Check for updates

Towards defining, assessing and modelling competency levels in stoichiometry†

Alexander Krause, [iD] *[a] Sebastian Polarz, [iD] [b] Anett Hoppe, [iD] [cd] Ralph Ewerth [iD] [cde] and Andreas Nehring [iD] [a]

Stoichiometry is a significant yet challenging topic in chemistry education. While extensive research has explored students' conceptions, difficulties, and learning approaches, this study adopts a competency-based approach to introduce a new model defining three competency levels in stoichiometry. The stoichiometry competency level model (StoiCoLe model) offers a framework for evaluating students' performance in algorithmic stoichiometry. To test the assumptions of the StoiCoLe model, a 40-item test was developed to measure and categorise student's competencies according to the model's levels. Using data from 289 students enrolled in an introductory chemistry course across three semesters, psychometric properties and model assumptions were analysed through Rasch-analysis and item processing times. The results indicate that there is sufficient psychometric reliability in the categorization of students according to the StoiCoLe model. However, both the item difficulty and the processing times are only partially consistent with the assumptions of the model and indicate an adaptation of the model. In line with prior studies, the majority of students exhibited lower competency levels. These findings are discussed in terms of how the competency-based approach can enhance relevant competencies and contribute to literature on chemistry education in stoichiometry.

Introduction

Stoichiometry is a fundamental area of chemistry, concerned with the quantitative ratios of substances in chemical reactions, as well as the composition of chemical compounds and mixtures (Brown *et al.*, 2021). It thus marks the significant transition of chemistry from a descriptive, qualitative science to a quantitative discipline. Competencies related to stoichiometry enable chemists to calculate masses, volumes or amounts of substances in chemical reactions, which is an essential requirement for conducting safe and effective laboratory work. To perform these calculations, chemists must possess a solid understanding of key stoichiometric concepts, such as the mole concept and stoichiometric laws, alongside practical knowledge of stoichiometric equations, quantities, units, and the language of chemical formulas. Consequently, teaching and learning in

stoichiometry involves both conceptual and mathematical tasks, which may vary in their emphasis on either aspect. Conceptual tasks require a deep understanding of relationships, concepts, and theories, while mathematical tasks can be solved with rule-based, step-by-step calculations – and are therefore often referred to as 'algorithmic' tasks (Bodner and Herron, 2002).

Stamovlasis *et al.* (2004, 2005) demonstrated that the competency to solve algorithmic or conceptual tasks are two independent dimensions: the mathematical steps required to solve algorithmic problems are not acquired solely by understanding the concepts, nor is a conceptual understanding acquired by solving algorithmic problems (Nurrenbern and Pickering, 1987; Sawrey, 1990; Mason *et al.*, 1997; Cracolice *et al.*, 2008).

Nevertheless, the use of mathematics is essential in achieving accurate stoichiometric outcomes when planning reactions or interpreting laboratory results. In the context of industrial processes, precise stoichiometric calculations can be of paramount importance. When large quantities of reactants are used and large amounts of energy are released, conceptual understanding alone is not sufficient, as deficits in algorithmic skills can lead to catastrophic consequences. Moreover, the competency to solve stoichiometric calculations constitutes the basis to solve algorithmic tasks in advanced areas such as acid–base reactions, kinetics, and chemical equilibrium. In particular, the transfer of algorithmic skills from familiar stoichiometric examples to novel quantitative challenges is critical in contemporary chemical research.

[a] *Institute of Science Education, Leibniz Universität Hannover, Hannover, Germany. E-mail: krause@idn.uni-hannover.de*

[b] *Institute of Inorganic Chemistry, Leibniz Universität Hannover, Hannover, Germany*

[c] *TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany*

[d] *Department of Mathematics and Computer Science, Philipps-Universität Marburg & Hessian Center for Artificial Intelligence (hessian.AI), Marburg, Germany*

[e] *L3S Research Center, Leibniz Universität Hannover, Hannover, Germany*

† Electronic supplementary information (ESI) available: Stoichiometry test instrument, results of the stoichiometry test, revised stoichiometry test instrument. See DOI: https://doi.org/10.1039/d5rp00077g

Considering the example of a complementary research method, the Thermogravimetric analysis (TGA) is used to monitor mass changes of a material sample under a controlled temperature programme and atmosphere (Saadatkhah *et al.*, 2020). The mass loss data provides insights into the composition and decomposition pathways of the sample, which may otherwise remain largely unknown. However, accurate interpretation of a TGA curve requires a high level of stoichiometric competency. Thus, both algorithmic and conceptual tasks are essential, and proficient learners must be capable of solving both types effectively (Bodner, 1987; Frank *et al.*, 1987; Smith *et al.*, 2010).

Stoichiometry has been a focal point in chemistry education research for over forty years. A significant emphasis of past research has been on identifying specific learning difficulties in stoichiometry and providing explanations for these challenges (Dierks, 1981; Schmidt, 1990; Huddle and Pillay, 1996; BouJaoude and Barakat, 2000; Arasasingham *et al.*, 2004; BouJaoude *et al.*, 2004; Agung and Schwartz, 2007; Dahsah and Coll, 2008; Gulacar and Fynewever, 2010; Scott, 2012; Gulacar *et al.*, 2013a, 2013b; Shadreck and Enunuwe, 2018). Building on this foundation, various methods and approaches have been developed to support learners in stoichiometry learning processes. These approaches range from visual representations, such as flow diagrams or matrixes (Tyndall, 1975; Kauffman, 1976; Berger, 1985; Cameron, 1985; Poole, 1986; Koch, 1995; Krieger, 1997; Olmsted, 1999; Ault, 2001; Wagner, 2001), visualisations and experientiality through models (Kashmar, 1997; Witzel, 2002; Molnar and Molnar-Hamvas, 2011; Ramesh *et al.*, 2020) and analogies from everyday life (Umland, 1984; Cain, 1986; Haim, 2003) to experiments (Arlotto, 1974; Figueira *et al.*, 1988; Martínez and Ibanez, 2020), mathematical methods (Garst, 1974; Tykodi, 1987; Mousavi, 2019; Carlson, 2022) problem-solving strategies (Schmidt, 1997; Schmidt and Jignéus, 2003; Hand *et al.*, 2007; Okanlawon, 2008; Rosenberg *et al.*, 2016; Gulacar *et al.*, 2022) and new, especially digital teaching formats (Cotes and Cotuá, 2014; Gayeta, 2017; Nufus *et al.*, 2020; Rasmawan, 2022).

In this paper, we contribute to the existing literature by applying a competency-based approach to the field of stoichiometry. We propose and test a model that delineates various competency levels. Based on the framework of algorithmic questions from Smith *et al.* (2010), the framework of sub-problems from Gulacar and Fynewever (2010) and the framework of the multidimensional analysis system from Dori and Hameiri (2003), our model is based on content knowledge on central stoichiometric concepts of algorithmic stoichiometry. It outlines the competency levels of learners and potential trajectories for their learning progression, particularly concerning complex, multi-step algorithmic tasks. This model offers a comprehensive perspective on the abilities students should develop and demonstrate as they advance in their competency in algorithmic stoichiometry. However, this means that the model does not allow any conclusions to be drawn about students' conceptual understanding of stoichiometry, as this would require items to capture concepts, such as free text explanations, multiple choice questions or drawings. In our approach, we measure the competency in algorithmic stoichiometry that relate to the outcome and the process-related application of the sub-competencies. The model therefore represents a significant synthesis of existing approaches and an essential advancement. Specifically, it refines certain sub-problems identified by Gulacar and Fynewever (2010) and structured them based on the framework of the multidimensional analysis system by Dori and Hameiri (2003). Our objective is to facilitate the development of effective instructional strategies and learning methods, as indicated above, through the precise delineation of (sub-)competencies. This endeavor also aims to address the research gap identified by Blömeke *et al.* (2015) concerning the definition and assessment of competencies in higher education for the particular topic of chemistry.

By modelling competency levels, our aim is to provide a framework for summarising and assessing the learning level and progress. This model facilitates the evaluation of competencies both at individual student level and across entire groups. Given the robust psychometric test results obtained in this study, the StoiCoLe model presents a promising avenue for assessing individual student competencies in the future.

Overall, the StoiCoLe model could contribute to a clear depiction of subsequent steps in competency development and, subsequently, assist in determining the types of support learners require to advance to higher competency levels, particularly in the algorithmic aspects of stoichiometry.

Theoretical background

The competency-based approach

Competencies are defined as "dispositions that are acquired and required to successfully cope with specific situations or tasks" (Koeppen *et al.*, 2008, p. 62). Therefore, a competent person is characterised by having necessary dispositions to solve domain-specific problems (Weinert, 2001; Hartig and Klieme, 2006). Specific dispositions are required in different situation making competence a context-specific construct, which is acquired by mastering domain-specific or real-life situations (Koeppen *et al.*, 2008; Blömeke *et al.*, 2015). To solve a problem, required dispositions are used leading to observable performances (Blömeke *et al.*, 2015). Consequently, understanding proficiency as a competence and learning as developing competences involves the idea of transfer: "competent" students can successfully perform within the same family of problems (Krell *et al.*, 2022).

Competency models and competency levels

Competency models provide detailed representations of competencies within a particular subject area (Klieme *et al.*, 2003; Schecker and Parchmann, 2006). Competency levels are "specified by cognitive processes and actions of a certain quality that (…) [learners] can master at that level, but not at lower levels (Klieme *et al.*, 2003, p. 76). Drawing on the work of Vorholzer and von Aufschnaiter (2020), we have selected the dimensions of "interconnectedness" and "cognitive processes" to charaterise competency levels in stoichiometry.

The competency classification in the dimension "interconnectedness" is based on the increasing interconnectedness of knowledge elements that are required for solving problems. As the complexity of stoichiometric problems increases, so does the need of interconnected knowledge elements to solve them. This increased number of elements is associated with increased mental demand, which contributes to an increase in task difficulty (Niaz, 1988, 1989).

This dimension is divided into two sub-dimensions (Commons *et al.*, 1998; Vorholzer and von Aufschnaiter, 2020). The sub-dimension "number of elements to be connected" reflects that an increase in the required content – such as facts, relationships or concepts – leads to enhanced complexity (Kauertz *et al.*, 2010; Vorholzer and von Aufschnaiter, 2020). The sub-dimension "connection character" attributes the increase in complexity to a change in the relationship between the elements. For instance, a shift from isolated facts to interconnected relationships represents a change in character, thereby increasing complexity. Furthermore, the way in which the elements are connected – whether unconnected, monocausal or interdependent – also represent a change in character (Vorholzer and von Aufschnaiter, 2020). Calculating the amount of substance, masses, volumes or other important properties of substances, requires connected and interdependent thinking.

The second dimension "cognitive processes" describes required cognitive processes (Schecker and Parchmann, 2006; Kauertz *et al.*, 2010; Vorholzer and von Aufschnaiter, 2020). According to sequential processes of information processing defined by Weinstein and Mayer (1986), Kauertz *et al.* (2010), describe four cognitive processes of increasing complexity: reproducing, selecting, organising and integrating. This aligns well with the demands in stoichiometry: students must reproduce and select central concepts of stoichiometry as well as corresponding calculations (such as the stoichiometric equations). Also, they must organise and integrate the concepts and calculations according to the specific property to be calculated.

The StoiCoLe model: describing levels of competency in stoichiometry

Following the principle that competencies necessitate content knowledge, we have established content knowledge as a foundational level required at all competency levels. This content knowledge includes, for instance, the recall of stoichiometric quantity equations, quantities and units as well as the chemical symbols of the elements or formulas of trivial substance names. While content knowledge is indispensable for solving problems at higher competency levels, it alone is insufficient; it serves as a prerequisite rather than a complete solution.

The identification of the competencies necessary for solving algorithmic problems in stoichiometry was based on the sub-problems defined by Gulacar and Fynewever (2010), which have been further refined for this study. In the StoiCoLe model, these essential competencies are termed "sub-competencies" and are split into three distinct categories: "Basic equations," "Extension equations," and "Reaction equations". The first competency level is characterised by the ability to apply the

Humans and many other living organisms obtain their energy from the metabolisation of carbohydrates. Determine the molar mass of glucose ($C_6H_{12}O_6$).

Molar Mass:
M (H) = 1,01 g/mol   M (C) = 12,01 g/mol   M (O) = 16,00 g/mol

Fig. 1 Example item for the first competency level.

sub-competencies within each category. For instance, demonstrating competency at this level includes the ability to determine the molar mass from a given chemical formula (see Fig. 1).

The following competency levels were defined on the basis of the dimension "interconnectedness" and "cognitive processes". In accordance with the framework proposed by Vorholzer and von Aufschnaiter (2020), the task complexity related to the interconnectedness dimension is influenced by both the number of sub-competencies required to solve the task and the character of their connection.

The sub-dimension "connection character" is reflected in the fact that an understanding of the connections between the sub-competencies becomes necessary. Furthermore, the increasing interconnectedness changes the required cognitive processes. While simple stoichiometric tasks require one or the selection of a few sub-competencies, more complex tasks demand the selection and organisation of several sub-competencies. Consequently, learners must possess not only the requisite content knowledge and sub-competencies but also the procedural knowledge needed to effectively select and organise them. In the example task for the second competency level (see Fig. 2), four sub-competencies of two categories and their organisation are required.

In order to solve this task, students have to be able to
1. determine the molar mass of glucose ($C_6H_{12}O_6$)
2. determine the amount of substance ($C_6H_{12}O_6$)
3. set up the amount of substance ratio
4. determine the volume (CO_2)

Additionally, students must integrate these sub-competencies into a coherent problem-solving process, which is often (mono)causal; thus, a lack of content knowledge or a missing sub-competency cannot be compensated for. For instance, determining the amount of substance of glucose is impossible without knowing how to determine molar mass. However, tasks can often be approached through various solution paths utilising different sub-competencies, depending on the available information and the learners' preferred methods. For example, tasks can be resolved using ratios (proportionality method) or different

Humans and many other living organisms obtain their energy from the metabolisation of carbohydrates. The reactions that take place can be simplified according to the following reaction equation.
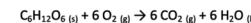
$$C_6H_{12}O_{6\,(s)} + 6\,O_{2\,(g)} \rightarrow 6\,CO_{2\,(g)} + 6\,H_2O_{\,(l)}$$

Determine the volume (L) of carbon dioxide under standard conditions (1 bar, 25 °C) that is released during the reaction of 1 kg glucose.

Molar Mass:
$M(\text{H}) = 1{,}01\,\frac{g}{mol}$ $M(\text{C}) = 12{,}01\,\frac{g}{mol}$ $M(\text{O}) = 16{,}00\,\frac{g}{mol}$

Constants:
$R = 83{,}145\,\frac{bar \cdot ml}{mol \cdot K}$ $N_A = 6{,}022 \cdot 10^{23}\,\frac{1}{mol}$

Fig. 2 Example item for the second competency level.

Humans and many other living organisms obtain their energy from the Metabolization of carbohydrates, from the reaction between glucose ($C_6H_{12}O_6$) with oxygen to form carbon dioxide and water.

Determine the volume (L) of carbon dioxide under standard conditions (1 bar, 25 °C) that is released during the reaction of 1 kg glucose.

Molar Mass:
$M(H) = 1,01\frac{g}{mol}$ $M(C) = 12,01\frac{g}{mol}$ $M(O) = 16,00\frac{g}{mol}$

Constants:
$R = 83,145\frac{bar \cdot ml}{mol \cdot K}$ $N_A = 6,022 \cdot 10^{23}\frac{1}{mol}$

Fig. 3 Example item for the third competency level.

quantity equations and the amount of substance, such as the mole method (Schmidt and Jignéus, 2003). In the example task at the second competency level, the volume of carbon dioxide can be calculated either using the general gas equation or the molar volume. If the density were provided, a third approach would be possible.

The third competency level is distinguished by the requirement to select and organise an even greater number of sub-competencies across all categories to solve the problem effectively. This is illustrated by the example task provided for the third competency level (see Fig. 3).

With each competency level, stoichiometric tasks increasingly reflect real-world problems encountered in laboratory or industry settings. At each level, problem-solving necessitates that students select and organise a larger number of sub-competencies into coherent steps:

1. *convert substance's names into chemical formulae*

2. *set up the reaction equation*

3. *balance the reaction equation*

4. determine the molar mass of glucose ($C_6H_{12}O_6$)

5. determine the amount of substance ($C_6H_{12}O_6$)

6. set up the amount of substance ratio

7. determine the volume (CO_2)

In summary, the StoiCoLe model is grounded in the idea that students demonstrate a higher stoichiometry competency as they progressively integrate sub-competencies into coherent solutions for increasingly complex stoichiometric problems. The model comprises three competency levels, as outlined in Fig. 4. According to this model, content knowledge forms the foundation for being competent in stoichiometry. This includes essential quantity equations, relevant stoichiometric quantities (*e.g.* mass, volume or amount of substance) and their associated units, as well as components of the chemical formula language (*e.g.* element symbols or chemical formula of substances with trivial names).

The first competency level contains the sub-competencies, which are divided into the three categories "Basic equation", "Extension equations" and "Reaction equations", with the exception of the sub-competency "Determine of amount of substance ratios". Hence, students at the first competency level are therefore able to successfully apply the sub-competencies independently of each other. The sub-competency categories serve to form the second and third competency level.


3. Competency level

Students are able to use sub-competencies of all three categories to solve stoichiometric problems.

2. Competency level

Students are able to use sub-competencies of two categories to solve stoichiometric problems.

1. Competency level

Students are able to use sub-competencies of the categorie....

Basic Equation	Extension Equation	Reaction Equation
• Apply and transform of: $n = \frac{m}{M}$ • Determine the molar mass	• Apply and transform of : $n = \frac{V}{V_M}$ $n = \frac{p \cdot V}{R \cdot T}$ $n = \frac{N}{N_A}$ $n = c \cdot V$	• Determine the chemical formula • Set up the reaction equation • Determine the oxidations numbers • Balance the reaction equation

• Determine the amount of substance ratio

Content knowledge level

Students have content knowledge about

Quantities & units; Stoichiometric equations, Chemical formula language

Fig. 4 Stoichiometric competency level model: StoiCoLe model.

Students at the second competency level are able to successfully use sub-competencies from two out of the three categories. This results in three competency categories for the second competency level: basic equation and extension equation, basic equation and reaction equation as well as extension equation and reaction equation.

Accordingly, at the third competency level, sub-competencies of all three categories are necessary to solve a stoichiometric task. Based on these considerations, we assume that reaching higher competency levels is more difficult for students. In addition, we assume an increase in processing time due to the higher task difficulty caused by the larger number of sub-competencies and their organisation.

Due to the broad subject area of stoichiometry, it must be clarified that this competency model "does not capture all attributes [...] of the original [subject area of stoichiometry], but only those [...] that appear [in this first approach as most] relevant" (Stachowiak, 1973, p. 132). For example, the limiting reagent, the percentage yield and the mathematical determination of chemical formulas were not taken into account in the StoiCoLe model.

Research questions

Based on these considerations, we sought to determine whether these theoretical assumptions are supported by empirical data. We focused on the following three research questions (RQ):

1. What psychometric properties are revealed by a test designed to assess the stoichiometric competencies outlined in the StoiCoLe model?

2. To what extent do item difficulty and processing time confirm the competency levels of the StoiCoLe model?

3. At which levels of the StoiCoLe model do students demonstrate competency?

Methods

Based on the StoiCoLe model, a test assessing competencies on all levels was created (see ESI†). The content knowledge (item 1–9) and sub-competencies of the first competency level (item 10–31) are each represented by two items. For the second competency level (items 32–36), each competency category is also represented by two test items, with the exception of the categories "extension equations" and "reaction equations," which are each represented by only one item. The third competency (item 37–40) level includes four test items.

In total, the test comprises 40 items. To ensure that all students address at least one task at the third competency level, the sequence of items was adjusted after the first competency level. Specifically, the linear sequence was modified to an alternating pattern, where two items from the second competency level are followed by one item from the third competency level.

Data collection

The developed test instrument for the StoiCoLe model was administered to students of the module "General Chemistry I" in the winter semesters 2021/22 and 2022/23 at the Leibniz Universität Hannover. To increase the sample size, particularly at the higher competency levels, we collected additional data as part of an experimental study in the 2024/25 semester. A shortened version of the stoichiometry test, consisting of 12 items, was compiled to reduce the processing time. To ensure that all the different item types from the content knowledge level (7 items) and the first competency level (11 items) were still covered, an item pool of 18 items was created so that each content knowledge or sub-competency is represented once. The items were distributed over five booklets, each with six item clusters and two items, according to a balanced incomplete block design (BIBD) (Frey *et al.*, 2009). In total, the five shorter test versions consisted of six items compiled from the content knowledge level and the first competency level, as well as two items each from the second competency level (item 32 and a modified version of item 34), the third level (items 39 and 40) and two new items that take the limiting reagent into account. Due to the random allocation of booklets and unanswered items, there is a slightly different number of responses for the items.

The implementation was carried out in accordance with local data protection regulations as required by the European Union (General Data Protection Regulation). Participation was voluntary and all students signed a consent form. Data collection was conducted as part of a regular classroom exercise, using the open-source online survey tool LimeSurvey (Limesurvey GmbH, 2003) in the first and second semester and a paper and pencil test in the third semester. By switching the data collection from digital to analogue format, we wanted to gain an additional insight into the processing steps and documentation of the students. Participants had to use a mobile device with internet access to carry out the test in LimeSurvey and a calculator to carry out the stoichiometric calculations in the first and second semester. In the third semester, participants only needed a pen and a calculator. All Participants were allowed to take handwritten notes. Other aids, such as recipe sheets or the use of the internet, were prohibited. The processing time was set at 105 minutes for the full stoichiometry test and 45 minutes for the shortened version. The processing time of individual items was only recorded *via* LimeSurvey in the first and second data collection.

Data sample

The demographic data of the sample are listed in Table 1.

Data analysis

Item difficulty was assessed using item response theory (IRT) and analysed in conjunction with student processing time. IRT is a probabilistic test theory that describes the probability of a certain response of a person to an item (Bond *et al.*, 2020). This study utilised the Rasch model, which is suited for dichotomous item responses. Therefore, the collected data were converted into a dichotomous format (0 incorrect item/1 correct item). Because not all students completed the test, it was difficult to determine whether incomplete responses were due to students' lack of ability or simply because they ran out of

Table 1 Demographic characteristics of the sample

Factor	Sample	
Age	$N = 280$	
	$M = 20.6$	$SD = 2.6$
Gender	$N = 289$	
Male	147	(51.9%)
Female	134	(46.4%)
Divers	3	(1.0%)
No response	5	(1.7%)
Semester	$N = 285$	
1. Semester	223	(78.3%)
3. Semester	45	(15.8%)
5. and higher semester	11	(3.9%)
Study program	$N = 289$	
Subject-related	219	(75.8%)
Bachelor of Chemistry	126	(43.6%)
Bachelor of Biochemistry	76	(26.3%)
Bachelor of Nanotechnology	14	(4,8%)
Interdisciplinary Bachelor of Science	2	(0.7%)
Bachelor of Physics with Chemistry	1	(0.4%)
Teaching-related	68	(23.5%)
Interdisciplinary Bachelor's degree	57	(19.7%)
Bachelor's degree in Technical Education	9	(3.1%)
Certificate program for a third subject	2	(0.7%)
Special program at the authors' university	2	(0.7%)

time. Therefore, only tasks with responses and tasks without a response followed by another task with response were included in the data analysis (Ludlow and O'Leary, 2000). We did not use imputation methods, as these are always associated with a bias in the results. Item and person parameters for the Rasch model were estimated using the package *TAM* (Robitzsch *et al.*, 2024) in the open-source software *R* (R Core Team, 2025) with the integrated development environment *RStudio* (Posit, 2025). The Rasch model evaluates a single stoichiometric competency, which is relevant to solve the tasks.

Item parameters were estimated using the *marginal maximum likelihood* method (MML), while personal parameters and their reliability were determined through the *Expected A Posteriori* method (EAP). Parameter visualisation was done *via* a Wright Map, which displays the distribution of both person and item parameters, providing a comparative overview of the data (Wilson, 2023). The visualisation was created with the package *WrightMap* (Torres Irribarra and Freund, 2024).

In line with the StoiCoLe model's assumption, item difficulty is expected to increase across the competency levels. To verify this assumption, the mean item difficulties for each competency level were determined and compared. Furthermore, the students' processing time was measured for each item in the first and second year. According to the competency classification, differences in the mean processing times are expected to enhance between the content knowledge level and the competency levels. Therefore, the mean processing time was determined and further compared. To prevent distortion from items that were skipped, only items with a response or a processing time of more than 60 seconds (especially for items

of the 2nd and 3rd competency level) were included in the analysis. The mean processing time was also determined and compared.

In order to verify whether the items can distinguish between different levels of difficulty, item reliability, item separation and item strata value were calculated (Wright and Masters, 1982). The separation and the strata value provide insights into how many distinct strata (groups) the continuously measured values (item and person parameters) can be divided into (Wright and Masters, 1982; Boone *et al.*, 2014). While the separation value is suitable for large normal distributions, the strata values are suitable for distributions that are heavy-tailed and should include extreme performance levels (Bond *et al.*, 2020). The person reliability, the person separation and the person strata value were also calculated in order to verify how many groups the sample of students can be divided into (Wright and Masters, 1982).

Finally, the competency level of each student was assigned based on their person parameters and the median of the item difficulty as the threshold value for the respective competency level. We have based the probability of success (65%) on other assessments, such as PISA, in order to determine with greater certainty whether learners can answer the items correctly (Boone *et al.*, 2014). The median of item difficulty was chosen for this purpose because (a) it is less susceptible to outliers and (b) with person parameters at the median, at least half of the tasks at that competency level should be solvable (see Hartig, 2007).

The final model validation was approached through an argumentation-based approach (Cronbach and Meehl, 1955; Kane, 1992). This type of "validation is the process of gathering supporting evidence for the intended test score interpretations" and checking it (Hartig *et al.*, 2020, p. 535). Consequently, the validity of test score interpretations was evaluated based on the discussed arguments related to item difficulty and processing time.

Results

First RQ: psychometric properties of the test instrument

The test instrument values of the Rasch-modelling are summarised in Table 2. Detailed mean square (MNSQ) infit and outfit values, along with associated standardized z-scores (ZSTD), are provided in Table 3.

Second RQ: analysing item difficulty and processing time of the competency levels of the StoiCoLe model

An analysis of variance (ANOVA) revealed significant differences in the mean item difficulty across the competency levels $F(1, 38) = 53.8$; $p < 0.001$ (see Fig. 5). Overall, the mean item difficulty increases from the content knowledge level to the third competency level (see Table 4). The mean processing time significantly differs between the competency levels $F(1, 38) = 72.03$; $p < 0.001$ (see Fig. 6). Overall, the average processing

Table 2 Test values of the unidimensional Rasch modelling

Model features	Test values
Person reliability	0.78
Observed variance	1.39
Separation index	1.88
Number of person strata	2.83
Person parameters	
Max	3.12
Min	−3.78
Average	0.00
SD	1.04
SE	0.53
Item reliability	0.99
Observed variance	3.06
Separation index	8.24
Number of item strata	11.32
Task parameters (65%)	
Max.	3.83
Min.	−3.07
Average	0.06
SD	1.74
SE	0.20

Fig. 5 Boxplot of item difficulty of content knowledge and competency levels.
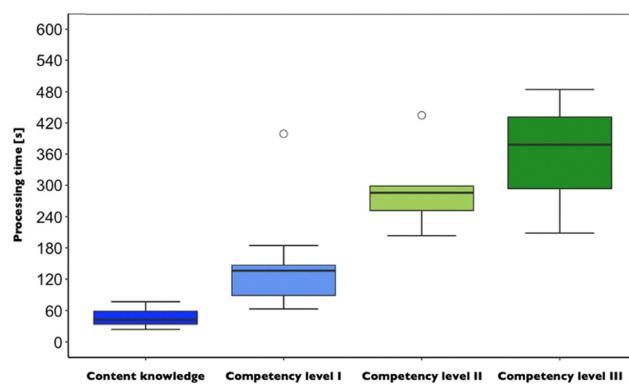
Fig. 6 Boxplot of processing time of content knowledge level and competency levels.

time increases from the content knowledge level to the third competency level (see Table 4).

post hoc tests using a Bonferroni correction showed no significant difference in item difficulty and processing time between the content knowledge level and the first competency level as well as the second and third competency levels (see Table 5).

Third RQ: competency assessment

The classification of the students' competencies is shown in Table 6. Overall, it can be seen that the number of students is decreasing across the competency levels. Using the median of the

item parameters (item difficulty) of each competency level with a success probability of 65% as the threshold value, 259 (86.5%) of the

Table 3 Infit and outfit values with standardised z-scores (ZSTD)

		Infit		Outfit				Infit		Outfit	
Item	Responses	MNSQ	ZSTD	MNSQ	ZSTD	Item	Responses	MNSQ	ZSTD	MNSQ	ZSTD
1	185	0.89	−0.71	0.92	−0.55	21	219	1.06	0.95	1.18	**2.55**
2	220	1.00	0.07	1.09	0.76	22	228	1.00	−0.03	1.02	0.21
3	185	1.00	0.05	0.96	−0.16	23	185	0.88	−0.86	0.81	−1.52
4	223	1.14	1.98	1.28	**3.78**	**24**	184	0.86	−1.76	0.78	**−2.86**
5	220	0.95	−0.25	0.81	−1.23	25	208	0.94	−0.64	0.94	−0.63
6	214	0.97	−0.30	0.99	−0.09	26	210	1.10	1.38	1.15	1.99
7	220	1.07	0.83	1.05	0.59	27	184	0.99	−0.36	1.00	0.06
8	206	1.14	**2.11**	**1.32**	**4.36**	28	184	0.98	−0.19	1.11	0.97
9	218	1.00	0.01	0.91	−0.88	29	219	0.96	−0.57	0.89	−1.48
10	237	1.02	0.17	**1.33**	1.95	**30**	203	1.08	1.15	1.15	**2.17**
11	185	1.01	0.11	**1.38**	1.23	31	183	1.08	0.65	1.27	1.91
12	185	1.17	**2.16**	**1.42**	**4.94**	**32**	282	0.90	−1.60	0.83	**−2.86**
13	213	0.97	−0.43	0.95	−0.74	**33**	183	0.92	−0.52	0.71	**−2.05**
14	215	1.08	0.85	1.24	**2.43**	**34**	153	0.89	−1.14	0.71	**−3.15**
15	185	1.01	0.16	1.05	0.55	35	145	1.02	0.19	1.08	0.71
16	220	0.99	−0.09	1.01	0.18	36	108	1.01	−0.13	0.75	−1.26
17	185	1.02	0.28	1.04	0.61	**37**	183	0.89	−0.83	0.72	**−2.40**
18	185	0.96	−0.36	1.02	0.21	**38**	132	0.85	−0.48	**0.51**	**−2.13**
19	207	0.94	−0.48	1.02	0.14	**39**	193	0.92	−0.35	**0.66**	−1.91
20	185	0.87	−1.70	0.78	**−3.07**	40	122	0.98	−0.18	0.96	−0.45

Note: Item with infit and outfit values outside the range between 0.7 and 1.3 are shown in bold. In addition, the items with a standardized z-score outside the range between −2 and 2 are shown in bold.

Table 4 Difference in the average item difficulty and the processing time across the competency levels

Level	Items	Item difficulty (50%)			Item difficulty (65%)			Processing time [s]		
	n	M	Mdn	SD	M	Mdn	SD	M	Mdn	SD
Content knowledge level	9	−1.82	−1.88	0.86	−1.20	−1.26	0.86	41.2	37.6	14.7
1st competency level	22	−1.07	−1.13	1.17	−0.45	−0.50	1.17	115.7	99.5	76.5
2nd competency level	5	1.72	1.50	0.66	2.34	2.12	0.66	295.4	286.0	86.4
3rd competency level	4	2.25	2.49	1.11	2.87	3.10	1.11	418.3	431.8	166.2

Table 5 Bonferroni corrected *post hoc* analysis of item difficulty and processing time

Level	Item difficulty			Processing time		
	1st competency level	2nd competency level	3rd competency level	1st competency level	2nd competency level	3rd competency level
Content knowledge level	0.49	0.001	0.001	0.16	>0.001	>0.001
1st competency level	—	0.001	0.001	—	>0.001	>0.001
2nd competency level	—	—	1.00	—	—	0.18

Table 6 Students' competency classification

Level	Success probability (50%)			Success probability (65%)		
	Mdn	Number of students	Achieved [%]	Mdn	Number of students	Achieved [%]
Content knowledge level	−1.88	278	[96.2]	−1.26	250	[86.5]
1st competency level	−1.13	247	[85.5]	−0.50	205	[70.9]
2nd competency level	1.50	16	[5.5]	2.12	5	[1.7]
3rd competency level	2.49	2	[<0.01]	3.10	1	[<0.01]

students achieved the content knowledge level, 205 (70.9%) the first competency level, five (1.7%) the second competency level and one (<0.01%) the third competency level. With a success probability of 50%, 278 (96.2%) of the students achieved the content knowledge level, 247 (85.5%) the first competency level, five (5.5%) the second competency level and one (<0.01%) the third competency level.

The distribution of the competency achieved by the students (person parameters) and the item difficulty (item parameters) with a 65% success probability is represented in a Wright Map (see Fig. 7).

Discussion

The objective of this study was to introduce the StoiCoLe model as a comprehensive framework for assessing algorithmic competencies in stoichiometry and to position students within defined competency levels. To achieve this, a competency test was developed according to the StoiCoLe model and its psychometric properties were investigated.

Rasch analysis indicated that the collected data exhibit adequate psychometric properties. The evaluation was based on accepted infit and outfit value ranges, *i.e.* the tolerated deviation of the responses from the Rasch model. The value range depends on the test type, the sample size and the desired measurement accuracy (Bond *et al.*, 2020). Item fit values for the mean square (MNSQ) infit and outfit values in a range of 0.7

to 1.3 are acceptable (Wright and Linacre, 1994; Bond *et al.*, 2020). The infit values for all items are within this acceptable range and thus show that the items have a good fit for people in the medium ability range. In contrast, there are six mean square outfit values outside the acceptable range, indicating that there are outliers within items when considering all students that require closer examination.

The items that show an underfit of the model are item 8, 10, 11 and 12. According to this, some students with competencies on a higher level show difficulties in rather easy tasks, such as naming the quantity equation of amount of substance and number of particles ($N = n \times NA$), determining the molar mass and determining the chemical formula.

The difficulties with the quantity equation between amount of substance and number of particles is probably due to the fact that this quantity equation is not or rarely used in stoichiometry due to the amount of substance as a measure of the number of particles. This relationship relevant to an understanding of the mole concept poses a difficulty for learners and is criticized for its lack of explication (Steiner, 1986; Karp, 1988). In the case of the determining of the molar mass, it is possible that some careless mistakes were made. Determining chemical formulas requires content knowledge, depending on whether it is a molecular or ionic compound. While both types of compounds require knowledge of the symbols associated with the elements, molecular compounds require knowledge of the Greek and Latin prefixes and the numbers they represent. For ionic
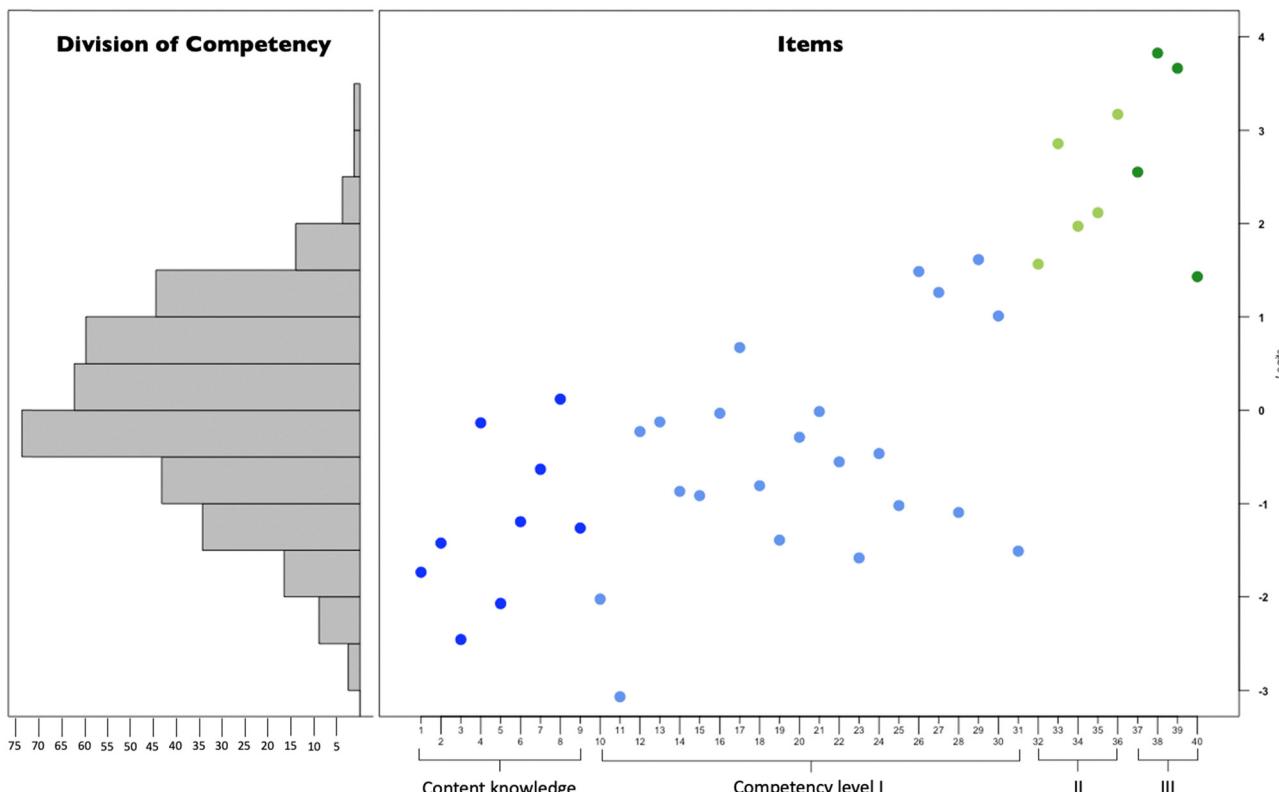
Fig. 7 Wright map of the unidimensional Rasch model.

compounds, the charge numbers of the atomic and molecular ions as well as the composition of molecular ions based on the name are necessary, *e.g.*, for the sulphate ion (SO_4^{2-}). This mirrors the difficulties with chemical formula language described by Gulacar and Fynewever (2010) and Taskin and Bernholt (2014).

The items that show an overfit of the model and thus a significantly lower variance than expected from the Rasch model are item 38 and 39 (3rd competency level). However, it should be borne in mind that the majority of students of the sample are at the content knowledge and 1st competency level and the group of students who are at the 2nd and 3rd competency level and for whom a distinction would be less strict is significantly smaller. If students are unable to solve tasks at a lower level, it is highly likely that they will also be unable to solve tasks at a higher level. Against this background, the relevance of the items for the StoiCoLe model and the generally good infit values, we see no reason to remove items from the test.

However, we revised the items with regard to the standardized *z*-value (ZSTD) of the infit and outfit, which indicates how likely the misfit of the Rasch model is (Bond *et al.*, 2020). Standardized *z*-values within the acceptable range of −2 to 2 can be neglected (Boone *et al.*, 2014; Bond *et al.*, 2020). While the mean square values show a decreasing deviation with increasing sample size, the standardized *z*-values show a significant deviation. Against this background, it is worth looking at the deviating standardized *z*-values and checking the

associated items (Bond *et al.*, 2020). A revised version of the stoichiometry test contains adaptations of the conspicuous items (see ESI†). These include item 3, as hydrogen peroxide (H_2O_2) is not, strictly speaking, a trivial name, item 14 and 15, which do not clearly reflect the sub-competency "Setting up the amount of substance ratio". Furthermore, item 35 was revised as it was possibly problematic because the reaction equation was not explicitly addressed.

The observed variance in person parameters indicates at a first glance that the test effectively differentiates between students. A closer look at the person reliability, the person separation index and the number of person strata indicate the existence of three different person ability levels (Fisher, 1992; Boone *et al.*, 2014). This might contradict the assumption of the StoiCoLe model with four levels and will be discussed later with regard to model adjustments. In contrast, item reliability, the item separation index and the number of item strata show that the items are capable of depicting and distinguishing between different levels of difficulty. However, item reliability should be viewed critically in view of the large sample size (Bond *et al.*, 2020).

Both item difficulty and processing time were utilised to validate the model assumptions. Notably, processing time partially confirms the competency level assumptions, although no significant difference was observed between the content knowledge level and the first competency level as well as between the second and third competency level. Additionally, some items, such as "Balance the reaction equation"

(1st competency level) and item 33 (2nd competency level), had processing time differences of $\Delta 399.5$ seconds and $\Delta 435.3$ seconds, respectively, which were significantly longer than those for most items at the respective competency levels.

Item difficulty also partially supports the StoiCoLe model. Specifically, within the content knowledge level and the 1st competency level, there is substantial variance and considerable overlap among the items. The notably higher item difficulty in the content knowledge level suggests that essential content, such as stoichiometric quantity equations, may not have been adequately learned or internalised. The differences between the sub-competencies are consistent with Gulacar and Fynewever (2010), who demonstrated that sub-problems in stoichiometry vary in difficulty. For example, the sub-competency "Determine the molar mass" has a notably low item difficulty ($M = -3.17$), whereas "Set up the reaction equation" has an item difficulty ($M = 0.76$) comparable to that of items at the 2nd and 3rd competency levels. However, it must be pointed out that the chemical formulas were not given in the two tasks for setting up the reaction equation and therefore the sub-competency "Determining the chemical formula" was also required for the solution. These items were, as already mentioned, adapted in the revised version of the stoichiometry test. Furthermore, there is considerable variance among items within the same sub-competency, such as "Determine the oxidation number" ($\Delta 2.71$) and "Balance the reaction equation" ($\Delta 2.52$). This variance between and in the sub-competencies has a corresponding effect on the difficulty of the higher competency levels, regardless of the number of sub-competencies and cognitive processes. This interaction is discussed later in relation to an adaptation of the StoiCoLe model.

There are three possible explanations for the high variance in item difficulty within the sub competencies. Firstly, it is likely that various factors influence item difficulty within a given sub-competency. Regarding the sub-competency "Balance the reaction equation", Segedinac et al. (2018) have shown that balancing a reaction equation becomes more challenging with an increasing number of substances or when forming a compound from elements obtained from different sources. Accordingly, other influential factors may affect item difficulty across different sub-competencies. In the sub-competency "Determine of oxidation numbers", increased difficulty could result from factors such as a higher number of different atoms, the presence of the same atoms with multiple oxidation states, or the type of chemical representation used (chemical formula item 28; versus structural formula item 29).

Second, the variation in item difficulty may be attributed to differences in item type and the approaches or methods employed by learners. For example, there are four different methods that can be applied to balance reaction equations, including trial and error or inspection, algebraic, oxidation number and half-reaction method (Herndon, 1997). While for small numbers of reactants and products (e.g. item 31 or 40) a trial-and-error approach (i.e. inspection method) might be the most effective, for larger numbers (e.g. item 30 or 38), the algebraic, oxidation number or half-reaction method might be

more effective. Supporting evidence for varying effectiveness of these methods is provided by of Charnock (2016) and Chibuye and Mupela (2019), who compare the inspection and algebraic methods in terms of their effect on performance and indicate that students taught the algebraic method performed significantly better. On the contrary, Kolb (1981), Tóth (1997) and Guo (1997) were able to show that even complex chemical equations can be solved using a systematical inspection method. In summary, the methods are sometimes described as superior or inferior (see Yde, 1989; Herndon, 1997). In contrast, Yalman (1959) and Kolb (1979) assert that there is no single superior method and that only the outcome is important, regardless of whether one, several or a combination of methods is used. We propose that the variation in item difficulty within the sub-competency "Balance the reaction equation" may be attributed to the use of unsystematical trial-and-error approaches by students, rather than systematic inspection or other systematic balancing methods. This idea is supported by the observed lower item difficulty items with reactions involving fewer reactants and products, as well as the longer processing times required for more complex reactions.

Third, competencies are acquired on concrete chemical content. Consequently, task difficulty may depend on two factors: whether the competency was acquired using the same content (Walpuski and Ropohl, 2014) and the extent to which learning transfer, including reasoning processes such as analogical reasoning, is involved (Alfieri et al., 2013). Future research could explore the aspects of learning transfer and its impact on stoichiometric competencies. It is possible that the sub-competency "Determine the oxidation number" was acquired based on potassium permanganate (item 28) as a learning content, resulting in lower item difficulty for this content. For a competency assessment, it is therefore important that students complete several items of a competency (level) in order to prevent misattribution of their abilities. However, accommodating the necessary time poses a challenge. To address this, we included at least two tasks per sub-competency or level in this study. In the future, further improvements should be made to shorten the stoichiometry test, for example by using a booklet design, as was already done in the third data collection.

Between the 1st and 2nd competency level, the results indicate a clear gradation between the majority of the items, aligning with the model assumptions. In particular, the item difficulties at the higher competency levels illustrate that, in addition to the increasing number of sub-competencies, the change in cognitive processes from the reproduction and application of sub-competencies to the selection and organisation of these to solve a stoichiometric problem could be the decisive factor causing difficulties, as otherwise significantly more students would have reached the higher competency levels.

However, there is an overlap between items of the 2nd and 3rd competency level, suggesting a lack of clear progression and contradicting the model's predictions. In this context, it should be taken into account that the number of students who

have the ability to solve items of the 2nd and 3rd competency level with a 65% success probability lies under 2% in the sample. A better differentiation between the 2nd and 3rd competency levels would therefore have required more students being competent on these levels. Nevertheless, clear differences can be observed in the items of the two competency levels. With regard to the second competency level, items with provided or simple reaction equation result in lower item difficulty. This result is in line with observations from Tang *et al.* (2014) and Schmidt (1990) that stoichiometric tasks with given chemical formula or reaction equation are easier. In addition, item 33 and 38 required the transfer of a systematic substance name, such as iron(II,III)-oxide into the corresponding chemical formula. Based on the students' responses, we can confirm the difficulties in dealing with chemical formula language described by Gulacar and Fynewever (2010) and Taskin and Bernholt (2014). On the other hand, items that required connecting stoichiometric quantity equations generally exhibited lower item difficulty. The minimal variation in item difficulty among applying stoichiometric quantity equations at the 1st competency level may reflect students' difficulties with some mathematical operations (Lazonby *et al.*, 1982; Scott, 2012).

On the 3rd competency level, especially item 40 displayed a particularly lower item difficulty compared to the other items. This is likely because item 40 featured an easier reaction equation, as suggested by Segedinac *et al.* (2018), relative to the other items at this level.

Distribution and competency level

Finally, students were assigned to their appropriate competency level based on the determined item difficulties and person parameters. The IRT modelling results demonstrate that the majority of students are assigned to the content knowledge level and 1st competency level and are able to recall relevant content knowledge and master the individual sub-competencies with a sufficiently high ($\geq$65%) success probability. Only a minority of students can select, apply and organise multiple sub-competencies effectively. Even when considering a success probability of 50%, which leads to a small increase in the number of people within the higher levels, the general pattern does not change and leads to the same conclusion.

The result confirms the high dependence of the success rate of algorithmic tasks on the type, as shown by Stamovlasis *et al.* (2004, 2005). While simpler algorithmic tasks—comparable to sub-competencies of the 1st competency level in our framework—show high success rates, more demanding algorithmic tasks, which align with higher competency levels, exhibit lower success rates. This challenges the notion that algorithmic tasks generally have a higher success rate compared to conceptual tasks, as reported in several studies (Nurrenbern and Pickering, 1987; Sawrey, 1990; Mason *et al.*, 1997; Cracolice *et al.*, 2008). Algorithmic tasks typically have higher success rates for two main reasons. First, traditional chemistry education has emphasised these types of problems. Second, students often memorise solution paths without fully understanding the underlying chemical concepts

(Lythcott, 1990; Pickering, 1990; Nakhleh, 1993; Smith and Metz, 1996; Smith *et al.*, 2010). Although Stamovlasis *et al.* (2004, 2005) differentiate between conceptual and algorithmic competencies and a high algorithmic competency not necessarily indicating high conceptual understanding, there is evidence that conceptual understanding positively influences algorithmic problem-solving abilities (Niaz, 1989, 1995a, 1995b; BouJaoude and Barakat, 2003). This is shown by the fact that conceptual learners solve algorithmic problems correctly and with fewer steps (Mason and Crawley, 1994; BouJaoude and Barakat, 2003). We agree that the primary goal in chemistry education is to build conceptual understanding, from which algorithms should be derived (Bodner, 1987; Niaz, 1989; BouJaoude and Barakat, 2000; Arasasingham *et al.*, 2004). However, this also means that both algorithmic and conceptual tasks are important and competent learners should be able to solve both types of tasks (Bodner, 1987; Frank *et al.*, 1987; Smith *et al.*, 2010). The difficulties students face in stoichiometry, as identified in other studies (Dierks, 1981; Schmidt, 1990; Huddle and Pillay, 1996; Boujaoude and Barakat, 2000, 2003; Arasasingham *et al.*, 2004; Dahsah and Coll, 2008; Shadreck and Enunuwe, 2018; Rosa *et al.*, 2022) emphasise the need for improved stoichiometric instruction, even for algorithmic tasks.

Future directions: linking the prior research with the competency-approach in stoichiometry

All in all, we conclude that the study results confirm the usability of the competency approach in the field of stoichiometry. Identifying students' competency levels in stoichiometry enables researchers and educators to gain a comprehensive view of learners, which can be valuable for assessing entry-level performance in chemistry programmes, for example. This approach also facilitates the grouping of students with varying competencies, which is crucial for effective collaborative learning in chemistry (Heeg *et al.*, 2020). On the basis of the competency level and process data, for example, collaborative group formation can be carried out using machine learning techniques (Costaguta, 2015). Generating individual feedback for students regarding the aims defined by the StoiCoLe model is conceivable, too. By placing a stoichiometry task within the competency model, one can predict its difficulty. The IRT-based modelling further supports the development of digital, adaptive learning environment (Deonovic *et al.*, 2018). However, further research on test validity of the competency assessment is needed.

As described in other fields of competency-based research (Neumann, 2020; Krell *et al.*, 2022), locating our work in existing literature shows that linking students' cognitive processes to their learning results, as measured in our stoichiometry test, remains a subject of future research. Specific questions like "Which reasoning processes are used to be "competent" on a certain level of the StoiCoLe model?", "To which extent is transfer of learning need to become more competent?" "Which strategies are used by competent learners?", "What are concrete difficulties leading to not reaching a competency level?" or "Which learning processes have to be initiated in order to reach the next competency level?" arise in

this context. While the estimated person parameters *via* the test answers (item output) make it possible to locate the learners in the model and identify learning difficulties and learning goals, the analysis of processes with the help of learning analytics represents an approach for future research.

How plausible are alternative competency levels?

The results of the IRT indicate that modifications to the model are required. Two aspects in particular are problematic:

1. the high variance within the sub-problems.
2. the overlap between the competency levels.

In view of the high variance within the sub-competencies, one approach would be to further differentiate the sub-competencies, but this would also increase the complexity of the StoiCoLe model and make it more difficult for learners and teachers to use. A sensible balance needs to be struck here.

In particular, the overlap between the second and third competency levels suggests that there is no clear gradation. However, this finding should be interpreted cautiously, as it may be influenced by the small number of students, who were presumably more competent, that engaged with items at these higher levels. A notable issue is the higher difficulty of tasks requiring the sub competencies "Determining the chemical formula" and "Balancing the reaction equation" of the Reaction equation category. Conversely, tasks that required establishing connections between stoichiometric equations were found to be less difficult.

Based on this, the categories "Basic equation" and "Expansion equation" of the 1st competency level should be combined in a new category for mathematical–chemical calculations in the future StoiCoLe model. This category should also include the mathematical determination of the ratio formula and sum formula. Consequently, students at the 2nd competency level would be able to solve stoichiometric tasks using only sub-competencies from the mathematical–chemical category. Considering the clear difficulty of the tasks in the category "Reaction equation", the 3rd competency level should be characterized by the students being able to solve stoichiometric problems using sub-competencies from the mathematical–chemical category and sub-competencies from the "Reaction equation" category. Additionally, incorporating limiting reagent and percentage yield as advanced competency levels should be considered in future version of the StoiCoLe model.

Study limitations

The main limitation of this study is the asymmetrical distribution of person ability in the study sample. The majority of the participants are at the lower levels of the StoiCoLe model, resulting in less variance and insight at the higher competency levels. While having a sample size with more participants on the 2nd and 3rd competency levels would be beneficial, the current analysis, supplemented by processing time data, supports the assumptions of the model. However, it should be noted that the processing time decreases with increasing stoichiometric competency and was not considered separately (Tang *et al.*, 2014). Given the clear definitions of the 2nd and

3rd competency level and the comparable cognitive demands of the items, it is reasonable to expect similar findings with additional data. Future research could build on these findings by including a greater number of items and participants with different levels of competency. Studies incorporating think-aloud protocols (Gulacar *et al.*, 2013a, 2013,b) or learning analytics approaches (Gobert *et al.*, 2012, 2013; Ifenthaler *et al.*, 2018; Teig, 2024) could provide deeper insights into the processes involved in solving stoichiometric tasks and further validate the competency model.

Implications

On the basis of the StoiCoLe model, the competency of learners in algorithmic stoichiometry can be assessed and more focused learning opportunities can be derived.

1. Using the StoiCoLe model, educators can make *a priori* assumptions about the difficulty of stoichiometric tasks.

2. It maps the terrain of algorithmic competencies in this topic: it helps educators to define goals of instruction based upon the different competency levels. It can also be used for designing a curriculum across study programs aiming at highest competency level as final stage of competency development.

3. Compared to the framework of sub-problems from Gulacar and Fynewever (2010) and the framework of the multidimensional analysis system from Dori and Hameiri (2003), our model can provide learners and teachers with a more comprehensible orientation and concretise learning objectives. Here are some examples: locating students on a concrete competency level gives content-oriented implications on what students are able to carry out in stoichiometry.

● Learners who fall below the content knowledge level need to focus on fundamental stoichiometric content, such as stoichiometric quantity equations, units, and the chemical formula language.

● Learners who have not achieved the 1st competency level struggle with sub-competencies essential for progressing to higher levels.

● Learners who have not achieved the 2nd or 3rd competency level struggle with selecting and organising the correct sub-competencies. Additionally, it can point out difficulties with sub competencies, too.

Due to the interconnected and hierarchical structure of stoichiometric tasks, mastering these individual sub-competencies is crucial for advancement and cannot be bypassed. This approach highlights the significant challenge in teaching stoichiometry: ensuring that foundational knowledge and sub-competencies are thoroughly mastered and integrated before advancing to more complex tasks.

For the use of the stoichiometry test instrument, we recommend the use of a shortened version of the revised stoichiometry test instrument with regard to the effective use of time. However, one must be aware that this reduces the reliability of the location in the StoiCoLe model. Further development of the stoichiometry test instrument by improving validity and reliability according to Arjoon *et al.* (2013) and Lazenby *et al.* (2023)

should be considered in future research. To search for existing instruments, we can recommend the Chemistry Instrument Review and Assessment Library (CHIRAL) (Barbera *et al.*, 2022).

In the field of chemistry education research, we see–in line with existing literature – great potential in the use of Item Response Theory approaches, such as the Rasch model, to test *a priori* assumptions of competency levels or learning progressions (Boone *et al.*, 2014). This procedure allows us to identify deviations from the model assumptions, which can then be investigated in more detail, for instance as part of a qualitative study. In the absence of any *a priori* assumptions, it is possible to conduct an explorative Rasch analysis (see for example Taskin *et al.*, 2015). Furthermore, Rasch analysis techniques can and should be used to determine and improve the quality of assessment instruments such as for concept inventories (see Barbera, 2013). Especially since the Rasch-based assessment evaluation, like the learning gain, is direct, comprehensible and interpretable (Pentecost and Barbera, 2013). In case of concept inventory evaluation, we see the potential to re-analyze existing test developments *via* rapid scaling to see if there are any thresholds in the response that indicate levels or fundamental advances in understanding. The idea of threshold concepts (*e.g.*, Meyer, 2016) suggests that there may be concepts that allow easy items to be answered correctly (which have no reference to threshold concepts) and then difficult items are only answered once a threshold concept (*e.g.*, the chemical equilibrium in the context of understanding chemical reaction) has been mastered. It may be possible to prove the existence of specific threshold concepts through *post hoc* Rasch analysis.

Conclusion

Based on a competency level approach, the StoiCoLe model addresses the assessment of algorithmic competencies in stoichiometry. The corresponding test shows good psychometric properties and allows to locate learners on the competency levels. The model assumptions were partially supported by the Rasch analysis and the processing time of the items, but require further development. The results indicate that the majority of students are positioned at the 1st competency level, exhibiting relatively low algorithmic competencies. While they can apply most of the individual sub-competencies successfully, selecting and organising these competencies remains a significant challenge. Only a small proportion of students achieve the higher competency levels. Future research should aim to support students in selecting and organising steps related to specific sub-competencies.

Ethical statement

All participants were informed about the use of their data and were given the option to withdraw from the study at any time. They provided consent prior to participation. The data collection was conducted anonymously, ensuring that individual identities could not be discerned. There were no further ethics requirements from our institution.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the ESI.† This encompasses the applied stoichiometry test, the dichotomous coded results and the revised stoichiometry test. Access to the stoichiometry test with the solutions facilitates comprehension of the presented results, as well as the ability to conduct and analyse the test independently. The results of the stoichiometry test enable verification of the data presented. The revised stoichiometry test illustrates necessary adjustments to some tasks.

Acknowledgements

Notes and references

Agung S. and Schwartz M. S., (2007), Students' Understanding of Conservation of Matter, Stoichiometry and Balancing Equations in Indonesia, *Int. J. Sci. Educ.*, **29**(13), 1679–1702, DOI: **10.1080/09500690601089927**.

Alfieri L., Nokes-Malach T. J. and Schunn C. D., (2013), Learning Through Case Comparisons: A Meta-Analytic Review, *Educ. Psychol.*, **48**(2), 87–113, DOI: **10.1080/00461520.2013.775712**.

Arasasingham R. D., Taagepera M., Potter F. and Lonjers S., (2004), Using Knowledge Space Theory To Assess Student Understanding of Stoichiometry, *J. Chem. Educ.*, **81**(10), 1517–1523, DOI: **10.1021/ED081P1517**.

Arjoon J. A., Xu X. and Lewis J. E., (2013), Understanding the state of the art for measurement in chemistry education research: examining the psychometric evidence, *J. Chem. Educ.*, **90**(5), 536–545, DOI: **10.1021/ED3002013**.

Arlotto R., (1974), Three Simple Stoichiometric Experiments for Introductory High School Chemistry, *Sci. Teacher*, **41**(5), 51–53.

Ault A., (2001), How to Say How Much: Amounts and Stoichiometry, *J. Chem. Educ.*, **78**(10), 1347–1349, DOI: **10.1021/ed078p1347**.

Barbera J., (2013), A psychometric analysis of the chemical concepts inventory, *J. Chem. Educ.*, **90**(5), 546–553, DOI: **10.1021/ED3004353/SUPPL_FILE/ED3004353_SI_002.DOCX**.

Barbera J., Harshman J. and Komperda R., (2022), The Chemistry Instrument Review and Assessment Library (CHIRAL). **https://chiral.chemedx.org**.

Berger S. A., (1985), Using a conversion matrix to simplify stoichiometric calculations from balanced equations, *J. Chem. Educ.*, **62**(5), 396–397, DOI: **10.1021/ed062p396.2**.

Blömeke S., Gustafsson J. E. and Shavelson R. J., (2015), Beyond Dichotomies, *Z. Psychol.*, **223**(1), 3–13, DOI: **10.1027/2151-2604/A000194**.

Bodner G. M., (1987), The role of algorithms in teaching problem solving, *J. Chem. Educ.*, **64**(6), 513–514, DOI: **10.1021/ED064P513**.

Bodner G. M. and Herron J. D., (2002), Problem-Solving in Chemistry, in *Chemical Education: Towards Research-based Practice*, Gilbert J. K., De Jong O., Justi R., Treagust D. F. and Van Driel J. H. (ed.), Springer, pp. 235–266, DOI: **10.1007/0-306-47977-X_11**.

Bond T. G., Yan Z. and Heene M., (2020), Applying the rasch model: fundamental measurement in the human sciences, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 1–348, DOI: **10.4324/9780429030499/ APPLYING-RASCH-MODEL-TREVOR-BOND-ZI-YAN-MORITZ-HEENE/ACCESSIBILITY-INFORMATION**.

Boone W. J., Yale M. S. and Staver J. R., (2014), *Rasch analysis in the human sciences*, Springer, DOI: **10.1007/978-94-007-6857-4**.

BouJaoude S. and Barakat H., (2000), Secondary school students' difficulties with stoichiometry, *School Sci. Rev.*, **81**(296), 91–98.

BouJaoude S. and Barakat H., (2003), Students' Problem Solving Strategies in Stoichiometry and their Relationships to Conceptual Understanding and Learning Approaches, *Electron. J. Sci. Educ.*, **7**(3), 1–42.

BouJaoude S., Salloum S. and Abd-El-Khalick F., (2004), Relationships between selective cognitive variables and students' ability to solve chemistry problems, *Int. J. Sci. Educ.*, **26**(1), 63–84, DOI: **10.1080/0950069032000070315**.

Brown T. L., LeMay H. E., Bursten B. E., Murphy C. J., Woodward P. M. and Stoltzfus M., (2021), *Chemistry: The Central Science in SI Units*, Global Edition, 15th edn, Pearson.

Cain L., (1986), S'Mores— A Demonstration of Stoichiometric Relationships, *J. Chem. Educ.*, **63**(12), 1048–1049, DOI: **10.1021/ed063p1048.2**.

Cameron D. L., (1985), A Pictorial Framework to Aid Conceptualization of Reaction Stoichiometry, *J. Chem. Educ.*, **62**(6), 510–511, DOI: **10.1021/ed062p510**.

Carlson E., (2022), Improving student success in stoichiometry via the scale factor method, *J. Chem. Educ.*, **99**(5), 1863–1867, DOI: **10.1021/acs.jchemed.1c01155**.

Charnock N. L., (2016), Teaching methods for balancing chemical equations: an inspection versus an algebraic approach, *Am. J. Educ. Res.*, **4**(7), 507–511, DOI: **10.12691/education-4-7-2**.

Chibuye B. and Mupela K., (2019), Effect of using Algebraic Method on Secondary School Students' Performance in Balancing Chemical Equations in Chemistry, *J. Educ. Pract.*, **10**(23), 55–62, DOI: **10.7176/JEP/10-23-08**.

Commons M. L., Trudeau E. J., Stein S. A., Richards F. A. and Krause S. R., (1998), Hierarchical Complexity of Tasks Shows the Existence of Developmental Stages, *Dev. Rev.*, **18**(3), 237–278, DOI: **10.1006/DREV.1998.0467**.

Costaguta R., (2015), Algorithms and machine learning techniques in collaborative group formation, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, pp. 249–258, DOI: **10.1007/978-3-319-27101-9_18**.

Cotes S. and Cotuá J., (2014), Using Audience Response Systems during Interactive Lectures To Promote Active Learning and Conceptual Understanding of Stoichiometry, *J. Chem. Educ.*, **91**(5), 673–677, DOI: **10.1021/ed400111m**.

Cracolice M. S., Deming J. C. and Ehlert B., (2008), Concept Learning versus Problem Solving: A Cognitive Difference, *J. Chem. Educ.*, **85**(6), 873–878, DOI: **10.1021/ed085p873**.

Cronbach L. J. and Meehl P. E., (1955), Construct validity in psychological tests, *Psychol. Bull.*, **52**(4), 281–302, DOI: **10.1037/H0040957**.

Dahsah C. and Coll R. K., (2008), Thai Grade 10 and 11 student's understanding of stoichiometry and related concepts, *Int. J. Sci. Math. Educ.*, **6**, 573–600, DOI: **10.1007/s10763-007-9072-0**.

Deonovic B., Yudelson M., Bolsinova M., Attali M. and Maris G., (2018), Learning meets assessment: on the relation between item response theory and Bayesian knowledge tracing, *Behaviormetrika*, **45**(2), 457–474, DOI: **10.1007/s41237-018-0070-z**.

Dierks W., (1981), Stoichiometric Calculations: Known Problems and Proposed Solutions at a Chemistry-Mathematics Interface, *Stud. Sci. Educ.*, **8**(1), 93–105, DOI: **10.1080/03057268108559888**.

Dori Y. J. and Hameiri M., (2003), Multidimensional analysis system for quantitative chemistry problems: symbol, macro, micro, and process aspects, *J. Res. Sci. Teach.*, **40**(3), 278–302, DOI: **10.1002/TEA.10077**.

Figueira A. R., Coch J. and Zeplca M., (1988), Teaching Stoichiometry, *J. Chem. Educ.*, **65**(12), 1060–1061, DOI: **10.1021/ed065p1060**.

Fisher W. P., (1992), Reliability statistics, *Rash Measurement Trans.*, **6**(3), 238–238.

Frank D. V., Baker C. A. and Herron J. D., (1987), Should Students Always Use Algorithms To Solve Problems? *J. Chem. Educ.*, **64**(6), 514–515, DOI: **10.1021/ed064p514**.

Frey A., Hartig J. and Rupp A. A., (2009), An NCME instructional module on booklet designs in large-scale assessments of student achievement: theory and practice, *Educ. Measurement: Issues Pract.*, **28**(3), 39–53, DOI: **10.1111/J.1745-3992.2009.00154.X**.

Garst J. F., (1974), The Extent of Reaction as a Unifying Basis for Stoichiometry in Elementary Chemistry, *J. Chem. Educ.*, **51**(3), 194–196, DOI: **10.1021/ed051p194**.

Gayeta N. E., (2017), Flipped Classroom as an Alternative Strategy for Teaching Stoichiometry, *Asia Pacific J. Multidisciplinary Res.*, **5**(4), 83–89.

Gobert J. D., Sao Pedro M. A., Baker R. S. J. D., Toto E. and Montalvo O., (2012), Leveraging Educational Data Mining for Real-Time Performance Assessment of Scientific Inquiry Skills within Microworlds, *J. Educ. Data Mining*, **4**(1), 104–143, DOI: **10.5281/zenodo.3554645**.

Gobert J. D., Sao Pedro M., Raziuddin J. and Baker R. S., (2013), From Log Files to Assessment Metrics: Measuring

Students' Science Inquiry Skills Using Educational Data Mining, *J. Learn. Sci.*, **22**(4), 521–563, DOI: **10.1080/10508406.2013.837391**.

Gulacar O. and Fynewever H., (2010), A research methodology for studying what makes some problems difficult to solve, *Int. J. Sci. Educ.*, **32**(16), 2167–2184, DOI: **10.1080/09500690903358335**.

Gulacar O., Mann H. K., Mann S. S. and Vernoy B. J., (2022), The Influence of Problem Construction on Undergraduates Success with Stoichiometry Problems, *Educ. Sci.*, **12**, 867–887, DOI: **10.3390/EDUCSCI12120867**.

Gulacar O., Overton T. L. and Bowman C. R., (2013a), A Closer Look at the Relationships between College Students' Cognitive Abilities and Problem Solving in Stoichiometry, *Int. J. Phys. Chem. Educ.*, **5**(2), 144–163, DOI: **10.51724/ijpce.v5i2.80**.

Gulacar O., Overton T. L., Bowman C. R. and Fynewever H., (2013b), A novel code system for revealing sources of students' difficulties with stoichiometry, *Chem. Educ. Res. Pract.*, **14**(4), 507–515, DOI: **10.1039/C3RP00029J**.

Guo C., (1997), A New and General Method for Balancing Chemical Equations by Inspections, *J. Chem. Educ.*, **74**(11), 1365–1366, DOI: **10.1021/ED074P1365**.

Haim L., (2003), Learning Stoichiometry with Hamburger Sandwiches, *J. Chem. Educ.*, **80**(9), 1021–1022, DOI: **10.1021/ed080p1021**.

Hand B., Yang O. E. M. and Bruxvoort C., (2007), Using writing-to-learn science strategies to improve year 11 students' understandings of stoichiometry, *Int. J. Sci. Math. Educ.*, **5**(1), 125–143, DOI: **10.1007/S10763-005-9028-1/METRICS**.

Hartig J., (2007), Skalierung und Definition von Kompetenzniveaus [Scaling and definition of competence levels], in Klieme E. and Beck B. (ed.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie Deutsch Englisch Schülerleistungen International*, Beltz, pp. 83–99, DOI: **10.25656/01:3143**.

Hartig J., Frey A. and Jude N., (2020), Validität von Testwertinterpretationen [Validity of test value interpretations], in Moosbrugger H. and Kelava A. (ed.), *Testtheorie und Fragebogenkonstruktion*, Springer, pp. 529–545, DOI: **10.1007/978-3-662-61532-4_21**.

Hartig J. and Klieme E., (2006), Kompetenz und Kompetenzdiagnostik [Competence and competence diagnostics], in Schweizer K. (ed.), *Leistung und Leistungsdiagnostik*, Springer, pp. 127–143, DOI: **10.1007/3-540-33020-8_9**.

Heeg J., Hundertmark S. and Schanze S., (2020), The interplay between individual reflection and collaborative learning-seven essential features for designing fruitful classroom practices that develop students' individual conceptions, *Chem. Educ. Res. Pract.*, **21**(3), 765–788, DOI: **10.1039/c9rp00175a**.

Herndon W. C., (1997), On Balancing Chemical Equations: Past and Present, *J. Chem. Educ.*, **74**(11), 1359–1362, DOI: **10.1021/ED074P1359**.

Huddle P. A. and Pillay A. E., (1996), An In-Depth Study of Misconceptions in Stoichiometry and Chemical Equilibrium at a South African University, *J. Res. Sci. Teaching*, **33**(1),

65–77, DOI: **10.1002/(SICI)1098-2736(199601)33:1<65::AID-TEA4>3.0.CO;2-N**.

Ifenthaler D., Gibson D. and Dobozy E., (2018), Informing learning design through analytics: applying network graph analysis, *Australasian J. Educ. Technol.*, **34**(2), 117–132, DOI: **10.14742/AJET.3767**.

Kane M. T., (1992), An argument-based approach to validity. *Psychol Bull*, **112**(3), 527–535, DOI: **10.1037/0033-2909.112.3.527**.

Karp S., (1988), Stoichiometric Calculations Using Moles, *J. Chem. Educ.*, **65**(12), 1115, DOI: **10.1021/ED065P1115.1**.

Kashmar R. J., (1997), The use of cut-out molecular models on the overhead projector to illustrate stoichiometry and limiting reactants, *J. Chem. Educ.*, **74**(7), 791–792, DOI: **10.1021/ED074P791**.

Kauertz A., Fischer H. E., Mayer J. and Sumfleth E., (2010), Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I [Modeling competence according to standards for science education in secondary schools], *Z. Didaktik Naturwissen.*, **16**, 135–153.

Kauffman G. B., (1976), A schematic summary of general chemistry stoichiometry, *J. Chem. Educ.*, **53**(8), 509, DOI: **10.1021/ED053P509**.

Klieme E., Avenarius H., Blum W., Döbrich P., Gruber H., Prenzel M., Reiss K., Riquarts K., Rost J., Tenorth H. E. and Vollmer H. J., (2003), Zur Entwicklung nationaler Bildungsstandards. Eine Expertise [On the development of national educational standards. An expertise], *Bundesministerium für Bildung und Forschung (BMBF),* 2nd edn, DOI: **10.25656/01:20901**.

Koch H., (1995), Simplifying Stoichiometry: students react with confidence to a new approach to quantitative chemistry problems, *Sci. Teacher*, **62**(8), 36–39.

Koeppen K., Hartig J., Klieme E. and Leutner D., (2008), Current issues in competence modeling and assessment, *J. Psychol.*, **216**(2), 61–73, DOI: **10.1027/0044-3409.216.2.61**.

Kolb D., (1979), More on balancing redox equations, *J. Chem. Educ.*, **56**(3), 181–184, DOI: **10.1021/ED056P181**.

Kolb D., (1981), Balancing complex redox equations by inspection, *J. Chem. Educ.*, **58**(8), 642–645, DOI: **10.1021/ED058P642**.

Krell M., Vorholzer A. and Nehring A., (2022), Scientific Reasoning in Science Education: From Global Measures to Fine-Grained Descriptions of Students' Competencies, *Educ. Sci.*, **12**(2), 97, DOI: **10.3390/EDUCSCI12020097**.

Krieger C. R., (1997), Stoogiometry: A Cognitive Approach to Teaching Stoichiometry, *J. Chem. Educ.*, **74**(3), 306–309, DOI: **10.1021/ed074p306**.

Lazenby K., Tenney K., Marcroft T. A. and Komperda R., (2023), Practices in instrument use and development in chemistry education research and practice 2010–2021, *Chem. Educ. Res. Pract.*, **24**, 882–895, DOI: **10.1039/D2RP00275B**.

Lazonby J. N., Morris J. E. and Waddington D. J., (1982), The muddlesome mole, *Educ. Chem.*, **19**(4), 109–111.

Limesurvey GmbH, (2003), LimeSurvey: An open source survey tool.

Ludlow L. and O'Leary M., (2000), What to do about missing data? *Rasch Measurement Trans.*, **14**(2), 751–751.

Lythcott J., (1990), Problem Solving and Requisite Knowledge of Chemistry, *J. Chem. Educ.*, **67**(3), 248–252, DOI: **10.1021/ed067p248**.

Martínez V. A. and Ibanez J. G., (2020), All Roads Lead to Rome: Triple Stoichiometry with a Lithium Battery, *J. Chem. Educ.*, **97**(11), 4103–4107, DOI: **10.1021/ACS.JCHEMED.0C00876/ASSET/IMAGES/MEDIUM/ED0C00876_M003.GIF**.

Mason D. and Crawley F. E., (1994), Differences between Algorithmic and Conceptual Problem Solving by Nonscience Majors in Introductory Chemistry [Paper presentation], National Association for Research in Science Teaching 67th Annual Meeting, Anaheim, California, United States.

Mason D. S., Shell D. F. and Crawley F. E., (1997), Differences in Problem Solving by Nonscience Majors in Introductory Chemistry on Paired Algorithmic-Conceptual Problems, *J. Res. Sci. Teach.*, **34**(9), 905–923, DOI: **10.1002/(SICI)1098-2736(199711)34:9<905::AID-TEA5>3.0.CO;2-Y**.

Meyer J. H. F., (2016), Threshold concepts and pedagogic representation, *Educ. Training*, **58**(5), 463–475, DOI: **10.1108/ET-04-2016-0066**.

Molnar J. and Molnar-Hamvas L., (2011), LEGO-Method – New Strategy for Chemistry Calculation, *US-China Educ. Rev. B*, **7**, 891–908.

Mousavi A., (2019), Stoichiometry of equations through the Inverse de Donder relation, *Chem. Teacher Int.*, **1**(1), 1–6, DOI: **10.1515/cti-2018-0006**.

Nakhleh M. B., (1993), Are Our Students Conceptual Thinkers or Algorithmic Problem Solvers? Identifying Conceptual Students in General Chemistry, *J. Chem. Educ.*, **70**(1), 52–55, DOI: **10.1021/ED070P52**.

Neumann K., (2020), Die Bedeutung instruktionaler Kohärenz für eine systematische Kompetenzentwicklung [The importance of instructional coherence for systematic competence development], *Unterrichtswissenschaft*, **48**, 1–10, DOI: **10.1007/S42010-020-00068-6**.

Niaz M., (1988), The information-processing demand of chemistry problems and its relation to pascual-leone's functional *m*-capacity, *Int. J. Sci. Educ.*, **10**(2), 231–238, DOI: **10.1080/0950069880100211**.

Niaz M., (1989), The relationship between M-demand, algorithms, and problem solving: a neo-Piagetian analysis, *J. Chem. Educ.*, **66**(5), 422–424, DOI: **10.1021/ED066P422**.

Niaz M., (1995a), Relationship between student performance on conceptual and computational problems of chemical equilibrium, *Int. J. Sci. Educ.*, **17**(3), 343–355, DOI: **10.1080/0950069950170306**.

Niaz M., (1995b), Relationship between student performance on conceptual and computational problems of chemical equilibrium, *Int. J. Sci. Educ.*, **17**(3), 343–355, DOI: **10.1080/0950069950170306**.

Nufus H., Susilawati S. and Linda R., (2020), Implementation of E-Module Stoiciometry Based on Kvisoft Flipbook Maker for Increasing Understanding Study Learning Concepts of Class X Senior High School, *J. Educ. Sci.*, **4**(2), 261–272, DOI: **10.31258/JES.4.2.P.261-272**.

Nurrenbern S. C. and Pickering M., (1987), Concept Learning versus Problem Solving Is There a Difference? *J. Chem. Educ.*, **64**(6), 508–510, DOI: **10.1021/ed064p508**.

Okanlawon A. E., (2008), The Modified GRASS Model: An Alternate Path to Solve Complex Stoichiometric Problems, *J. Turk. Sci. Educ.*, **5**(2), 11–25.

Olmsted J., (1999), Amounts Tables as a Diagnostic Tool for Flawed Stoichiometric Reasoning, *J. Chem. Educ.*, **76**(1), 52–54, DOI: **10.1021/ED076P52**.

Pentecost T. C. and Barbera J., (2013), Measuring learning gains in chemical education: a comparison of two methods, *J. Chem. Educ.*, **90**(7), 839–845, DOI: **10.1021/ED400018V**.

Pickering M., (1990), Further Studies on Concept Learning versus Problem Solving, *J. Chem. Educ.*, **67**(3), 254–255, DOI: **10.1021/ed067p254**.

Poole R. L., (1986), Teaching Stoichiometry: A Two Cycle Approach, *J. Chem. Educ.*, **63**(1), 57–58, DOI: **10.1021/ed066p57**.

Posit, (2025), *RStudio: Integrated development for R (Version 2024.12.0.467), [Computer software], Posit Software, PBC*, **https://posit.co**.

Ramesh V. B., Selvam A. A. A., Kulkarni S., Dattatreya Manganahalli A. and Bettadapur K. R., (2020), Designing and Using an Atomic Model Kit with H, C, N, and O Model Atoms Having a Mass Ratio of 1 : 12 : 14 : 16 to Teach the Concept of Mole and Associated Stoichiometric Relationships, *J. Chem. Educ.*, **97**(4), 986–991, DOI: **10.1021/ACS.JCHEMED.9B00665**.

Rasmawan R., (2022), Stoichiometry E-Book Based on Creative Problem Solving (CPS) to Solve Conceptual Problems, *J. Pendidikan Kimia Indonesia*, **6**(1), 70–78, DOI: **10.23887/JPK.V6I1.43441**.

R Core Team, (2025), *R: A language and environment for statistical computing (Version 4.4.2), [Computer software], R Foundation for Statistical Computing*, **https://www.R-project.org**.

Robitzsch A., Kiefer T. and Wu M., (2024), *TAM: Test Analysis Modules (Version 4.2-21) [Computer software]. CRAN*, **https://cran.r-project.org/packages=TAM**.

Rosa V., States N. E., Corrales A., Nguyen Y. and Atkinson M. B., (2022), Relevance and equity: should stoichiometry be the foundation of introductory chemistry courses? *Chem. Educ. Res. Pract.*, **23**(3), 1–24, DOI: **10.1039/d1rp00333j**.

Rosenberg D., Jansen W. and Busker M., (2016), Strukturierte Lösungshilfen für Aufgaben zur Stöchiometrie [Structured solution aids for stoichiometry tasks], *Prax. Naturwiss., Chem. Sch.*, **65**(4), 20–24, DOI: **10.25656/01:17390**.

Saadatkhah N., Carillo Garcia A., Ackermann S., Leclerc P., Latifi M. and Samih S., et al., (2020), Experimental methods in chemical engineering: thermogravimetric analysis—TGA, *Can, J. Chem. Eng.*, **98**(1), 34–43, DOI: **10.1002/CJCE.23673**.

Sawrey B. A., (1990), Concept Learning versus Problem Solving: Revisited, *J. Chem. Educ.*, **67**(3), 253–254, DOI: **10.1021/ed067p253**.

Schecker H. and Parchmann I., (2006), Modellierung naturwissenschaftlicher Kompetenz [Modelling scientific competence], *Z. Didaktik Naturwiss.*, **12**(1), 45–66.

Schmidt H. J., (1990), Secondary school students' strategies in stoichiometry, *Int. J. Sci. Educ.*, **12**(4), 457–471, DOI: **10.1080/0950069900120411**.

Schmidt H. J., (1997), An alternate path to stoichiometric problem solving, *Res. Sci. Educ.*, **27**(2), 237–249, DOI: **10.1007/BF02461319**.

Schmidt H. J. and Jignéus C., (2003), Students' strategies in solving algorithmic stoichiometry problems, *Chem. Educ. Res. Pract.*, **4**(3), 305–317, DOI: **10.1039/B3RP90018E**.

Scott F. J., (2012), Is mathematics to blame? An investigation into high school students' difficulty in performing calculations in chemistry, *Chem. Educ. Res. Pract.*, **13**(3), 330–336, DOI: **10.1039/c2rp00001f**.

Segedinac M. T., Horvat S., Rodić D. D., Rončević T. N. and Savić G., (2018), Using knowledge space theory to compare expected and real knowledge spaces in learning stoichiometry, *Chem. Educ. Res. Pract.*, **19**(3), 670–680, DOI: **10.1039/c8rp00052b**.

Shadreck M. and Enunuwe O. C., (2018), Recurrent Difficulties: stoichiometry problem-solving, *Afr. J. Educ. Stud. Math. Sci.*, **14**, 25–31, DOI: **10.4314/AJESMS.V14I0**.

Smith K. J. and Metz P. A., (1996), Evaluating Student Understanding of Solution Chemistry through Microscopic Representations, *J. Chem. Educ.*, **73**(3), 233–235, DOI: **10.1021/ED073P233**.

Smith K. C., Nakhleh M. B. and Lowery Bretz S., (2010), An expanded framework for analyzing general chemistry exams, *Chem. Educ. Res. Pract.*, **11**(3), 147, DOI: **10.1039/C005463C**.

Stachowiak H., (1973), *Allgemeine Modelltheorie [General model theory]*, Springer.

Stamovlasis D., Tsaparlis G., Kamilatos C., Papaoikonomou D. and Zarotiadou E., (2004), Conceptual Understanding versus Algorithmic Problem Solving: A Principal Component Analysis of a National Examination, *Chem. Educ.*, **9**(6), 398–405.

Stamovlasis D., Tsaparlis G., Kamilatos C., Papaoikonomou D. and Zarotiadou E., (2005), Conceptual understanding versus algorithmic problem solving: further evidence from a national chemistry examination, *Chem. Educ. Res. Pract.*, **6**(2), 104–118, DOI: **10.1039/B2RP90001G**.

Steiner R. P., (1986), Teaching stoichiometry, *J. Chem. Educ.*, **63**(12), 1048, DOI: **10.1021/ed063p1048.1**.

Tang H., Kirk J. and Pienta N. J., (2014), Investigating the effect of complexity factors in stoichiometry problems using logistic regression and eye tracking, *J. Chem. Educ.*, **91**(7), 969–975, DOI: **10.1021/ED4004113/ASSET/IMAGES/LARGE/ED-2013-004113_0001.JPEG**.

Taskin V. and Bernholt S., (2014), Students' Understanding of Chemical Formulae: a review of empirical research, *Int. J. Sci. Educ.*, **36**(1), 157–185, DOI: **10.1080/09500693.2012.744492**.

Taskin V., Bernholt S. and Parchmann I., (2015), An inventory for measuring student teachers' knowledge of chemical representations: design, validation, and psychometric analysis, *Chem. Educ. Res. Pract.*, **16**(3), 460–477, DOI: **10.1039/C4RP00214H**.

Teig N., (2024), Uncovering Student Strategies for Solving Scientific Inquiry Tasks: Insights from Student Process Data in PISA, *Res. Sci. Educ.*, **54**(2), 205–224, DOI: **10.1007/S11165-023-10134-5/FIGURES/2**.

Torres Irribarra D. and Freund R., (2024), *Wright Map: IRT item-person map with ConQuest integration. (Version 1.4). [Computer software]*. CRAN, **https://CRAN.R-project.org/package=WrightMap**.

Tóth Z., (1997), Balancing Chemical Equations by Inspection, *J. Chem. Educ.*, **74**(11), 1363–1364, DOI: **10.1021/ED074P1363**.

Tykodi R. J., (1987), Reaction Stoichiometry and Suitable "Coordinate Systems", *J. Chem. Educ.*, **64**(11), 958–960, DOI: **10.1021/ed064p958**.

Tyndall J. R., (1975), A Logic Diagram for Teaching Stoichiometry, *J. Chem. Educ.*, **52**(8), 492–492, DOI: **10.1021/ED052P492**.

Umland J. B., (1984), A Recipe for Teaching Stoichiometry, *J. Chem. Educ.*, **61**(12), 1036–1037, DOI: **10.1021/ed061p1036**.

Vorholzer A. and von Aufschnaiter C., (2020), Dimensionen und Ausprägungen fachinhaltlicher Kompetenz in den Naturwissenschaften – ein Systematisierungsversuch [Dimensions and levels of scientific competence – An attempt at systematization], *Z. Didaktik der Naturwissenschaften*, **26**(1), 1–18, DOI: **10.1007/S40573-020-00107-W**.

Wagner E. P., (2001), A Study Comparing the Efficacy of a Mole Ratio Flow Chart to Dimensional Analysis for Teaching Reaction Stoichiometry, *Sch. Sci. Math.*, **101**(1), 10–22, DOI: **10.1111/J.1949-8594.2001.TB18185.X**.

Walpuski M. and Ropohl M., (2014), Statistische Verfahren für die Analyse des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit [Statistical methods for analysing the influence of task characteristics on difficulty], in Krüger D., Parchmann I. and Schecker H. (ed.), *Methoden in der naturwissenschaftsdidaktischen Forschung*, Springer, pp. 385–398, DOI: **10.1007/978-3-642-37827-0_30**.

Weinert F. E., (2001), Vergleichende Leistungsmessung in Schulen: Eine Umstrittene Selbstverständlichkeit [Comparative performance assessment in schools: a controversial matter of course], in *Leistungsmessungen in Schulen*, Weinert F. E. (ed.), Beltz, pp. 17–32.

Weinstein C. E. and Mayer R. E., (1986), The teaching of learning strategies, in Wittrock M. C. (ed.), *Handbook of Research on Teaching*, Macmillan, pp. 315–327.

Wilson M., (2023), *Constructing measures: an item response modeling approach*, 2nd edn, Taylor & Francis, DOI: **10.4324/9781003286929**.

Witzel J. E., (2002), Lego Stoichiometry, *J. Chem. Educ.*, **79**(3), 352A–352B, DOI: **10.1021/ed079p352**.

Wright B. D. and Linacre J. M., (1994), Reasonable mean-square fit values, *Rasch Measurement Trans.*, **8**(3), 370–370.

Wright B. D. and Masters G. N., (1982), *Rating Scale Analysis Rasch Measurement*, Mesa Press.

Yalman R. G., (1959), Writing oxidation-reduction equations: a review of textbook materials, *J. Chem. Educ.*, **36**(5), 215–218, DOI: **10.1021/ED036P215**.

Yde P. B., (1989), Abolish the half reaction method, *Int. J. Math. Educ. Sci. Technol.*, **20**(4), 533–542, DOI: **10.1080/0020739890200406**.