Received: 15 March 2022 | Accepted: 29 October 2022

DOI: 10.1111/bjet.13288

REVIEW ARTICLE

British Journal of Educational Technology BERA

Formative assessment strategies for students' conceptions—The potential of learning analytics

Judith Stanja[1] | Wolfgang Gritz[2] | Johannes Krugel[1] | Anett Hoppe[2] | Sarah Dannemann[3]

[1]Didactics of Electrical Engineering and Computer Science, Leibniz University Hannover, Hannover, Germany

[2]TIB Leibniz Information Centre for Science and Technology & L3S Research Centre, Leibniz University Hannover, Hannover, Germany

[3]Biology Education, Martin-Luther-University, Halle (Saale), Germany

Correspondence

Judith Stanja, Didactics of Electrical Engineering and Computer Science, Leibniz University Hannover, Appelstraße 9A, Hannover 30167, Germany.
Email: stanja@dei.uni-hannover.de

Funding information

Niedersächsisches Ministerium für Wissenschaft und Kultur

Abstract

Formative assessment is considered to be helpful in students' learning support and teaching design. Following Aufschnaiter's and Alonzo's framework, formative assessment practices of teachers can be subdivided into three practices: eliciting evidence, interpreting evidence and responding. Since students' conceptions are judged to be important for meaningful learning across disciplines, teachers are required to assess their students' conceptions. The focus of this article lies on the discussion of learning analytics for supporting the assessment of students' conceptions in class. The existing and potential contributions of learning analytics are discussed related to the named formative assessment framework in order to enhance the teachers' options to consider individual students' conceptions. We refer to findings from biology and computer science education on existing assessment tools and identify limitations and potentials with respect to the assessment of students' conceptions.

KEYWORDS

biology education, computer science education, formative assessment, learning analytics, students' conceptions/explanations, synthesis paper, teacher support

INTRODUCTION

Learning processes may aim at the acquirement of skills, knowledge, practices or the individuals' understanding. One responsibility of teachers is to support students' learning processes

Practitioner notes

What is already known about this topic

- Students' conceptions are considered to be important for learning processes, but interpreting evidence for learning with respect to students' conceptions is challenging for teachers.
- Assessment tools have been developed in different educational domains for teaching practice.
- Techniques from artificial intelligence and machine learning have been applied for automated assessment of specific aspects of learning.

What does the paper add

- Findings on existing assessment tools from two educational domains are summarised and limitations with respect to assessment of students' conceptions are identified.
- Relevent data that needs to be analysed for insights into students' conceptions is identified from an educational perspective.
- Potential contributions of learning analytics to support the challenging task to elicit students' conceptions are discussed.

Implications for practice and/or policy

- Learning analytics can enhance the eliciting of students' conceptions.
- Based on the analysis of existing works, further exploration and developments of analysis techniques for unstructured text and multimodal data are desirable to support the eliciting of students' conceptions.

(Gropengießer, 2006). Several fields of domain-specific educational research have found that students' conceptions are important in learning processes and need to be considered in teaching if the students' understanding is to be addressed (Amin, 2015). Formative assessment (FA) is seen as a possibility to support learning processes by considering individual students' conceptions (von Aufschnaiter & Alonzo, 2018). Nevertheless, teacher profession research has found that the FA of students' conceptions is a challenging task even for experienced teachers (von Aufschnaiter & Alonzo, 2018). This article focuses on the discussion of learning analytics (LA), for example, using machine learning (ML) methods, as potential support tools for the assessment of students' conceptions in class. In this article, we contribute to the work on *learning analytics for assessment* (Gašević et al., 2022) by

- identifying challenges of FA regarding students' conceptions,
- carving out some relevant data that needs to be analysed and presenting ideas on how LA can inform and support FA of students' conceptions.

Similar to the FaSMEd (Formative Assessment in Science and Mathematics Education) model that has been built to characterise and analyse technology enhanced FA (Cusi et al., 2019), we also combine the dimensions, *FA strategies* and *functionalities of technology*. As for the *agents* dimension, we focus on support for the *teacher*. In contrast to the FA strategies' dimension of the FaSMEd model (Cusi et al., 2019), we base our analysis on a model of *formative assessment as practices*, which has been widely used in science education (Alonzo, 2018; von Aufschnaiter & Alonzo, 2018) (see Section 2). The model by Alonzo (2018) and von Aufschnaiter and Alonzo (2018) is suitable to characterise the involved practices and to discuss possible contributions of LA to the further development of technology-enhanced FA.

In Section 3, we look at a selection of assessment tools used in biology education and computer science education. The examples from the two educational disciplines illustrate the current state of FA with regard to different targets of assessments and highlight domain-specific differences. Potential contributions of LA to the identified challenges and desirable further developments from an educational perspective are discussed in Section 4, together with research questions that could guide future work. The article concludes in Section 5.

THEORETICAL BACKGROUND

In this section, we expose the characteristics of students' conceptions and their relevance for understanding processes. Based on this, a model for FA is introduced that can be used to characterise the involved practices for teachers and, thus, can serve as a promising reference point for LA. Finally, we provide information on the application of learning analytics to the educational field.

The importance of students' conceptions in learning processes

In education, the need to consider students' conceptions in teaching is widely accepted across disciplines (Amin et al., 2011; Bascandziev et al., 2018; diSessa, 2014; Gropengießer, 2006; Groß et al., 2019; Hammann & Asshoff, 2014; Kattmann, 2015; Krüger et al., 2018; Vosniadou, 2013). Multiple studies and models since the 1970s reflect the importance that is attributed to students' conceptions (Amin, 2015; Potvin et al., 2020)—not only as a prerequisite but also as a helpful and necessary resource for understanding. Different theoretical approaches have been used in research (Amin, 2015; Potvin et al., 2020). The used theory frames how learning is conceptualised and modelled and therefore affects the evaluation of the appropriateness of assessment techniques. Furthermore, the design of supporting materials and tools for FA purposes is determined by the conceptualisation of learning. The conceptualisation of learning has implications for learner modelling, what shall be elicited and how it can be interpreted (see Section 2.2).

The theory of experience-based understanding assumes that humans actively construct a basic understanding of their environment and reality based on their experiences (Krüger, 2007). This experience-based understanding further determines learning (Krüger, 2007, p. 83)—even in areas that do not allow experiences. This theory is based on the cognitive linguistic theory by Lakoff and Johnson (Lakoff & Johnson, 1980). According to the theory of experience-based understanding, students' conceptions are conceptualised as embodied cognitions that are constructed based on sensomotorical experiences. In contexts that do not allow experiences, these embodied cognitions are used in a metaphorical mapping process in order to make sense (Krüger, 2007). In many cases, we are not aware of these mapping processes. In everyday life, biological phenomena are often explained using metaphors that are contrary to a biological understanding of the phenomenon (Gropengießer, 2006). This understanding differs from other conceptualisations—for example, with regard to the value of students' conceptions. Some researchers use the terms misconceptions, naive ideas, non-normative or alternative ideas to refer to 'inaccurate explanatory elements' from a normative scientific perspective (Nehm & Ha, 2011). In the frame of the theory of experience-based understanding, these conceptions are termed as "everyday conceptions" and the development of conceptions does not incorporate a judgemental perspective. Everyday conceptions have been useful to the learners in their day-to-day life and may represent potential as well as obstacles for individualised learning (Gropengießer & Marohn, 2018; Kattmann, 2015; Zabel & Gropengiesser, 2011). Therefore, we use the term "everyday conception" to refer to conceptions that differ from biological ones. To support individual learning processes, it is considered as important to diagnose both, the students' everyday and biological conceptions (Gropengießer, 2006; von Aufschnaiter & Alonzo, 2018). For the both of them combined, we use the term "students' conceptions". This means, that learning is not understood as a substitution of 'inaccurate' conceptions but rather an enrichment, modification or differentiation (Baalmann et al., 2004). The structure of students' learning paths with respect to their development of conceptions may differ individually and have found to be non-hierarchical and complex (Zabel & Gropengiesser, 2011). As students' conceptions necessarily need to be considered in the teaching design, formative assessments of students' conceptions are necessary to support learning as understanding.

Formative assessments

Formative assessments—also called assessments for learning—aim at support of learning and teaching (Zhai et al., 2021) by assessing a learner's state and inferring next steps. In this article, we focus on FA for the support of teaching. According to the work of Alonzo (2018) and von Aufschnaiter and Alonzo (2018), FA is seen as a process consisting of the following three practices:

1. eliciting,
2. interpreting and
3. responding.

Eliciting is about the collection of evidence for students' learning using tasks and questions (eg, in classroom discussions or with the use of tasks or instruments). Alonzo argues for interpretable evidence to gain more precise and actionable information than the number of correct answers or a norm-referenced score (Alonzo, 2018). With respect to students' conceptions, findings on scientifically" wrong" but common ideas have to be included (Alonzo, 2018). Alonzo and Aufschnaiter subdivide *interpreting* into two practices: First, *analysing* what students are saying, writing or doing and what this indicates about their thinking; second, *identifying implications* for learning based on the previous analysis, and *responding* then regards feedback to students or adaptation of instruction. Educational research can inform about hindering, productive or necessary conceptions for further learning. The practice of interpreting is particularly challenging for teachers (Alonzo, 2018). Often, characterisations of students' understanding are simplified and dichotomous (right/wrong; "gets it"/"doesn't get it"). Alonzo and Aufschnaiter have pointed out that dichotomous characterisations of students' understanding are problematic since they do not uncover students' learning resources and learning needs and therefore have negative effects for all three practices (limited focus on vocabulary or facts, holistic judgements as "right"/"wrong" instead of nuanced information, difficulties in/no orientation for responding) (von Aufschnaiter & Alonzo, 2018).

There are many challenges that arise when the theoretical demand to consider students' conceptions shall be implemented by teachers in practice. This includes the frequent assessment of students' conceptions. Assuming that teachers usually teach a large number of students, time is a critical factor. Besides time, the variety of topics and the application of corresponding educational models to gain differentiated insights as well as the analysis process itself pose more challenges for teachers. To mitigate the challenges of FA, educational research has suggested to provide support for teachers in form of resources like frameworks or models that could be used to reason about students' conceptions or assessment instruments (Alonzo, 2018). Section 3 will present and discuss current assessment tools in biology and computer science education and identify limitations and options for a further development of data-based support from an educational perspective.

Learning analytics

According to Siemens et al. (2011), 'Learning analytics (LA) is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.' In a similar sense, Tempelaar et al. (2013) describe the goals of LA: 'to apply the outcomes of analysing data gathered by monitoring and measuring the learning process, as feedback to assist directing that same learning process.'

Data

Wise (2019) distinguishes between three different types of data:

- Activity, that is, data describing what the learners are doing (eg, log files).
- Artefact, that is, data created by learners (eg, quiz answers).
- Association, that is, data describing connections between entities (eg, students interacting with other students or resources).

The data generated can be used to derive conclusions about the student or instructor (eg, engagement (Yilmaz & Yilmaz, 2022), sentiment (Kaliisa & Dolonen, 2022), etc.) in order to improve the learning process. A limiting aspect to consider using LA methods when analysing the data is that only a slice of reality is represented (Selwyn, 2019). For example, it is unlikely that notes on paper or conversations in the classroom will be fully captured and processed and ready for automatic analysis. When it comes to ML models, important requirements are the availability and representativity of training data for the target application to ensure generalisability of the models (Nehm, Ha, et al., 2012).

Common analytic methods

Data collected during learning can be used in a variety of ways. Typical classes of methods used in LA are as follows:

- Prediction.
- Structure discovery.
- Temporal analysis.
- Visual analytics.
- Natural language processing (NLP).

For example, predictive models (eg, classification and regression) can be used for early alert systems to inform learners about their current progress (Arnold & Pistilli, 2012) or to forecast student success (Gardner & Brooks, 2018). Structural analysis (eg, clustering or network analysis) on the other hand, can detect isolated and active students (Saqr & Alamro, 2019). Temporal analyses (eg, hidden Markov models) attempt to analyse time-related aspects of students' behaviour patterns, for example, of online assignment submissions (Kokoç et al., 2021). Visual analysis can be used for learners and instructors to convey information more easily, for example, via dashboards (Broos et al., 2020). NLP can be used for the automatic evaluation of texts like essays and provide feedback to instructors or students (Knight et al., 2020).

Learning analytics and formative assessments

There is a variety of analysis techniques for automatic assessments, which include item response theory, diagnostic classification models, hidden Markov models, factor analysis and deep learning based models (Minn, 2022) or combinations of techniques (eg, Liu et al., 2021 or Misiejuk et al., 2021). They have been used for assessments of different kinds of learners' abilities like skills, factual knowledge, and problem solving abilities (Minn, 2022) or perceptions of peer feedback (Misiejuk et al., 2021). Whether and how they could be used for the assessments of students' conceptions still needs further research. ML techniques have been used for the development of assessments to" (1) evaluate complex constructs in science, (2) enhance inferences about student comprehension and (3) advance automation and accuracy of scoring" (Zhai et al., 2021, pp. 193–194; see also Bertolini et al., 2021; Nehm, Ha, et al., 2012).

Jensen et al. (2021) attempt to increase learning and retention with interventions of formative assessment quizzes that students are likely to answer correctly. The quizzes are asked immediately after watching a video on its content. They consist of three items and the tasks are to fill in blanks, describe graphs or explain why a statement is true or false. Students get instant feedback on how many answers were correct after completing the quizzes. Also with students in mathematics, Zheng et al. (2019) used assessments embedded in an adaptive system as a type of formative assessment for mathematics students in grades 6–8. They used a platform which had assessment modalities like equation solving or proofs and subsequently provides insights for teachers like identified knowledge gaps. In contrast, Öncel et al. (2021) attempted to estimate vocabulary using essays with the goal of providing more individualised feedback to learners on their writing skills. Therefore, they derive linguistic properties at four different levels: descriptive, lexical, syntax and cohesion and automatically predict the students' vocabulary score using several ML algorithms. In addition, there are also approaches that attempt to externalise students' implicit knowledge by using concept maps (Giabbanelli & Tawfik, 2020; Kim et al., 2019; Kim & McCarthy, 2021; Wu et al., 2012). Wu et al. (2012) compare concept maps of students with those of teachers using an existing key word list and provide hints on what students need to change. However, this assumes that the students use the same vocabulary as in the reference model. Therefore, Giabbanelli and Tawfik (2020) use a thesaurus database in order not to depend on an exact match between student and expert concept maps. Other works even analysed the effect of automatically generated knowledge structures (KS) from text as a reflection method to improve summary writings (Kim et al., 2019; Kim & McCarthy, 2021). They compared interventions using these KS with video instructions and multiple choice questions and found, that the reflection with KS improved the similarity to expert KS the most (Kim et al., 2019; Kim & McCarthy, 2021).

The presented work attempts to automatically identify, analyse or present students' knowledge gaps. Graphs are used to map concepts and their connections to each other, but to the best of our knowledge, no approach deals with eliciting students' conceptions.

ASSESSMENTS IN EDUCATION

This section present a selection of existing assessment tools used in biology education and computer science education. We use the examples to illustrate the current state of data-based FA with respect to students' conceptions and to identify desirable further developments from educational perspectives. We chose two topics that are of great interest in the respective field:

1. The biology education part addresses three successively developed assessment tools regarding the topic of evolutionary change. Students' conceptions in this topic has been widely researched in biology education. All tools have been recommended for teachers to be used in assessments with the purpose to gain insights in students' thinking and to support teaching.
2. The computer science education part focuses on students' skills and conceptions concerning programming and the automated assessment of students' code. We address the potential and limitations of existing automated assessments of code.

Assessments in biology education

The review of Kuschmierz et al. (2020) gives an overview of instruments that have been used for knowledge and understanding assessment in the domain of evolution. Existing instruments provide a score associated with levels of knowledge or understanding from a scientific perspective. The main item formats that were used are multiple choice, binary choice or scale items. Only few instruments used open-response formats. By design, closed formats have limited explanatory power for students' conceptions (Laskowski et al., 2018; von Aufschnaiter & Alonzo, 2018). However, assessment research provided information on frequencies of used biological conceptions (as variation) and some typical everyday conceptions (as need for adaptation). Moreover, they found that students' conceptions depend on the contexts used (like plants vs. animals or gain vs. loss of traits) (Anderson et al., 2002; Nehm & Ha, 2011). The biological conceptions and everyday conceptions considered in these tools subsume conceptions described in biology education (Hammann & Asshoff, 2014; Kampourakis, 2020; Weitzel, 2006).

In order to demonstrate assessment instruments on biological topics, we will refer to the CINS (Conceptual inventory of natural selection (Anderson et al., 2002)), the ACORNS (Assessing Contextual Reasoning about Natural Selection (Nehm, Beggrow, et al., 2012)) and EvoGrader (Moharreri et al., 2014). Table 1 gives a brief overview of the three instruments. All of them are tools to inform teachers about their students' conceptions (Anderson et al., 2002; Moharreri et al., 2014; Nehm, Beggrow, et al., 2012). They all address biological conceptions as well as everyday conceptions.

The CINS was designed to support constructivist and socioconstructivist learning (Anderson et al., 2002). Therefore, open question formats were used to gather information about everyday conceptions when the test was developed. The test itself uses multiple choice question formats. Everyday conceptions are used as distractors. The answers are evaluated as correct/incorrect. This question format has been criticised as it may lead to inaccurate scoring (Nehm & Schonfeld, 2008). The ACORNS instrument answered the critique of forced choice by using open response formats. In these formats, students have to give a

TABLE 1 Assessment instruments used in biology education

Assessment	Topic	Constructs	Question format
CINS (Anderson et al., 2002)	Evolution	10 key ideas	Multiple choice
ACORNS (Nehm, Beggrow, et al., 2012)	Evolution	7 key concepts, 6 naive ideas	Constructed response
EvoGrader (Moharreri et al., 2014)	Evolution	6 key concepts, 3 naive ideas	Constructed response

short text-based answer. Here, the presence/absence of biological conceptions and everyday conceptions is assessed based on students' expressions and the total number of both, biological conceptions and everyday conceptions is determined. A closer look reveals that ACORNS aims at grading of students. The inclusion of biological conceptions has positive score impact and the usage of everyday conceptions has negative score impact (Bertolini et al., 2021). Thus, the assessment itself is rather normative than aiming at an individual support of learning processes. Based on ACORNS items and ACORNS-like items, a free online automated assessment tool—the EvoGrader[1] (Moharreri et al., 2014)—has been developed. The automated scoring replaces the time-consuming manual scoring. Answers are scored for the presence/absence of biological conceptions and everyday conceptions. The tool uses automated scoring models that have been built using large human-scored data sets. EvoGrader (Moharreri et al., 2014) provides information on groups of learners. The information provided are visualisations of percentages of biological and everyday conceptions as well as their co-occurrences and a summarised categorisation of students' answers. The website proposes usages of the tool and provides teaching suggestions (learning activities like comparing and contrasting different types of explanations). Looking at the CINS, ACORNS and the EvoGrader, we can observe differences regarding

- the question formats used (which have changed from multiple choice formats to open response formats),
- the scoring (from non-automated to automated) and
- the information provided (either the individual presence or absence of conceptions or summarised frequencies of conceptions, co-occurrences and the summarised evaluation for a learning group also using visualisations).

We can see differences between the assessment tools that answer some practical needs of teachers as well as observed problems. Scoring, as offered by the above tools, however, is not considered helpful for the support of individual learning processes at all according to von Aufschnaiter and Alonzo (2018). Desirable future developments include more fine-grained assessments that provide nuanced information on students' individual biological conceptions and everyday conceptions in order to foster individualised learning (see Section 4).

Assessment in computer science education

Students' (mis-)conceptions have also been studied for various topics in computer science education at least since the 1980s (Bayman & Mayer, 1983).

A very important topic in computer science education is traditionally programming because of its long tradition and wide-spread use. The conceptions of learners (including preconceptions and misconceptions) in this field have been investigated regarding various programming concepts. A literature review of learners' misconceptions and other difficulties is provided, for example, by Qian and Lehman (2017).

There are various (often independent) initiatives to develop concept inventories for computer science in general and for introductory programming in particular (Caceffo et al., 2016, 2019; Henry & Dumas, 2020; Rachmatullah et al., 2020).

Caceffo et al. (2016, 2019) use a combination of different methods (theoretical identification of fundamental concepts, analysis of students' exam solutions, and interviews) to develop a concept inventory for basic programming concepts such as parameters, variables, recursion, loops, structures, pointers and Boolean expressions. They furthermore investigate the application of the items in different programming languages and highlight the difficulty of translating misconceptions into another programming language.

Rachmatullah et al. (2020) developed a concept inventory using a block-based programming language. The items focus on similar programming concepts (variables, loops, etc.) but on a more basic level compared to Caceffo et al. (2019).

Concept inventories usually consist of single-choice or multiple choice questions. However, in the context of learning programming, a very frequent exercise format is the submission of learners' source code. Such source code is a very open answer format, but more structured/formalised compared to natural language texts.

The assessment of learners' solutions is nowadays often supported by automated assessment systems (Keuning et al., 2018), in particular due to the high and increasing number of learners in this field which makes manual assessment infeasible. Keuning et al. (2018) provide an extensive systematic literature review of automated feedback generation for programming exercises (arguably one of the most important aspects of automatic FA). They state that most automatic feedback systems mainly report errors of the submitted solutions and that only very few systems explicitly take conceptions of the learners into account.

Keuning et al. (2018) furthermore find that most automatic assessment systems are not easily and flexibly adaptable by teachers to a specific learning setting. Flexible adaptions using custom models or rules are reported to be 'too specialised and time-consuming' (Keuning et al., 2018).

A recent literature review by Paiva et al. (2022) continues in this research direction and gives an overview of the state-of-the-art of automatic assessment in computer science education (with a focus on programming exercises). They additionally also consider aspects of LA and find that only 3 out of 30 assessment tools offer more than simple statistics for the teachers; those three tools provide in particular insights into the program development process of the learners (Paiva et al., 2022, Table 7). The authors furthermore conclude that a significant gap in this field is the lack of a standardised data format; this has partly been addressed by the proposal of Price et al. (2020).

For programming solutions, it is possible to infer (mis-)conceptions and competencies of learners from the submitted source code; this is often complemented by other methods like interviews or think-aloud etc. (Teague & Lister, 2014). A recent project aims to empirically identify competencies and (mis-)conceptions for object-oriented programming by performing a large-scale analysis of students' source code, complemented by manual qualitative code analyses and interviews (Krugel et al., 2020). The students' source code is stored in a graph-oriented data base which efficiently supports an automatic static analysis by using complex queries to identify solutions that contain certain sub-patterns (Koegl et al., 2022). However, the automatic elicitation and interpretation of students' conceptions regarding programming concepts as well as adequate automatic responses are still open problems.

Comparison between biology and computer science assessments

Here, we reflect on assessments with respect to the *target* (construct being assessed). The examples from Biology education and Computer Science education illustrate that learning subjects and targets of assessments might differ in nature:

1. Does the focus lie on the development of domain-specific skills, knowledge, the acquirement of practices or understanding/sense making of students?
2. What is assessed by existing tools?

The examples from biology education show steps from a normative to a more constructivist understanding of learning by integrating everyday conceptions in the reference model on students' conceptions. Particularly the last example, EvoGrader, supports insights for the design of group teaching/learning by providing explanatory visualisations. In biology, developments of FA tools aiming at individualised learning support are desirable. Moreover, the biological and everyday conceptions used in existing biology assessments are each rather a collection of selected conceptions (see Section 3.1) and provide a rather rough picture. More fine-grained, precise and more complex models of students' conceptions might be more appropriate to represent an individuals' conceptions and to assess what different students think. In CS education, tools have been developed that allow to evaluate source code provided by students. However, adaptability of the tools and usability by teachers seem to be an issue (Keuning et al., 2018). In CS education research, inferences from source code to students' conceptions or competencies are supported by the human-based analysis of additional data (text data). Therefore, further developments of FA tools targeted at students' conceptions could also profit from multimodal LA. For both subjects, more differentiated insights into students' conceptions, would necessarily need adapted forms of assessments.

DISCUSSION/SYNTHESIS

Challenges of formative assessment for students' conceptions

In Section 2.2, we have formulated the following basic challenges for teachers when they are required to consider students' conceptions in teaching:

1. FA of individual students' conceptions is time-consuming, as it has to be done regularly for a large number of students, and is therefore hardly practicable.
2. The analysis of students' statements aiming at differentiated insights about their thinking is challenging for teachers.

To meet these challenges would allow to address individual students' conceptions and to support meaningful learning processes. This has been described as an essential educational aim for different subjects (Reinfried et al., 2009) for geography, history and biology. With respect to *question formats*, current approaches to assessment often use closed-response type question formats like binary or multiple choice or short-answer formats that are easy to implement and score at large scale (Laskowski et al., 2018; Nehm, Ha, et al., 2012). These formats are limited in capturing the complexity and variety of students' conceptions. Assessments that target the construction of complex scientific explanations in science class also suffer in this regard (Nehm, Ha, et al., 2012). In non-technology-based education research, open-response formats (like asking for explanations) have often been used to assess students' conceptions (Nehm, Ha, et al., 2012). When scored manually, some challenges of open-response formats according to Nehm, Ha, et al. (2012) are rubric development and validation costs, grading time or grading fatigue. On the other hand, in order to gain insights into students' conceptions and their complexity, open question formats are more appropriate than closed formats since they allow students to express their ideas. This results in more unstructured data (text data) and therefore calls for different analysis techniques.

The review of Zhai et al. (2021) on ML-based science assessments shows that almost all (except one) considered studies aiming at the assessment of students' conceptual understanding used constructed-response formats where students have to give short text-based answers. Limitations and challenges with respect to the application of ML techniques have already been discussed in Section 2.3. Even a method that allows a succeeding categorisation of students' statements would offer a fundament for individualised learning processes. Furthermore, even if prominent conceptions are subject-specific (see Section 3.3), this method could be adapted for the conceptual structures of different subjects.

Clustering and classification of text data

In education research, the work of Sherin (2013) is an example for the application of statistical NLP techniques in the domain of physics education for the analysis of students' explanations about the existence of seasons. He wanted to explore whether the analysis would provide interpretable results similar to those found with more traditional human-based qualitative analyses (Sherin, 2013). He selected passages from transcribed one-on-one standardised clinical interviews and preprocessed the transcripts by removing annotations (eg, gestures), interviewer prompts, punctuation and stop words as well as by segmenting the remaining text in overlapping 100 word chunks. Sherin used a vector space model (bag-of-words assumption) and clustering to identify types of explanations. He interpreted the resulting clusters with respect to existing findings on typical explanation types and observed that not all of the elements that appeared in traditional human analyses could be captured by his analysis technique. He used visualisations (bar charts) to identify the dominant explanation types for each segment and explanation type shifts in the course of the interviews. Sherin's approach to use NLP techniques for the analysis of students' explanations seems promising and should be further explored for FA purposes.

The example shows that rather simple techniques could be applied to separate different explanation types and detect changes in explanation types. In the example of the four seasons, explanation types consist of closer-farther, side-based and tilt-based connections between the earth and the sun. If these classes are known, both supervised and unsupervised approaches (as in Sherin, 2013) could be used, depending on whether sample data is available or not. Both technology groups could be used to sort and select student texts in formative assessment, and inform teacher actions. This classification may ease the step of *eliciting* in formative assessment. In this approach, however, the step of interpretation still needs to be performed by a human with respect to existing educational theory. The necessity of this manual step might represent a drawback. Moreover, for the studied topic, educational research had revealed only three explanation types. It would be interesting to see, whether this procedure may be used for other topics to generate meaningful clusters of explanation types also for more complex topics. Such a topic could be, for instance, evolution in biology education, for which nine explanation pattern where identified and described (Zabel & Gropengiesser, 2011).

Challenges of written texts

Sherin (2013) used transcribed interviews for his analyses. An open question is how the analysis would have worked on written text. Not at least since written texts from students come with further challenges:

- students' typing errors,
- punctuation mistakes,
- incomplete sentences,
- idiosyncratic usage of technical terms,
- ambiguous referents and
- verbosity of students' impacts explanation performance measures (Federer et al., 2015).

NLP techniques could be used to handle or mitigate some of these problems (Jurafsky et al., 2009; Jurafsky & Martin, 2021). As an example, Lee, Kim, et al. (2020) supposes an approach to detect typing errors by using a deep learning language model. Another example is the transformer-based method of Švec et al. (2021) for reconstructing punctuation in automatically generated speech transcripts.

Incorporating semantics using deep learning-based language models

Sherin's work (2013) was published in 2013, long before the rise of large deep learning based language models such as BERT (Devlin et al., 2019) or GPT-3 (Brown et al., 2020). These are able to represent natural text in numerical vectors and ideally encapsulate semantics. However, these are trained on domain-independent texts. It would be exciting to test the above proposals using these, or specialised language models such as BioBERT (Lee, Yoon, et al., 2020), SciBERT (Beltagy et al., 2019) or EduBERT (Clavié & Gal, 2019). While BERT (Devlin et al., 2019) is based on large corpora consisting of Wikipedia articles and books, BioBERT (Lee, Yoon, et al., 2020) additionally used abstracts and full texts from biomedical papers on PubMed[2] for training. SciBERT (Beltagy et al., 2019) instead was trained on a large corpus from SemanticScholar (Ammar et al., 2018), which consists of papers from the computer science and biomedical domain. The authors of EduBERT (Clavié & Gal, 2019) aimed to build a language model that can be used for LA tasks. For this purpose, it was trained on forum data from MOOCs and courses at universities. New methods could use their representations to determine similarity to other student or reference solutions. Therefore, these technologies could provide support for the *eliciting* step, but also for the *classification* of the students' answers.

Visual analytics

Visual dashboards could provide visual summaries of student information to support the teacher in choosing successful interventions. Kim et al. (2019) and Kim and McCarthy (2021) automatically extracted and visualised graph structures from students' written text to facilitate students' own reflections. They have colour-coded the connections between concepts to indicate whether connections are (incorrectly) present or missing compared to a reference model. A similar view, which provides aggregated information the essays of a whole class could be a valuable support for a teacher as it allows to obtain a quick overview over the contents of the students' writings. Future work could focus on the extraction and labelling of the links between terms and provide insights about the similarities and differences among students (eg, one specific link between concepts is missing among the whole class).

Multimodal solutions

In this article, we focused on the analysis of textual student artefacts. Another interesting direction for future research efforts could be the analysis of both text and image (Shelton et al., 2016). For example, students could draw a concept map or an image of a conception or process in class and use a text to explain their illustration (Smith et al., 2019). Analyses of commonalities and differences between text and image are widely used in research on students' conceptions (Kattmann, 2015).

CONCLUSION AND FUTURE DIRECTIONS

LA has already been successfully used for assessment purposes (eg, automated scoring, visualisations of students' data) and to inform teachers. In this article, we focused on potential contributions of LA to support teachers FA practices with regard to the assessment of students' conceptions. We synthesised findings on existing assessment tools for FA support in biology and CS education and identified limitations with regard to the assessment of students' conceptions. The need to use open question formats results in more unstructured text data. NLP techniques for the analysis of unstructured text data with the purpose of eliciting students' explanation types seems to be promising (Sherin, 2013), but should further be explored. Chen et al. (2020) also argue to seek the potential of NLP in promoting precision and personalised education and NLP has been one technological focus in the AI in Education field (Feng & Law, 2021). Moreover, using different modalities of data, like drawings, concept maps, text or even source code (or their combination), seems promising to develop more informative assessments targeted at students' conceptions. The approach to apply different modalities of data has already been successfully used in educational research on students' conceptions (eg, see Gropengießer, 1997 or Kattmann, 2015). Therefore, the use of multimodal LA should be further explored. Future work on the application of LA techniques as support for FA on students' conceptions should include a thorough investigation of quality criteria of methods and measurements as has been done for other applications of LA (eg, see Winne, 2020). When aiming at teachers support, explanatory models with interpretable structures (Rosé et al., 2019) from which insights about the students' conceptions might be derived are preferable. These models could be built based on educational findings on students' conceptions or student data. Visual learning analytics (VLA) has been used to support teachers FA practices by providing visualisations of data for the learning group (in biology education) or for individual learners (in CS education). Future work could continue to study how FA practices of teachers may be supported by VLA techniques (Echeverria et al., 2018).

ACKNOWLEDGEMENTS

This work has been partly supported by the Ministry of Science and Education of Lower Saxony, Germany, through the Graduate training network "LernMINT: Data-assisted classroom teaching in the STEM subjects". We would like to thank the reviewers for their valuable feedback. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

There is no conflict of interest associated with this research.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ETHICS STATEMENT

Ethical approval is not applicable, because this article presents a synthesis based on published studies.

ORCID

Judith Stanja https://orcid.org/0000-0002-4137-9294
Wolfgang Gritz https://orcid.org/0000-0003-1668-3304
Johannes Krugel https://orcid.org/0000-0003-0967-4872
Anett Hoppe https://orcid.org/0000-0002-1452-9509

ENDNOTES

[1] http://www.evograder.org/.

[2] https://pubmed.ncbi.nlm.nih.gov/.

REFERENCES

Alonzo, A. C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education*, *31*(2), 104–112. https://doi.org/10.1080/08957347.2017.1408630

Amin, T. G. (2015). Conceptual metaphor and the study of conceptual change: Research synthesis and future directions. *International Journal of Science Education*, *37*(5–6), 966–991. https://doi.org/10.1080/09500693.2015.1025313

Amin, T. G., Smith, C. L., & Wiser, M. (2011). Student conceptions and conceptual change. In *Handbook of research on science education* (Vol. II). Routledge. https://doi.org/10.4324/9780203097267.ch4

Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H.-H., Peters, M. E., Power, J., Skjonsberg, S., Wang, L. L., … Etzioni, O. (2018). Construction of the literature graph in semantic scholar. In S. Bangalore, J. Chu-Carroll, & Y. Li (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1–6, 2018, Volume 3 (Industry Papers)* (pp. 84–91). Association for Computational Linguistics. https://doi.org/10.18653/v1/n18-3011

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*(10), 952–978. https://doi.org/10.1002/tea.10053

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (Vancouver, British Columbia, Canada) (LAK '12)* (pp. 267–270). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung—Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*(1), 7–28.

Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cognitive Psychology*, *104*, 1–28. https://doi.org/10.1016/j.cogpsych.2018.03.002

Bayman, P., & Mayer, R. E. (1983). A diagnosis of beginning programmers' misconceptions of BASIC programming statements. *Communications of the ACM*, *9*, 677–679. https://doi.org/10.1145/358172.358408

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3615–3620). Association for Computational Linguistics, Hong Kong, China. https://doi.org/10.18653/v1/D19-1371

Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology. *Journal of Science Education and Technology*, *30*(2), 193–209. https://doi.org/10.1007/s10956-020-09888-8

Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2020). Learning dashboards at scale: Early warning and overall first year experience. *Assessment & Evaluation in Higher Education*, *45*(6), 855–874. https://doi.org/10.1080/02602938.2019.1689546

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M.'. A. Ranzato, R. Hadsell, M.-F. Balcan, & H.-T. Lin (Eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6–12, 2020, Virtual*. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Caceffo, R., Frank-Bolton, P., Souza, R., & Azevedo, R. (2019). Identifying and validating Java misconceptions toward a CS1 concept inventory. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education (Aberdeen, Scotland UK) (ITiCSE '19)* (pp. 23–29). Association for Computing Machinery. https://doi.org/10.1145/3304221.3319771

Caceffo, R., Wolfman, S., Booth, K. S., & Azevedo, R. (2016). Developing a computer science concept inventory for introductory programming. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (Memphis, Tennessee, USA) (SIGCSE '16)* (pp. 364–369). Association for Computing Machinery. https://doi.org/10.1145/2839509.2844559

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Clavié, B., & Gal, K. (2019). EduBERT: Pretrained deep language models for learning analytics. *CoRR*. http://arxiv.org/abs/1912.00690

Cusi, A., Morselli, F., & Sabena, C. (2019). The use of polls to enhance formative assessment processes in mathematics classroom discussions. In G. Aldon & J. Trgalová (Eds.), *Technology in mathematics teaching* (pp. 7–30). Mathematics Education in the Digital Era, Vol. 13. Springer International Publishing. https://doi.org/10.1007/978-3-030-19741-4_1

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/n19-1423

diSessa, A. A. (2014). A history of conceptual change research. In R. Keith Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 88–108). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.007

Echeverria, V., Martinez-Maldonado, R., Shum, S. B., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, *5*(3), 72–97. https://doi.org/10.18608/jla.2018.53.6

Federer, M. R., Nehm, R. H., Opfer, J. E., & Pearl, D. (2015). Using a constructed-response instrument to explore the effects of item position and item features on the assessment of Students' written scientific explanations. *Research in Science Education*, *45*(4), 527–553. https://doi.org/10.1007/s11165-014-9435-9

Feng, S., & Law, N. (2021). Mapping artificial intelligence in education research: A network–based keyword analysis. *International Journal of Artificial Intelligence in Education*, *31*(2), 277–303. https://doi.org/10.1007/s40593-021-00244-4

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127–203.

Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. *Computers in Human Behavior*, *134*, 107304. https://doi.org/10.1016/j.chb.2022.107304

Giabbanelli, P. J., & Tawfik, A. A. (2020). Reducing the gap between the conceptual models of students and experts using graph-based adaptive instructional systems. In C. Stephanidis, D. Harris, W.-C. Li, D. D. Schmorrow, C. M. Fidopiastis, P. Zaphiris, A. Ioannou, X. Fang, R. A. Sottilare, & J. Schwarz (Eds.), *HCI International 2020—Late breaking papers: Cognition, learning and games—22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12425)* (pp. 538–556). Springer. https://doi.org/10.1007/978-3-030-60128-7_40

Gropengießer, H. (1997). *Didaktische Rekonstruktion des Sehens: Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung: Zugl.: Oldenburg, Univ., Diss., 1997*. Carl von Ossietzky-Univ. Zentrum für Pädag.

Gropengießer, H. (2006). *Wie man Vorstellungen der Lerner verstehen kann: Lebenswelten, Denkwelten, Sprechwelten* (2 aufl. ed.). Beiträge zur didaktischen Rekonstruktion, Vol. 4. Didaktisches Zentrum Carl-von-Ossietzky-Univ.

Gropengießer, H., & Marohn, A. (2018). Schülervorstellungen und conceptual change. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (pp. 49–67). Springer. https://doi.org/10.1007/978-3-662-56320-5_4

Groß, J., Hammann, M., Schmiemann, P., & Zabel, J. (Eds.). (2019). *Biologiedidaktische Forschung: Erträge für die Praxis*. Springer Spektrum. https://doi.org/10.1007/978-3-662-58443-9

Hammann, M., & Asshoff, R. (2014). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten*. Klett/Kallmeyer.

Henry, J., & Dumas, B. (2020). Approach to develop a concept inventory informing teachers of novice programmers' mental models. In *IEEE Frontiers in Education Conference, FIE 2020*, Uppsala, Sweden, October 21–24, 2020, IEEE, pp. 1–9. https://doi.org/10.1109/FIE44824.2020.9274045

Jensen, E., Umada, T., Hunkins, N. C., Hutt, S., Huggins-Manley, A. C., & D'Mello, S. K. (2021). What you do predicts how you do: Prospectively modeling student quiz performance using activity features in an online learning environment. In *LAK21: 11th International Learning Analytics and Knowledge Conference (Irvine, CA, USA) (LAK21)* (pp. 121–131). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448151

Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/ed3book_sep212021.pdf

Jurafsky, D., Martin, J. H., Norvig, P., & Russell, S. J. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, prentice hall series in artificial intelligence*. Prentice Hall, Pearson Education International.

Kaliisa, R., & Dolonen, J. A. (2022). CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-022-09598-7

Kampourakis, K. (2020). Students' "teleological misconceptions" in evolution education: Why the underlying design stance, not teleology per se, is the problem. *Evolution: Education and Outreach*, *13*(1), 1–12. https://doi.org/10.1186/s12052-019-0116-z

Kattmann, U. (2015). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht: [zusätzliche Stichwörter zum Download]*. Aulis Verlag.

Keuning, H., Jeuring, J., & Heeren, B. (2018). A systematic literature review of automated feedback generation for programming exercises. *ACM Transactions on Computing Education*, *19*(1), 43. https://doi.org/10.1145/3231711

Kim, K., Clarianay, R. B., & Kim, Y. (2019). Automatic representation of knowledge structure: Enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development*, *67*, 105–122. https://doi.org/10.1007/s11423-018-9626-6

Kim, M. K., & McCarthy, K. S. (2021). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *37*(3), 684–704. https://doi.org/10.1111/jcal.12516

Knight, S., Shibani, A., Abel, S., Gibson, A., & Ryan, P. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *The Journal of Writing Research*, *12*, 141–186. https://doi.org/10.17239/jowr-2020.12.01.06

Koegl, A., Hubwieser, P., Talbot, M., Krugel, J., Striewe, M., & Goedicke, M. (2022). Efficient structural analysis of source code for large scale applications in education. In I. Kallel, H. M. Kammoun, & L. Hsairi (Eds.), *IEEE Global Engineering Education Conference, EDUCON 2022, Tunis, Tunisia, March 28–31, 2022* (pp. 24–30). IEEE. https://doi.org/10.1109/EDUCON52537.2022.9766748

Kokoç, M., Akçapinar, G., & Hasnine, M. N. (2021). Unfolding students' online assignment submission behavioral patterns using temporal learning analytics. *Educational Technology & Society*, *24*(1), 223–235. https://www.jstor.org/stable/26977869

Krugel, J., Hubwieser, P., Goedicke, M., Striewe, M., Talbot, M., Olbricht, C., Schypula, M., & Zettler, S. (2020). Automated measurement of competencies and generation of feedback in object-oriented programming courses. In *2020 IEEE Global Engineering Education Conference, EDUCON 2020*, Porto, Portugal, April 27–30, 2020 (pp. 329–338). IEEE. https://doi.org/10.1109/EDUCON45650.2020.9125323

Krüger, D. (Ed.). (2007). *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden; mit 12 Tabellen*. Springer.

Krüger, D., Parchmann, I., & Schecker, H. (Eds.). (2018). *Theorien in der naturwissenschaftsdidaktischen Forschung*. Springer. https://doi.org/10.1007/978-3-662-56320-5

Kuschmierz, P., Meneganzin, A., Pinxten, R., Pievani, T., Cvetković, D., Mavrikaki, E., Graf, D., & Beniermann, A. (2020). Towards common ground in measuring acceptance of evolution and knowledge about evolution across Europe: A systematic review of the state of research. *Evolution: Education and Outreach*, *13*(1), 1–24. https://doi.org/10.1186/s12052-020-00132-w

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.

Laskowski, P., Karayev, S., & Hearst, M. A. (2018). How do professors format exams? An analysis of question variety at scale. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (London, United Kingdom) (L@S '18)* (p. 10). Association for Computing Machinery, New York, NY . https://doi.org/10.1145/3231644.3231667.Article 54

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinform*, *36*(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682

Lee, J.-H., Kim, M., & Kwon, H.-C. (2020). Deep learning-based context-sensitive spelling typing error correction. *IEEE Access*, *8*, 152565–152578. https://doi.org/10.1109/ACCESS.2020.3014779

Liu, M., Kitto, K., & Shum, S. B. (2021). Combining factor analysis with writing analytics for the formative assessment of written reflection. *Computers in Human Behavior*, *120*, 106733. https://doi.org/10.1016/j.chb.2021.106733

Minn, S. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence*, *3*, 100050. https://doi.org/10.1016/j.caeai.2022.100050

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior*, *117*, 106658. https://doi.org/10.1016/j.chb.2020.106658

Moharreri, K., Ha, M., & Nehm, R. H. (2014). EvoGrader: An online formative assessment tool for automatically evaluating written evolutionary explanations. *Evolution: Education and Outreach*, *7*(1), 1–14. https://doi.org/10.1186/s12052-014-0015-2

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection: Diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher*, *74*(2), 92–98. https://doi.org/10.1525/abt.2012.74.2.6

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*(3), 237–256. https://doi.org/10.1002/tea.20400

Nehm, R. H., Ha, M., & Mayfield, E. (2012). Transforming biology assessment with machine learning: Automated scoring of written evolutionary explanations. *Journal of Science Education and Technology*, *21*(1), 183–196. https://doi.org/10.1007/s10956-011-9300-9

Nehm, R., & Schonfeld, I. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, *45*(12), 1131–1160. https://doi.org/10.1002/tea.20251

Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing. In *LAK21: 11th International Learning Analytics and Knowledge Conference (Irvine, CA, USA) (LAK21)* (pp. 620–625). Association for Computing Machinery, New York, NY. https://doi.org/10.1145/3448139.3448207

Paiva, J. C., Leal, J. P., & Figueira, Á. (2022). Automated assessment in computer science education: A state-of-the-art review. *ACM Transactions on Computing Education*, *22*(3), 40. https://doi.org/10.1145/3513140

Potvin, P., Nenciovici, L., Malenfant-Robichaud, G., Thibault, F., Sy, O., Mahhou, M. A., Bernard, A., Allaire-Duquette, G., Sarrasin, J. B., Foisy, L.-M. B., Brouillette, N., St-Aubin, A.-A., Charland, P., Masson, S., Riopel, M., Tsai, C.-C., Bélanger, M., & Chastenay, P. (2020). Models of conceptual change in science learning: Establishing an exhaustive inventory based on support given by articles published in major journals. *Studies in Science Education*, *56*(2), 157–211. https://doi.org/10.1080/03057267.2020.1744796

Price, T. W., Hovemeyer, D., Rivers, K., Gao, G., Bart, A. C., Kazerouni, A. M., Becker, B. A., Petersen, A., Gusukuma, L., Edwards, S. H., & Babcock, D. (2020). ProgSnap2: A flexible format for programming process data. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (Trondheim, Norway) (ITiCSE '20)* (pp. 356–362). Association for Computing Machinery, New York, NY. https://doi.org/10.1145/3341525.3387373

Qian, Y., & Lehman, J. (2017). Students' misconceptions and other difficulties in introductory programming: A literature review. *ACM Transactions on Computing Education*, *18*(1), 24. https://doi.org/10.1145/3077618

Rachmatullah, A., Akram, B., Boulden, D., Mott, B., Boyer, K., Lester, J., & Wiebe, E. (2020). Development and validation of the middle grades computer science concept inventory (MG-CSCI) assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, *16*, em1841. https://doi.org/10.29333/ejmste/116600

Reinfried, S., Mathis, C., & Kattmann, U. (2009). Das Modell der Didaktischen Rekonstruktion. Eine innovative Methode zur fachdidaktischen Erforschung und Entwicklung von Unterricht. *Beiträge zur Lehrerbildung*, *27*(3), 404–414. https://doi.org/10.25656/01:13710

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, *50*(6), 2943–2958. https://doi.org/10.1111/bjet.12858

Saqr, M., & Alamro, A. (2019). The role of social network analysis as a learning analytics tool in online problem based learning. *BMC Medical Education*, *19*(1), 1–11. https://doi.org/10.1186/s12909-019-1599-6

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, *6*(3), 11–19. https://doi.org/10.18608/jla.2019.63.3

Shelton, A., Smith, A., Wiebe, E., Behrle, C., Sirkin, R., & Lester, J. (2016). Drawing and writing in digital science notebooks: Sources of formative assessment data. *Journal of Science Education and Technology*, *25*(3), 474–488. https://doi.org/10.1007/s10956-016-9607-7

Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, *22*(4), 600–638. https://doi.org/10.1080/10508406.2013.836654

Siemens, G., Gasevic, D., Haythornthwaite, C., Shane, D., Buckingham Shum, S., Ferguson, R., Duval, E., Katrien, V., & Baker, R. S. (2011). *Open learning analytics: An integrated & modularized platform*. Open University Press Maidenhead.

Smith, A., Leeman-Munk, S. P., Shelton, A., Mott, B. W., Wiebe, E. N., & Lester, J. C. (2019). A multimodal assessment framework for integrating student writing and drawing in elementary science learning. *IEEE Transactions on Learning Technologies*, *12*(1), 3–15. https://doi.org/10.1109/TLT.2018.2799871

Švec, J., Lehečka, J., Šmídl, L., & Ircing, P. (2021). Transformer-based automatic punctuation prediction and word casing reconstruction of the ASR output. In K. Ekštein, F. Pártl, & M. Konopík (Eds.), *Text, speech, and dialogue* (pp. 86–94). Springer International Publishing.

Teague, D., & Lister, R. (2014). Longitudinal think aloud study of a novice programmer. In J. L. Whalley & D. J. D'Souza (Eds.), *Sixteenth Australasian Computing Education Conference, ACE 2014, Auckland, New Zealand, January 2014 (CRPIT, Vol. 148)* (pp. 41–50). Australian Computer Society. http://dl.acm.org/citation.cfm?id=2667495

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). Formative assessment and learning analytics. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge—LAK '13* (p. 205). ACM Press. https://doi.org/10.1145/2460296.2460337

von Aufschnaiter, C., & Alonzo, A. C. (2018). Foundations of formative assessment: Introducing a learning progression to guide preservice physics teachers' video-based interpretation of student thinking. *Applied Measurement in Education*, *31*(2), 113–127. https://doi.org/10.1080/08957347.2017.1408629

Vosniadou, S. (Ed.). (2013). *International handbook of research on conceptual change*. Routledge. https://doi.org/10.4324/9780203154472

Weitzel, H. (2006). *Biologie verstehen: Vorstellungen zu Anpassung: Zugl.: Hannover, Univ., Diss., 2006 u.d.T.: Weitzel, Holger: Anpassung verstehen lernen: Lehr-Lern-Forschung nach dem Modell der didaktischen Rekonstruktion unter kognitionslinguistischer Perspektive* (1 aufl. ed.). Beiträge zur didaktischen Rekonstruktion, Vol. 15. Didaktisches Zentrum Univ.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior*, *112*, 106457. https://doi.org/10.1016/j.chb.2020.106457

Wise, A. F. (2019). *Learning analytics: Using data-informed decision-making to improve teaching and learning* (pp. 119–143). Springer International Publishing. https://doi.org/10.1007/978-3-319-89680-9_7

Wu, P.-H., Hwang, G.-J., Milrad, M., Ke, H.-R., & Huang, Y.-M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, *43*(2), 217–232. https://doi.org/10.1111/j.1467-8535.2010.01167.x

Yilmaz, F. G. K., & Yilmaz, R. (2022). Learning analytics intervention improves students' engagement in online learning. *Technology, Knowledge and Learning*, *27*(2), 449–460. https://doi.org/10.1007/s10758-021-09547-w

Zabel, J., & Gropengiesser, H. (2011). Learning progress in evolution theory: Climbing a ladder or roaming a landscape? *Journal of Biological Education*, *45*(3), 143–149. https://doi.org/10.1080/00219266.2011.586714

Zhai, X., Shi, L., & Nehm, R. H. (2021). A meta-analysis of machine learning-based science assessments: Factors impacting machine-human score agreements. *Journal of Science Education and Technology*, *30*(3), 361–379. https://doi.org/10.1007/s10956-020-09875-z

Zheng, G., Fancsali, S. E., Ritter, S., & Berman, S. (2019). Using instruction-embedded formative assessment to predict state summative test scores and achievement levels in mathematics. *Journal of Learning Analytics*, *6*(2), 153–174. https://doi.org/10.18608/jla.2019.62.11

How to cite this article: Stanja, J., Gritz, W., Krugel, J., Hoppe, A., & Dannemann, S. (2023). Formative assessment strategies for students' conceptions—The potential of learning analytics. *British Journal of Educational Technology*, *54*, 58–75. https://doi.org/10.1111/bjet.13288