactions on Education, 61*(2), 109–118.

Dahsah, C., & Coll, R. K. (2007). Thai Grade 10 and 11 students' conceptual understanding and ability to solve stoichiometry problems. *Research in Science & Technological Education, 25*(2), 227–241.

DeepL (n. d.). DeepL Pro API. https://www.deepl.com/en/pro-api. Accessed 15 Jan 2024.

Dröse, J., & Prediger, S. (2018). Strategien für Textaufgaben. Fördern mit Info-Netzen und Formulierungsvariationen. *Mathematik Lehren, 206*, 8–12.

Flowers, P. & Theopold, K. (2019). [ETextbook] Chemistry-2e. https://openstax.org/details/books/chemistry-2e

Hafsah, T., Rosnani, H., Zurida, I., Kamaruzaman, J., & Yin, K. Y. (2014). The influence of students' concept of mole, problem representation ability and mathematical ability on stoichiometry problem solving. *Scottish Journal of Arts, Social Sciences and Scientific Studies, 21*(1), 3–21.

King, E. C., Benson, M., Raysor, S., Holme, T. A., Sewall, J., Koedinger, K. R., ... & Yaron, D. J. (2022). The open-response chemistry cognitive assistance tutor system: Development and implementation.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.

Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of Educational Data Mining, 43*, 43–56.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78.

Kurzweil, P. (2023). Reaktionen, Katalyse, Thermochemie. In: Chemie. Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-41568-6_6

Leung, F. K., Clarke, D., Holton, D., & Park, K. (2014). How is algebra taught around the world?. Algebra teaching around the world, 1–15.

Mandina, S., & Ochonogor, C. (2018). Comparative effect of two problem solving instructional strategies on students' achievement in stoichiometry. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(12), em1621.

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011a). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human Computer Studies, 69*(1–2), 70–79. https://doi.org/10.1016/j.ijhcs.2010.09.001

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011b). Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers & Education, 56*(3), 574–584. https://doi.org/10.1016/j.compedu.2010.09.019

McLaren, B. M., van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in classroom experiments. *Computers in Human Behavior, 55*, 87–99.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*(2), 153–191.

Melis, E., & Siekmann, J. (2004, June). Activemath: An intelligent tutoring system for mathematics. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 91–101). Berlin, Heidelberg: Springer Berlin Heidelberg.

Molnár, J., & Molnár-Hamvas, L. (2011). LEGO-method–New strategy for chemistry calculation. *US-China Educ. Rev. B, 7*, 891–908.

Neuendorf, K. A. (2018). 18 Content analysis and thematic analysis. Advanced research methods for applied psychology: Design, analysis and reporting, 211.

Niaz, M., & Montes, L. A. (2012). Understanding stoichiometry: Towards a history and philosophy of chemistry. *Educación Química, 23*, 290–297.

Pardos, Z. A., Tang, M., Anastasopoulos, I., Sheel, S. K., & Zhang, E. (2023, April). Oatutor: An open-source adaptive tutoring system and curated content library for learning sciences research. In Proceedings of the 2023 chi conference on human factors in computing systems (pp. 1–17).

Posamentier, A. S., & Krulik, S. (2008). Problem-solving strategies for efficient and elegant solutions, grades 6–12: A resource for the mathematics teacher. Corwin press.

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., & ZuBelzen, A. U. (2020). A framework to foster problem solving in STEM and computing education. *Research in Science & Technological Education, 38*(1), 105–130.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. *In International Conference on Machine Learning* (pp. 28492–28518). PMLR.

Schmidt, H. J. (1997). An alternate path to stoichiometric problem solving. *Research in Science Education, 27*, 237–249.

Schmidt, H. J., & Jignéus, C. (2003). Students strategies in solving algorithmic stoichiometry problems. *Chemistry Education Research and Practice, 4*(3), 305–317.

Shadiev, R., Yang, M. K., Atamuratov, D., Shadiev, N., Fayziev, M., Gaevskaia, E., ... & Oshanova, N. (2022, August). Familiarization strategies to facilitate mobile-assisted language learning in unfamiliar learning environments: A study of strategies development and their validation. In International conference on innovative technologies and learning (pp. 213–217). Cham: Springer International Publishing.

Selting, M., Auer, P., & Barth-Weingarten, D. (2011). A system for transcribing talk-in-interaction: GAT 2. Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion, 12, 1–51.

Shakya, A., Rus, V., Fancsali, S., Ritter, S., & Venugopal, D. (2022, July). NeTra: A neuro-symbolic system to discover strategies in math learning. In Proceedings of the third workshop of the learner data institute, The 15th International conference on educational data mining (EDM 2022).

Siegler, R. S. (1991). Strategy choice and strategy discovery. *Learning and Instruction, 1*(1), 89–102.

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, 18*(6), 565–579.

Tenison, C., & MacLellan, C. J. (2014, June). Modeling strategy use in an intelligent tutoring system: implications for strategic flexibility. In *International conference on intelligent tutoring systems* (pp. 466–475). Springer International Publishing.

Tóth, Z., & Sebestyén, A. (2009). Relationship between students' knowledge structure and problem-solving strategy in stoichiometric problems based on the chemical equation. *International Journal of Physics and Chemistry Education, 1*(1), 8–20.

Van der Westhuizen, L. (2015). The development of the conceptual understanding of first-year chemistry university students in stoichiometry using thinking skills, visualization and metacognitive strategies (Doctoral dissertation, North-West University).

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.

Yuriev, E., Naidu, S., Schembri, L. S., & Short, J. L. (2017). Scaffolding the development of problem-solving skills in chemistry: Guiding novice students out of dead ends and false starts. *Chemistry Education Research and Practice, 18*(3), 486–504.