JOURNAL OF CHEMICAL EDUCATION

Open Access

This article is licensed under CC-BY 4.0

pubs.acs.org/jchemeduc Article

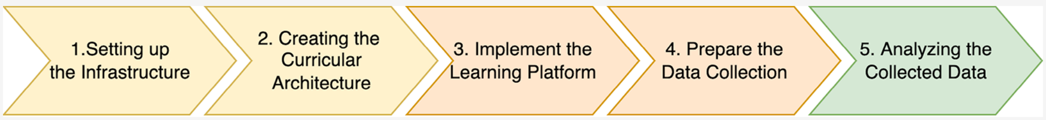
Exploring Data Mining in Chemistry Education: Building a Web-Based Learning Platform for Learning Analytics

Marvin Roski,* Ralph Ewerth, Anett Hoppe, and Andreas Nehring

Cite This: *J. Chem. Educ.* 2024, 101, 930−940

Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: The integration of learning analytics and artificial intelligence methods into education is part of the latest developments and significantly affects chemistry education (research): researchers might face the challenge of collecting and analyzing content-rich data sets involving interdisciplinary approaches from computer science, chemistry, and chemistry education. Developing a learning platform offers a higher degree of freedom compared to using existing Learning Management Systems. This paper presents a step-by-step overview of how we designed and created the web-based learning platform "I_3Learn". This concrete example showcases how the content management system WordPress can be utilized to develop an online learning infrastructure that enables the application of learning analytics methods. Our approach allows collection of time-accurate log files concerning individual usage behaviors without programming skills. The step-by-step overview enables chemistry education researchers to set up a customized web-based learning platform and lay the data foundation for learning analytics. The value of the I_3Learn platform is shown through a compelling use-case scenario involving 580 9th- and 10th-grade learners in German chemistry classrooms. In this scenario, we collected and analyzed half a million log files, and we report exemplary results and insights from the data.

KEYWORDS: *Internet/Web-Based Learning, Ionic Bonding*

Downloaded via 89.245.22.238 on March 26, 2024 at 06:49:18 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

■ INTRODUCTION

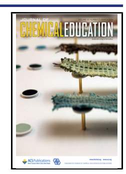
Web-based learning is essential for numerous educational settings, such as blended or remote learning. In addition, so-called Massive Open Online Courses (MOOCs) are established in the field of (scientific) education, providing access to education for many learners.[1] With the increasing technological progress in recent decades comes the exploration of artificial intelligence integrations in (science) education.[2] Within this development, learning analytics deals with analyzing learning process data to optimize the teaching processes,[3] for instance, by automating the analysis of learners' scientific argumentation.[4]

Mining educational data opens the door to integrating artificial intelligence into teaching and learning processes and enables new approaches to address ongoing challenges in chemistry education: The COVID-19 pandemic has also demonstrated the need for remote learning tools for chemistry education.[5,6] A comprehensive analysis of the effective use of these educational resources can be enhanced by examining log files, especially when utilizing the provided digital features, which may not be immediately evident.[7] Fine-grained log file data also allows for tracking and understanding learner interaction with assessment tasks, identifying factors that influence learning performance, and drawing conclusions on latent constructs such as learning strategies.[8]

Furthermore, the pursuit of integrating diversity, equity, inclusion, and respect (DEIR) is an essential focus of chemistry education.[9] While the risk of artificial intelligence to contribute to discrimination is widely discussed, extensive and ethically informed data collection can also provide the basis for individualized learning in order to improve support for vulnerable groups, identify learning barriers, deconstruct stigmatizing categories, and reconstruct categories based on learning behaviors.[10−15]

Consequently, technological means enable and require data collection on an unprecedented scale, presenting interested researchers with the challenge of organizing the collection, analysis, and interpretation of the resulting data.[16] One way to accomplish this is to collect log files to track learner behavior in a (web-based) learning platform.[17] Unfortunately, not all systems that can be used to design learning platforms can collect or extract log files, or educators might be confronted with investing many resources into developing competencies in programming and data analysis. The open source content

Received: August 9, 2023
Revised: January 16, 2024
Accepted: January 19, 2024
Published: February 13, 2024



ACS Publications

© 2024 The Authors. Published by American Chemical Society and Division of Chemical Education, Inc.

https://doi.org/10.1021/acs.jchemed.3c00794
J. Chem. Educ. 2024, 101, 930−940

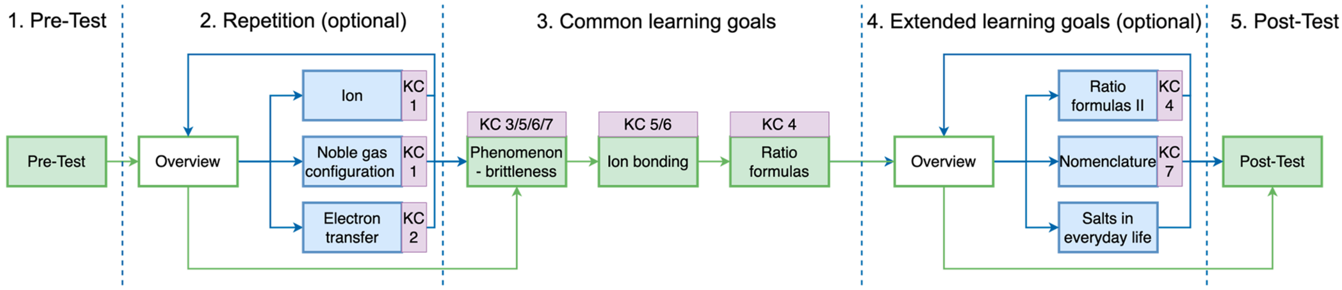


Figure 1. Simplified structure and order of the chapters in the I_3Learn learning platform. The chapters (blue or green boxes) are arranged in five sections: (1) pretest, (2) chapters with content that is a prerequisite to the topic of ion bonding and can be repeated voluntarily, (3) chapters on achieving the common learning goal (ion bonding), (4) chapters on extended learning goals that high-performing learners can achieve, and (5) the post-test.

management system WordPress[18] served as the foundation for developing the learning platform "I_3Learn" (acronym: inclusive, intelligent, and individual learning), designed to support chemistry learning and teaching, specifically in the field of ion bonding. In Germany, 580 9th and 10th grade students used I_3Learn, resulting in 493,511 data log files. The log file data enable a second-by-second reconstruction of the students' usage behavior on the learning platform. This can be used as a basis for learning analytics to improve learning, understand learning, and use learning material through statistical and artificial intelligence methods.[3,19,20]

This review is divided into four further sections. First, requirements and preliminary considerations for a web-based platform are presented in the next section. Second, the learning platform I_3Learn is introduced as an exemplary implementation of a WordPress-based learning platform for K-12 chemistry education. Furthermore, this section highlights key conceptual considerations within the architecture and illustrates how learning analytics, implemented through a web-based learning platform, can provide valuable insights. The third section provides a step-by-step overview of creating a learning platform with WordPress to collect data for learning analytics. The main steps and aspects for creating a learning platform are shown. The creation process requires no in-depth technological knowledge, while granting considerable freedom in design and content. The last section concludes the article, points out limitations, and provides an outlook on future research directions.

Requirements and Preliminary Considerations

Due to the SARS-CoV-2 pandemic, many educators and researchers were forced to switch to a remote learning setting to ensure access to education and research.[21] One way to address this issue is to use WordPress to design a custom web-based learning platform that provides the basis for collecting extensive log files about the usage behavior of learners. According to WordPress, 43% of all Web sites are created using this open source content management system.[22] It is highly accessible due to its clear graphical interface. With WordPress, it is possible (a) to design learning content freely, considering and implementing characteristics of chemistry education, and to integrate digital and real-world lab experiments; (b) to create a web-based learning platform usable from all devices (computers, smartphones, and tablets); and (c) to record log files of the usage behavior for learning analytics.

When designing a web-based learning platform, we placed high demands on it. This means high demands on data security, the possibility to choose free and individual learning chapters (many known Learning Management Systems (LMSs) only offer a hierarchical structure of the learning chapters), the possibility to switch between multiple representations (text and videos) without complications, the integration of accessibility tools (such as a read-aloud function or a completely customizable interface), and the possibility to record every interaction of the learners as a log file. No existing LMS that could meet all of these requirements can be identified.

A general prerequisite for a learning platform is the possibility of hosting it on in-house servers to have physical access to the data and comply with the data protection regulations of the country in which the learning platform is to be established, such as the General Data Protection Regulation (GDPR) from the European Union (EU), which complicates the processing of data on servers outside the EU.[23] In this regard, external cloud-based providers are likely out of the question, especially when data collection requires approval by external authorities such as (local) education government departments. Furthermore, a WordPress-based system contributes to data protection, since the learners do not have to log in with university or commercial accounts. In contrast, entirely anonymous and independent accounts that cannot be traced back to individuals can be used with the approach described in this paper.

Developing a learning platform gives more freedom than using an existing LMS, especially concerning integrating assistance systems, designing the learning platform, or managing learners. As in this approach, other existing LMSs, such as Moodle,[24] one of the most widespread and well-known LMSs in the world,[25,26] can also rely on software to track learner activity. However, design freedom is more restricted compared to that in a WordPress environment, where limitations are absent, allowing for the creation of the entire structure of the learning platform without constraints. In addition, Moodle must also be hosted on its server; therefore, in terms of effort, there is no difference. Other platforms, such as the social web annotation tool Perusall,[27] already offer learning analytics capabilities based on artificial intelligence but focus more on collaborative learning.

Furthermore, on an institutional level, there may be the restriction that no comparable LMS may be hosted since universities, for example, have usually established their system on campus, and the coexistence of a separate LMS is not permitted. Using these existing LMSs at the institutional level can also have restrictions: For example, the institution may set the privacy setting so that no data can be recorded and

https://doi.org/10.1021/acs.jchemed.3c00794
J. Chem. Educ. 2024, 101, 930−940

downloaded, or external persons, such as students from K-12 education, do not have access to a university or higher education LMS. Open-source content management systems have the potential to fill this gap because the default institutional settings do not apply to a learning platform created for a specific research project. Furthermore, it must be noted that data protection is essential in the everyday use of an institutional LMS. At the same time, this makes it even more important to have a separate coexisting web-based learning platform for research purposes, as offered in this step-by-step overview.

■ THE LEARNING PLATFORM I$_3$LEARN

Using WordPress, a web-based learning platform called "I$_3$Learn" has been developed for four learning sessions, each lasting 45 min. The platform's primary objective is to facilitate independent and remote learning for the learners. "I$_3$Learn" was utilized to study ion bonding, a topic recognized as presuppositional, which demands comprehension of fundamental concepts such as atom and ion distinctions, noble gas configuration, and electron transfer.[28] Additionally, the platform aimed to offer access to differentiated instruction.[29] Accordingly, the structure allows learners to work on chapters with different focuses (see Figure 1). The individual learning chapters (blue or green boxes in Figure 1) are divided into five sections: The first and fifth sections form the pretest and posttest. The second section incorporates the repetition chapters, where learners can freely repeat the prerequisite concepts for ion bonding (chapters: ion, noble gas configuration, and electron transfer). The third section includes chapters that achieve the minimum learning goal (chapters: phenomenon brittleness, ion bonding, and ratio formula), and the fourth section includes chapters that enable the achievement of an extended learning goal (chapters: extended ratio formulas, nomenclature, and salts in everyday life). Learners can choose the chapters voluntarily and in any order in the second and fourth sections. The learning chapters in the third section have a linear structure and, consequently, a restriction on the free chapter selection. This ensures that all learners receive sufficient guidance in this central part to achieve the minimum learning goal. In addition to these macro choices, learners can also make micro choices about their learning path within the chapters. This includes choosing between text- or video-based resources or tasks with or without assistance. In addition to the described learning platform, learners can find advanced organizers, further learning tasks without separate assistance, and other typical science educational offerings such as the learning opportunity of developing hypotheses. I$_3$Learn offers several million possible learning paths because of the various decision options available.

Moreover, the individual chapters and subchapters have been organized and assigned to so-called Knowledge Components (KC).[30] The seven KCs in I$_3$Learn were engineered based on well-researched student conceptions of ion formation and bonding[31−38] (see Table 1). In addition, the curriculum for integrated comprehensive schools in Lower Saxony, Germany, where I$_3$Learn was implemented, was used to construct the KCs by considering the described competencies for ion bonding.[39] Each chapter or page in a chapter can be assigned to one KC. The dedicated separation and mapping of the KCs within I$_3$Learn enable a more precise analysis of the learning paths, knowledge gains, and the modeling of students. For example, student models would

Table 1. Knowledge Components (KCs) Integrated into I$_3$Learn[31−38,42]

Knowledge Component	Description	Learner's (alternative) conceptions	Learning goals according to the curriculum[42]
KC 1	Structure of atoms and ions	Terminology problems;[34] no distinction between metals and nonmetals[31]	Learners explain the structure of atoms and ions with Bohr's atomic model
KC 2	Formation of ions as electron transfer	Electrons are shared;[31,37,38] ions occur as molecules;[31,34−36] ion bonding are covalent;[35] reverse electron transfer;[32] attraction between atoms[32]	Students explain the formation of ion bonds from elements as an electron transfer reaction; learners justify the formation of ions with the noble gas rule; learners distinguish between ionic and covalent bonding
KC 3	Characteristics salts	–	Learners explain the properties of ionic bonds using a bonding model
KC 4	Ratio formulas of salts	–	Learners represent chemical issues in appropriate formula notation (ratio formulas)
KC 5	Bonding as an attraction between ions	No difference between the formation and bonding of ions;[34] ions are shared or transferred;[31] bonding types can only be named if they are labeled or the charge is known[31]	Learners explain the cohesion of ions in salts using the effect of electrical interactions
KC 6	The lattice structure of salts	Lack of understanding of three-dimensional structures;[35] ion molecules form a lattice;[31] similar distance on figures indicate the same type of bonding;[31] bindings are represented by lines[31]	Learners visualize the structure of bonds using suitable visual models
KC 7	Differentiation between macroscopic and submicroscopic level	Difficulty in distinguishing levels of representation[33]	–

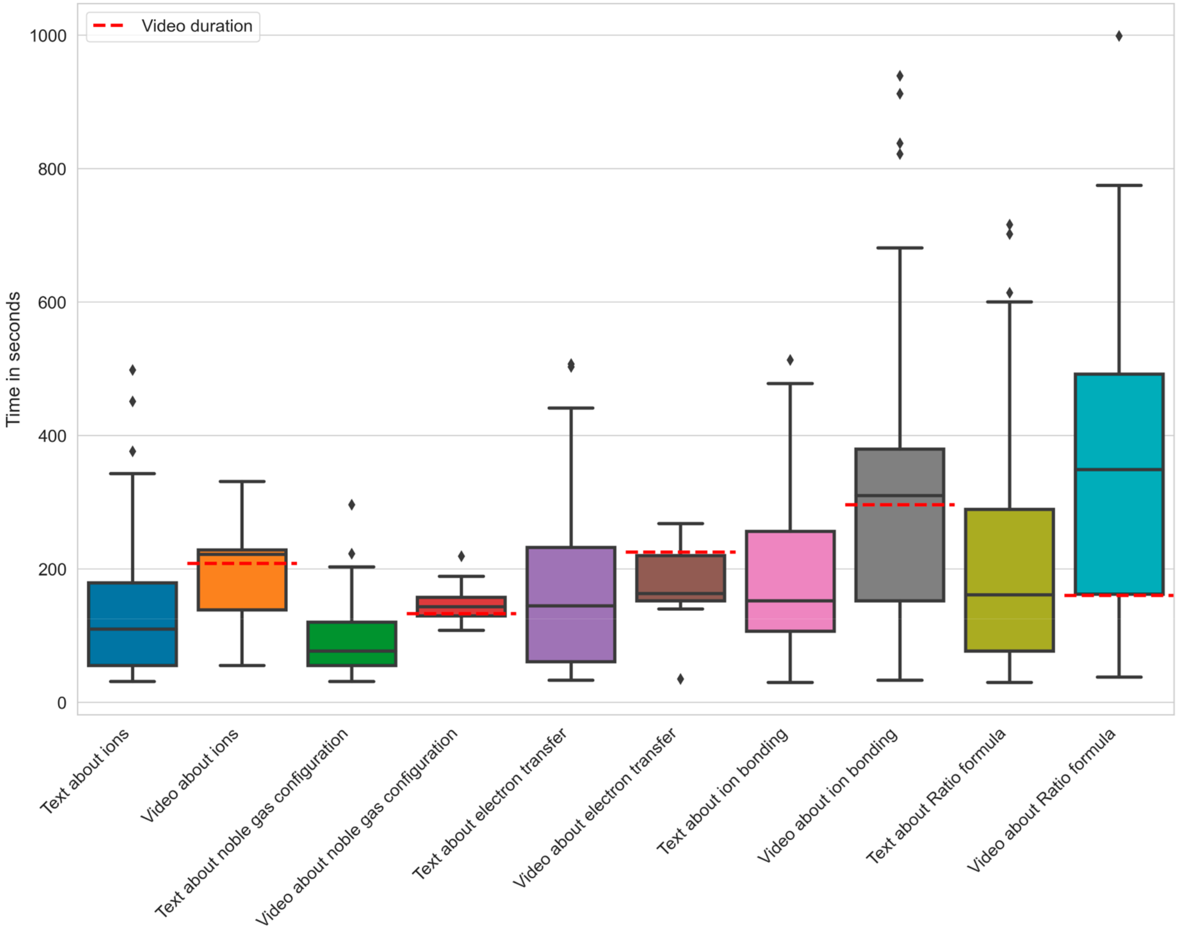


Figure 2. Boxplots of the average duration and distribution of the usage time of texts and videos in I_3Learn. The red dotted line beside the video's boxplots shows the corresponding video's total duration.

allow for the integration of a cognitive tutor, which predicts the probability for a learner whether the next learning task can be solved (e.g., Bayesian Knowledge Training).[30,40] In this way, tasks at different levels of understanding can be provided individually.[41]

Furthermore, I_3Learn was developed to focus on the engagement of all learners and the integration of the Universal Design for Learning (UDL) guidelines.[43] The guidelines are exemplified in the selection option between text and video (implementation of the UDL guideline 1, "Provide options for perception").[43] In total, the following elements were implemented in I_3Learn based on the UDL guidelines: multiple representations (text and video), learning tasks with and without help, self-assessment, customizable interface (font, font size, and contrast), read-aloud function, glossary, simple language, device independence, free chapter selection, everyday phenomenon, and advanced organizers. A demo version of an English I_3Learn version can be viewed by the following link. No registration is necessary, but anonymized usage data is collected but cannot be assigned to an individual: http://i3lern.idn.uni-hannover.de/en/jce/.

Using I_3Learn Data for Learning Analytics

In the period from January to June 2022, a total of 580 9th and 10th graders worked with I_3Learn. Teachers could choose how to integrate I_3Learn into their lessons to ensure natural integration, consider external factors such as the COVID-19 pandemic, and address the typical characteristics of different types of public schools. Half a million data points were collected, providing insight into the behavior of the learners. Each log file contains records such as pseudonymized learner IDs, timestamps, types of interactions, and details about the interaction. Thus, it is possible to reconstruct in detail which learners watched videos, read texts, or worked on tasks for how long. The data include interaction data, such as counts or duration of interactions, and personal data, such as self-concept in chemistry or reading ability. We provide a detailed description of all data collected in another publication (Roski et al.[44]). Data collection was approved by the Ministry of Science and Culture, Lower Saxony, Germany.

Furthermore, in the context of the pretest/post-test and learning tasks within the chapters, it is possible to reconstruct precisely when the learners saw and answered the question. Additionally, the system logs the learners' responses to questions and when they adapt them over time. The KC-based structure further makes it possible to filter the learners' interactions and focus on specific content. The data allows statistical and artificial intelligence-based methods to gain insights into the learners' usage behavior of a web-based learning platform. To date, the following use cases of the acquired data have been explored: (a) predicting dropout,[44] (b) identifying patterns of usage behavior in UDL-based learning opportunities,[10] or (c) using Bayesian Knowledge Tracing to predict learners' success with learning tasks within the KCs.[40]

https://doi.org/10.1021/acs.jchemed.3c00794
J. Chem. Educ. 2024, 101, 930−940

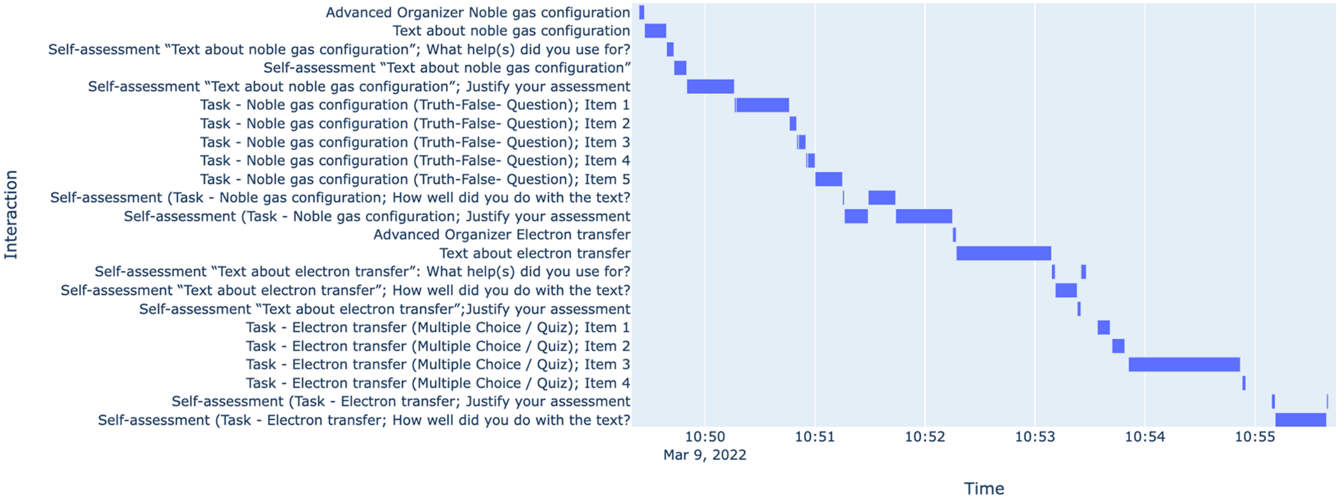


Figure 3. Learning progression of learner A in the repetition chapters. The blue bars show the usage time (x-axis) of the different content elements (y-axis).

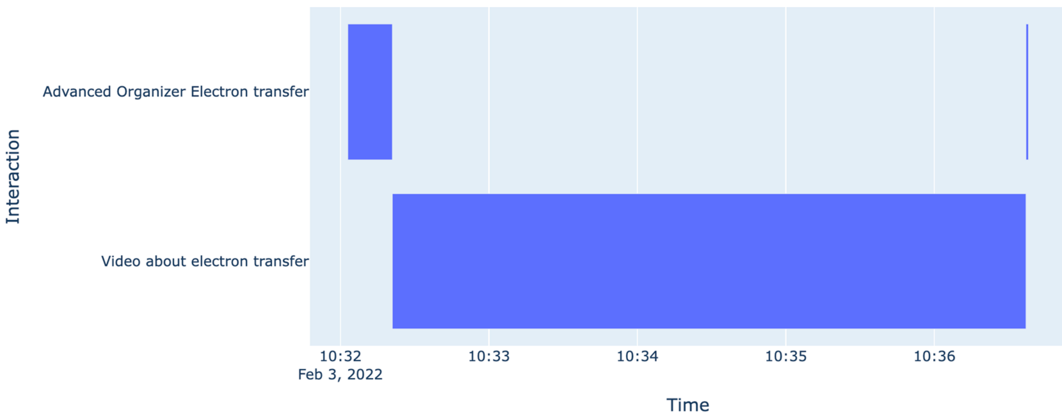


Figure 4. Learning progression of learner B in the repetition chapters. The blue bars show the usage time (x-axis) of the different content elements (y-axis).

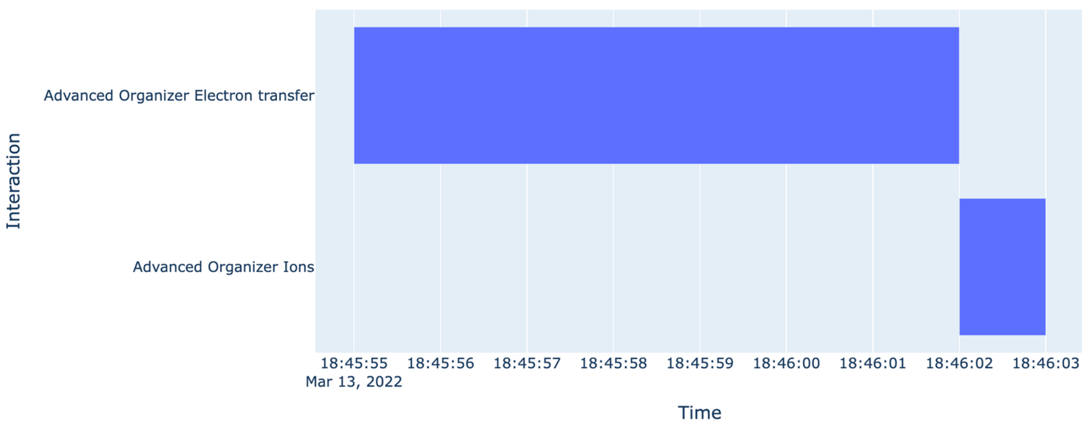


Figure 5. Learning progression of learner C in the repetition chapters. The blue bars show the usage time (x-axis) of the different content elements (y-axis).

The following are examples of how learning analytics can be used to analyze the usage of I_3Learn. The first example focuses on average usage behavior, while the second focuses on individual analysis.

Analyzing Usages of Multiple Representations

Learners can decide at five points in I_3Learn whether to watch a video or read a text. The content of the texts and videos is based on "ions", "noble gas configuration", "electron transfer", "ion bonding", and "ratio formulas". Texts and videos are the same in terms of content. The spoken content in the video is the slowly recited text. The contents relate to the chemistry curriculum of the federal state of Lower Saxony, Germany.[42] The duration of use can be examined, and the behavior of the learners can be analyzed. As shown in Figure 2, the duration of

https://doi.org/10.1021/acs.jchemed.3c00794
J. Chem. Educ. 2024, 101, 930−940

use of text and video of the same content differs considerably. In terms of content, the video and text do not differ. The text is delivered in slow speech in the video, and the same images and animations are used.

However, there are some peculiarities concerning the duration of the video use. In the three chapters of the repetition option (ion, noble gas configuration, and electron transfer; see Figure 1), many learners spend less than the total duration time in the video for ions and electron transfer, whereas, in noble gas configuration, many learners watch the whole video. The situation is also different for ratio formulas, the last chapter of the common learning goal (Figure 1). There, the actual duration of use considerably exceeds the video length. The data only include learners who have viewed the content for more than 30 s, as it can be assumed that a usage time below this is insufficient to perceive the content adequately. This finding can serve as an indication for further research. For one thing, the learning materials offered could evaluate whether learners have difficulties with the video on ratio formulas or whether the topic itself is a hurdle and requires more attention. Furthermore, the individual duration of use of the video could be considered and serve as a starting point for offering individual help if the duration of viewing goes far beyond the duration of the video.

Analyzing Individual Learners

In addition, usage behavior can also be considered at the level of individual learners. Figure 3 displays the learning behavior of learner A within the repetition chapters. The diagram shows how the learner engages with different content elements (*y*-axis) over time (*x*-axis). The log files also provide information about the learner's responses, which are not included here.

Looking at learners B (Figure 4) and C (Figure 5) in detail, it is immediately apparent that they are not engaged with the repetition chapters. Learner B clicks on the "Advanced Organizer: Electron Transfer" and watches the video for about 4 min (video duration 3:45 min), which comes after the advanced organizer. A short re-entry on the "Advanced Organizer: Electron Transfer" is in the log files afterward because the learner probably used the "back" function of the browser to go back to the overview page and thus generated the new entry. The learner did not use the learning tasks within the chapter or work on other chapters (see Figure 4).

While learner B is still watching a video in full, learner C spends only a few seconds on the review chapters (Figure 5). First, the learner views "Advanced Organizer: Electron Transfer" for a few seconds, then "Advanced Organizer: Ion". There is no interaction with any other content of the repetition chapters. Furthermore, a particular focus should be put on the *x*-axis. While learner B (Figure 4) spends several minutes in the individual parts of I_3Lern, learner C (Figure 5) engages for only a few seconds.

Summary of Applying Learning Analytics with the Data from I_3Learn

The shown global and individual examples of the usage behavior of I_3Learn illustrate some possibilities that go along with the detailed log file collection. The results primarily describe examples for a statistical evaluation of the raw data and still need to correspond to the comprehensive evaluation based on the understanding of learning analytics. Using further analysis approaches, it is possible to document individual learning paths, as well as the entire behavior of a cohort. This, on the one hand, allows for statistical analysis, like identifying "successful" learning paths, but, on the other hand, creates the basis for artificial intelligence-based tools. Individual support can be offered to learners who do not take advantage of the repetition chapters but show weaknesses in the subject knowledge test (pretest) and the learning tasks. A recommendation system could be used to make concrete suggestions for how they should continue learning. Also, "gaming the system" behavior,[45] as in the example of learner C, can be detected based on the data and appropriate individual support can be initiated. In addition, unsupervised machine learning approaches can cluster learner groups with similar behavioral patterns, or successful learning paths can be identified for groups with different prerequisites and abilities. In addition, a dashboard can be used to reflect usage by both the learner and the teacher. The possibilities are extensive because I_3Learn's integrated infrastructure can capture many learner interactions. Furthermore, learning barriers within the platform can also be identified if, for example, learners interact conspicuously with certain content parts.

STEP-BY-STEP OVERVIEW FOR USING WORDPRESS

The following is a step-by-step overview of creating a web-based learning platform using WordPress, such as I_3Learn. The essential steps are described as a starting guideline. The first step describes general requirements for a WordPress Web site and is thus universal and not specific to chemistry or science education.

Step One: Setting up the Infrastructure

First, the technical setup is established to make the Web site publicly available, involving two essential steps: first, we initialized a server to host the learning platform, and second, we installed WordPress on this server. Two options are available: Researchers can either delegate the setup, operation, and installation to a service provider or manually handle everything themselves. Opting for a service provider offers the advantage of receiving a finished product for a low monthly fee (ca. 25 Euro/USD per month[46]). However, limited access to certain WordPress features may be a disadvantage, as providers often modify their products. This service's target audience is not research projects but small businesses seeking a simple web presence. In addition, the server would not be physically accessible, which is a disadvantage with regard to data protection.

Furthermore, a disadvantage is that research data are stored on external servers. Implementing the server and the WordPress installation on the research institution hardware is recommended, allowing sensitive data to remain physically in the research project.

On the other hand, setting up a server by nonexperts is the biggest hurdle in creating a web-based learning platform, but it is doable. One of the best ways is to install a so-called LAMP stack is to install a Linux-based system. The letters stand for Linux, Apache (Web server), MySQL (database), and PHP (programming language). Afterward, WordPress can be installed.[47] After the installation and successful login with the admin account created during the installation process, the backend of the WordPress site can be accessed. This is a crucial control center and dashboard for a web-based learning platform. From here, content is created, learner accounts are moderated, and research data are accessed. In addition, it is recommended to set up a Secure Socket Layer (SSL) to ensure

Table 2. Essential Functions That the First Plugins Should Fulfill

Necessary function	Explanation
Backup	Allows a backup of the learning platform.
Hide admin login	Hiding the admin (account with access to the WordPress backend) login is a security feature, as it makes it more difficult for attackers to brute force the password and thus gain access to the learning platform.
Manage Learners	A plugin for managing learners' membership. This includes registration, logging in, and, if necessary, allowing learners to unlock individual sections of the learning platform only after achieving specific learning goals.
Removing the blog function	WordPress primary function is to offer a blog. This differs from a page-based Web site (learning platform) in that new articles (blogs) are published and presented chronologically on the main page. However, this does not correspond to the intended structure of a learning platform. This plugin turns off the blog function.
Removing the toolbar	The toolbar gives simplified access to the WordPress backend when the users view the so-called frontend, i.e., the learning platform from the learners' point of view. Learners need this toolbar and should not have access to the backend.

that the data exchange between the learner's device and the server occurs via an encrypted connection.

Step Two: Creating the Curricular Architecture

The second step is to create a learning platform. First, researchers must be clear about the structure and purpose of the learning platform. The following guiding questions can be decisive: What learning goals are related to the learning platform? Should the learning platform be structured linearly, or will learners always return to an overview page from which the following sections can be selected? Should the learners document experiments? Should students engage in inquiry learning? Should researchers develop and test hypotheses? Should the learning platform supplement classroom teaching or only remote learning? Should there be sections that the learners cannot select at first or that they should work on first, such as a pretest? Should the learning path be split up again within individual chapters, for example, if an alternative learning video is also offered for the text? In addition, structuring with the help of KCs can also be considered at this point.[30]

Creating an architecture of the intended learning platform is advisable for use as an overview. In implementing I_3Learn, designed with four lessons of 45 min each, more than 198 individual pages were created, which in their entirety form the learning platform.

Before the page design can begin, a so-called "theme" that significantly impacts the global design should be selected. WordPress and third parties provide over 11,000 themes,[48] which primarily determine the design of the starting page and the appearance of the header of the learning platform. The theme can be changed in the WordPress backend via the tab "Design". However, this is based on visual preferences alone.

Furthermore, the first must-have plugins can be installed during this step. These plugins have no direct influence on the learning experience but serve to ensure essential functions. A selection of the first necessary plugins can be found in Table 2. A plugin can be installed in the WordPress backend via "Plugins" and then "Install Plugins". An entire directory of all 60,000 freely available plugins[49] can be downloaded and installed immediately. Usually, these plugins are free of charge. However, some of the features include premium features that must be purchased separately. No specific plugins are recommended in the following but only the tasks they should fulfill. With the search function of WordPress, it is easy to find suitable plugins.

Step Three: Implement the Learning Platform

The page design can be done via the WordPress backend via the tab "Pages". Each page can be provided with its Uniform Resource Locator (URL), which is helpful to connect the individual pages, for example, by the buttons mentioned. The page design is usually done in a drag-and-drop manner so that the page design is comparatively simple; that is, no programming knowledge is required for the visual design of the learning platform. The graphical interface and the corresponding integrated instructions help beginners create a modern and appealing design for learning platforms. It is very similar to the formatting of a document in word processing software. It is important to note that changes to the start page are not made via the "Pages" tab, although the start page is also stored there, but via the "Design" tab. Otherwise, the layout of the homepage will be messed up. The built-in function of WordPress can be used to design individual pages of the learning platform. In addition, numerous other plugins extend the page building with additional functions. A recommended function is the possibility to hide content. This is used in the plugin to make the design responsive and to allow the content to be used on different devices (computer, tablet, and mobile phone). It can also be used to hide the content on all devices. The use of this is discussed below.

Step Four: Prepare the Data Collection

One of the essential points in designing the learning platform is the ability to collect data. For the integration of tasks, the plugin "H5P"[50] was used in I_3Learn, which allows flexible integration of different task types and formats. The H5P plugin is essential for data collection. H5P is an open-source application, with which tasks and queries can be carried out. The advantage is that all interactions (e.g., clicking) with content built with H5P are saved. This includes not only the final answer but also all previous ones. It does not matter whether the learner adds or removes content from the open or closed answers in the learning process. Everything is stored in the log files. It is also recorded when the learners can see the H5P content so that the time between "seen" and "answered" can be reconstructed. "Seen" in this context means the page has been accessed in the learning platform. It can be assumed that learners begin interacting with the page's content.

For this reason, it is essential that a page is not overloaded with many tasks and texts but is well separated on different pages. This enables a more precise tracking of the learning behavior. For example, if the content contains only one task, it is possible to track exactly when it was accessed, answered, and left, because accessing the next task on the next page is also recorded. A selection of practical H5P elements can be seen in Table 3. Furthermore, not only tasks can be created via H5P, but also interactive videos or documentation tools. I_3Learn integrated multiple choice questions, tasks with open answer options, cloze tests, drag and drop tasks (arranging an ion grid), or even game-based elements such as quizzes.

Table 3. Selection of H5P Elements to Be Integrated into a Learning Platform

H5P element	Explanation	Example usage in I_3Learn
Accordion	A standard interface allowing users to hide and reveal pieces of information	Used as a glossary to provide important terms on ions bonding for learners. Also well suited to be used as a "tracking pixel".
Documentation Tool	Learners develop hypotheses and can confirm or falsify them	Learners hypothesize about a virtual experiment on the brittleness of salt crystals subjected to a force applied by a hammer. They can confirm or refute these in the course of the chapter.
Drag and Drop	Arrangement of objects by drag and drop. Possibility to arrange chemical issues two-dimensionally. "Drag the word" is the alternative, but with words.	Learners arrange anions and cations into an ion lattice.
Fill in the Blanks	Allows inserting free text into existing sentence structures. A certain amount of spelling errors is tolerated.	Integrated questions that require only a small amount of open-ended response from the learner, such as learning tasks on ratio formulas
Interactive Book	Allows the design of smaller learning environments within an H5P integration. Other H5P elements can be integrated here.	It has been used in I_3Learn when learners are to study salt crystals or the phenomenon of brittleness on their own. The interactive book offers various information, such as illustrations. Learners can explore the material provided in their way.
Interactive Video	Other H5P elements, such as multiple-choice questions, can be placed within the video. Allows features like bookmarks, fast forward and rewind of 10 s, slowing down the video, and integrating subtitles.	In I_3Learn, videos were deployed using this H5P element, and features such as slowing down the speed of the video were implemented to implement UDL guidelines.
Single Choice Set	Multiple choice questions with an increased gaming character.	Learning tasks for the learners within the review chapters allow them to check their knowledge level independently.
Questionnaire	Allows open and closed response formats without providing learners with rule-based feedback. Requires a Learning Record Storage (LRS) or LRS-like environment, such as the "H5PxAPIkatchu" plugin (see Step Four: Prepare the Data Collection).	To collect pre- and post-test data. In I_3Learn, demographic data, self-concept and interest in chemistry, subject knowledge, reading ability, and cognitive ability of learners were collected.

The texts of a learning platform can also be integrated into the learning platform independently of H5P. This can quickly be done via the page editor. This has the advantage that with the help of other plugins, the texts can be individualized or text-to-speech functions can be implemented. Working with texts or generally pages where no H5P tasks are integrated can still be tracked. For this, the H5P element "accordion" can be integrated on a page but hidden with the help of the mentioned page builder plugin. This way, the page access is still recorded in the log files. This approach is like the so-called "tracking pixels", often used in online marketing.[51] According to this principle, all of the pages of the learning platform should contain at least one H5P element, either in the form of a task (or a similar element that learners should interact with) or as a hidden element so that the opening of each page is captured. This enables a seamless recording of the learning paths and the time spent in the learning platform.

Step Five: Analyzing the Collected Data

The data collection provides log files in xAPI[52] format. xAPI is a well-known structured data format from learning analytics.[53] The log files can be accessed in two ways: The first consists of using a Learning Record Storage (LRS) offered by various providers. These are primarily cloud-based offerings. An alternative is using the plugin "H5PxAPIkatchu".[54] This is not a full-fledged alternative to an LRS but is feasible for data collection like I_3Learn. This plugin can provide the log files as CSV files. A corresponding analysis can then be done with typical spreadsheet applications. Depending on the size of the log files, it may be helpful to use the software library Pandas.[55] Here, knowledge of Python is necessary. However, it is also possible to perform statistical and artificial intelligence-based analysis with programs with a graphical interface and, therefore, does not require any programming skills. An overview of programs for learning analytics and artificial intelligence is given by Slater et al.[56]

In order to understand the data and make it understandable for other researchers, it is recommended to create a "data manual" (see an example in Figure 6): After finalizing the learning platform, a test run should be made and the log files should be examined after every few interactions. This ensures, on one hand, that all interactions to be recorded are also recorded and, on the other hand, that all interactions can also be explained. Each H5P element integrated into the learning platform is assigned a specific "Content ID", which can be determined via this test run and documented in the data manual.

■ CONCLUSIONS AND LIMITATIONS

This paper presents a step-by-step overview of how a web-based learning platform with interaction data tracking can be realized without programming skills. We have contributed to interested chemistry education researchers a method for setting up an infrastructure and a feasible way to design a learning platform from the ground up to meet high educational standards while collecting fine-grained log file data in a standard data format. In addition, we demonstrated successful data mining with the web-based learning platform I_3Learn.

Integrating learning analytics in chemistry education requires extensive data.[57] The chemistry education research community can collect rich and comprehensive data that can potentially analyze and optimize teaching−learning processes using the described approach. The absence of publicly available

Pre-Test (What I_3Learn needs to learn about you)

Page	Content on Page	Description	xAPI Verb	Answer options	H5P Content ID or Subcontent ID
General questions; Full page with info text and questions		Only viewed	Consumed		156
	Question: What grade are you?	Only viewed	Attempted	[0] 9th year, [1] 10th year	bc8c80d2-b122-41c6-82ed-8fbb67b529d5
		Answered	Interacted		
	Question: What language do you speak most often at home?	Only viewed	Attempted	[0] German, [1] other language	6cd9c77d-f577-41ae-9652-3e27f05ac17b
		Answered	Interacted		

Figure 6. Excerpt of the data manual for I_3Learn. At the beginning of the pretest, demographic information is requested. The manual allows the log file entries in the raw data set to be identified because (a) each question has a unique "Content ID", (b) each entry is tagged with an xAPI verb that indicates the interaction made (column "Description"), and (c) the learner's answer in the form of numbers which are explained by the manual.

and high-quality data sets, especially in chemistry education, makes it necessary to collect data about learning processes independently.[16,58,59] The integration of learning analytics includes statistical analysis of learning and usage behavior, such as using text and videos within the learning platform. It also enables the use of artificial intelligence, e.g., to predict dropout (Roski et al.[44]) or group (cluster) students according to their learning behavior (Roski et al.[10]). In order to ensure clarity and the actual intention of this step-by-step overview, a more detailed data analysis was omitted. In addition, the creation of larger data sets allows researchers to address further science education questions. For example, based on the aforementioned insights from the I_3Learn data, it is possible to investigate how the learning success of learners differs depending on the preference for video or text resources—but this is also beyond the scope of this article.

In our case study, I_3Learn has been established in the field of K12 education in the chemistry classroom for the topic of ion bonding. However, the described approach using WordPress also allows for integration in higher education, lifelong learning, in-service training in the vocational sector, and other (chemistry) domains.

In addition, three comments are necessary. First, the web-based platform was used only to mine educational data. The artificial intelligence-based applications described were trained separately and "offline". Currently, no artificial intelligence-based applications are integrated into the web-based learning platform. Second, when collecting fine-grade log file data from learners, ethical dimensions must be considered when collecting data.[60] The collection of second-by-second interaction data also allows for the dedicated monitoring of learners. Also, it allows learners to be assessed and seemingly punished for less-than-optimal behavior based on the data. Furthermore, the use of artificial intelligence methods can cause bias and fairness issues that need to be minimized.[61,62] Naturally, we have taken this into account in our approach (see Roski et al.[61]). Third, there are other methods of collecting similar data, for example, via existing LMSs. We are just presenting a more independent and open way.

■ AUTHOR INFORMATION

Corresponding Author

Marvin Roski − *Chemistry Education, Institute of Science Education, Leibniz University Hannover, Hannover 30167, Germany;* orcid.org/0000-0002-0160-2199; Email: roski@idn.uni-hannover.de

Authors

Ralph Ewerth − *TIB - Leibniz Information Centre for Science and Technology, Hannover 30167, Germany; L3S Research Center, Leibniz University Hannover, Hannover 30167, Germany*

Anett Hoppe − *TIB - Leibniz Information Centre for Science and Technology, Hannover 30167, Germany; L3S Research Center, Leibniz University Hannover, Hannover 30167, Germany*

Andreas Nehring − *Chemistry Education, Institute of Science Education, Leibniz University Hannover, Hannover 30167, Germany;* orcid.org/0000-0002-8723-5552

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.3c00794

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This paper results from the PhD program LernMINT, funded by the Ministry of Science and Culture, Lower Saxony, Germany. We state explicitly that we are not affiliated with WordPress or any private person or company that provides plugins and other software. Furthermore, the authors do not receive any financial benefit from WordPress, creators of plugins, or companies of the mentioned software. The authors themselves covered any costs.

■ REFERENCES

(1) Drachsler, H.; Kalz, M. The MOOC and Learning Analytics Innovation Cycle (MOLAC): A Reflective Summary of Ongoing Research and Its Challenges. *J. Comput. Assist Learn* **2016**, *32* (3), 281−290.

(2) Korkmaz, C.; Correia, A.-P. A Review of Research on Machine Learning in Educational Technology. *EMI Educ Media Int.* **2019**, *56* (3), 250−267.

(3) Lang, C.; Siemens, G.; Wise, A.; Gasevic, D. *Handbook of Learning Analytics* **2017**, DOI: 10.18608/hla17.

(4) Wilson, C. D.; Haudek, K. C.; Osborne, J. F.; Buck Bracey, Z. E.; Cheuk, T.; Donovan, B. M.; Stuhlsatz, M. A. M.; Santiago, M. M.; Zhai, X. Using Automated Analysis to Assess Middle School Students' Competence with Scientific Argumentation. *J. Res. Sci. Teach* **2024**, *61*, 38.

(5) Link, R. D.; Gallardo-Williams, M. We Should Keep Developing Digital Laboratory Resources in the Postpandemic Era. *J. Chem. Educ.* **2022**, *99* (2), 519−520.

(6) Khouri, N. G.; Fontana, M.; Dias, I. L. R.; MacIel, M. R. W.; MacIel Filho, R.; Mariano, A. P. Chemical Engineering Teaching in COVID-19 Times: Successfully Adapting a Capstone Design Course to a Remote Format. *J. Chem. Educ.* **2021**, *98* (12), 3794−3803.
(7) Roski, M.; Walkowiak, M.; Nehring, A. Universal Design for Learning: The More, the Better? *Educ Sci. (Basel)* **2021**, *11* (4), 164.
(8) Costa, D. R.; Chen, C. W. Exploring the Relationship between Process Data and Contextual Variables among Scandinavian Students on PISA 2012 Mathematics Tasks. *Large Scale Assess Educ* **2023**, *11* (1), 5 DOI: 10.1186/s40536-023-00155-x.
(9) Wilson-Kennedy, Z. S.; Winfield, L. L.; Nielson, J.; Arriaga, E. A.; Kimble-Hill, A. C.; Payton-Stewart, F. Introducing the Journal of Chemical Education's Special Issue on Diversity, Equity, Inclusion, and Respect in Chemistry Education Research and Practice. *J. Chem. Educ.* **2022**, *99* (1), 1−4.
(10) Roski, M.; Sebastian, R.; Ewerth, R.; Hoppe, A.; Nehring, A. Learning Analytics and the Universal Design for Learning (UDL): A Clustering Approach. *Accepted with major revsions*.
(11) Stinken-Rösner, L.; Rott, L.; Hundertmark, S.; Menthe, J.; Hoffmann, T.; Nehring, A.; Abels, S. Thinking Inclusive Science Education from Two Perspectives: Inclusive Pedagogy and Science Education. *Ristal* **2020**, *3*, 30−45.
(12) Bowen, R. S.; Cooper, M. M. Grading on a Curve as a Systemic Issue of Equity in Chemistry Education. *J. Chem. Educ.* **2022**, *99* (1), 185−194.
(13) Ryu, M.; Bano, R.; Wu, Q. Where Does CER Stand on Diversity, Equity, and Inclusion? Insights from a Literature Review. *J. Chem. Educ.* **2021**, *98* (12), 3621−3632.
(14) Baker, R.; Siemens, G. Educational Data Mining and Learning Analytics. *The Cambridge Handbook of the Learning Sciences* **2014**, 253 https://learn.canvas.net/courses/33.
(15) Baker, R. Using Learning Analytics in Personalized Learning. In *Handbook on personalized learning for states, districts, and schools*; Murphy, M., Redding, S., Twyman, J., Eds.; **2016**; pp 165−174.
(16) Dietze, S.; Siemens, G.; Taibi, D.; Drachsler, H. Editorial: Datasets for Learning Analytics. *Journal of Learning Analytics* **2016**, *3* (2), 307−311.
(17) Fischer, C.; Pardos, Z. A.; Baker, R. S.; Williams, J. J.; Smyth, P.; Yu, R.; Slater, S.; Baker, R.; Warschauer, M. Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education* **2020**, *44* (1), 130−160.
(18) WordPress. https://wordpress.org (accessed 2023-08-01).
(19) Ifenthaler, D.; Drachsler, H. Learning Analytics. In *Handbuch Bildungstechnologie*; Springer: Berlin, 2020; pp 515−534. DOI: 10.1007/978-3-662-54368-9_
(20) Phil Long, B.; Siemens, G. Penetrating the Fog: Analytics in Learning and Education. *Educause Review* **2011**, *46* (5), 31.
(21) Patricia Aguilera-Hermida, A. College Students' Use and Acceptance of Emergency Online Learning Due to COVID-19. *Int. J. Educ Res. Open* **2020**, *1*, 100011.
(22) *About WordPress*. https://wordpress.org/about/ (accessed 2023-04-13).
(23) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation); 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (accessed 2023-10-12).
(24) moodle. https://moodle.com (accessed 2023-08-01).
(25) Mohd Kasim, N. N.; Khalid, F. Choosing the Right Learning Management System (LMS) for the Higher Education Institution Context: A Systematic Review. *Int. J. Emerg Technol. Learn* **2016**, *11* (6), 55−61.
(26) Moodle: Making quality online education accessible for all. https://moodle.com/about/ (accessed 2023-08-01).
(27) *Perusall*. https://www.perusall.com (accessed 2023-10-11).
(28) Hilbing, C.; Barke, H.-D. Ionen Und Ionenbindung: Fehlvorstellungen Hausgemacht! Ergebnisse Empirischer Erhebungen Und Unterrichtliche Konsequenzen. *CHEMKON* **2004**, *11* (3), 115−120.
(29) Hall, T.; Vue, G.; Strangman, N.; Meyer, A. *Differentiated Instruction and Implications for UDL Implementation Effective Classroom Practices Report Differentiated Instruction and Implications for UDL Implementation*; 2004; Vol. *3*.
(30) Pelánek, R. Bayesian Knowledge Tracing, Logistic Models, and beyond: An Overview of Learner Modeling Techniques. *User Model User-adapt Interact* **2017**, *27* (3−5), 313−350.
(31) Luxford, C. J.; Bretz, S. L. Development of the Bonding Representations Inventory to Identify Student Misconceptions about Covalent and Ionic Bonding Representations. *J. Chem. Educ.* **2014**, *91* (3), 312−320.
(32) Vrabec, M.; Prokša, M. Identifying Misconceptions Related to Chemical Bonding Concepts in the Slovak School System Using the Bonding Representations Inventory as a Diagnostic Tool. *J. Chem. Educ.* **2016**, *93* (8), 1364−1370.
(33) Meltafina, M.; Wiji, W.; Mulyani, S. Misconceptions and Threshold Concepts in Chemical Bonding. *J. Phys. Conf Ser.* **2019**, *1157* (4), No. 042030.
(34) Barke, H.-D. *Chemiedidaktik Diagnose Und Korrektur von Schülervorstellung*; Springer, 2006.
(35) Butts, B.; Smith, R. HSC Chemistry Students' Understanding of the Structure and Properties of Molecular and Ionic Compounds. *Res. Sci. Educ* **1987**, *17* (1), 192−201.
(36) Tan, K.-C. D.; Treagust, D. F. *Evaluating Students' Understanding of Chemical Bonding*; 1999.
(37) Nicoll, G. A Report of Undergraduates' Bonding Misconceptions. *Int. J. Sci. Educ* **2001**, *23* (7), 707−730.
(38) Coll, R. K.; Taylor, N. Mental Models in Chemistry: Senior Chemistry Students Mental Models of Chemical Bonding. *Chem. Educ. Res. Pract.* **2002**, *3* (2), 175−184.
(39) Niedersächisches Kultusministerium. *Die Arbeit in Den Schuljahrgängen 5 Bis 10 Der Integrierten Gesamtschule (IGS)*; 2021.
(40) Roski, M.; Hoppe, A.; Nehring, A. Individuelles Lernen Durch Bayesian Knowledge Tracing in Der Webbasierte Lernplattform "I3Lern" Analysieren Und Unterstützen. In *Lernen, Lehren und Forschen in einer digital geprägten Welt*; Gesellschaft für Didaktik der Chemie und Physik Jahrestagung 2022, **2023**.
(41) Sao Pedro, M. A.; Baker, R. S. J. d.; Gobert, J. D. Incorporating Scaffolding and Tutor Context into Bayesian Knowledge Tracing to Predict Inquiry Skill Acquisition. *Proceedings of the 6th International Conference on Educational Data Mining, EDM 2013* **2013**.
(42) Niedersächsiches Kultusministerium. *Kerncurriculum Für Die Integrierte Gesamtschule Schuljahrgänge 5−10. Naturwissenschaften*; 2012.
(43) CAST. *Universal Design for Learning Guidelines version 2.2*.
(44) Roski, M.; Sebastian, R.; Ewerth, R.; Hoppe, A.; Nehring, A. Dropout Prediction in a Web Environment Based on Universal Design for Learning. In *Artificial Intelligence in Education*; Wang, N., Rebolledo-Mendez, G., Matsuda, N., Santos, O. C., Dimitrova, V., Eds.; Springer Nature: Switzerland, Cham, 2023; pp 515−527. DOI: 10.1007/978-3-031-36272-9_42.
(45) Baker, R. S. J. D. *Gaming the System: A Retrospective Look*; 2011.
(46) wordpress.com Pricing. https://wordpress.com/pricing/ (accessed 2023-10-11).
(47) Get WordPress. https://wordpress.org/download/ (accessed 2023-08-01).
(48) WordPress: Themes. https://wordpress.org/themes/ (accessed 2023-08-01).
(49) WordPress: Plugins. https://wordpress.org/plugins/ (accessed 2023-08-01).
(50) *H5P*. https://h5p.org (accessed 2023-08-01).
(51) Fabian, B.; Bender, B.; Weimann, L. E-Mail Tracking in Online Marketing-Methods, Detection, and Usage. 2015.
(52) *xAPI.com*. https://xapi.com (accessed 2023-08-01).
(53) Berg, A.; Scheffel, M.; Drachsler, H.; Ternier, S.; Specht, M. The Dutch XAPI Experience. In *ACM International Conference*

Proceeding Series; Association for Computing Machinery, 2016; Vol. *25–29*-April-2016, pp 544–545. DOI: 10.1145/2883851.2883968.
(54) *SNORDIAN's H5PxAPIkatchu*. https://wordpress.org/plugins/h5pxapikatchu/ (accessed 2023-08-01).
(55) *pandas*. https://pandas.pydata.org (accessed 2023-08-01).
(56) Slater, S.; Joksimović, S.; Kovanovic, V.; Baker, R. S.; Gasevic, D. Tools for Educational Data Mining: A Review. *J. Educ Behav Stat* **2017**, *42* (1), 85–106.
(57) Siemens, G. Learning Analytics: The of a Discipline. *American Behavioral Scientist* **2013**, *57* (10), 1380–1400.
(58) Romero, C.; Ventura, S. Educational Data Mining and Learning Analytics: An Updated Survey. *Wiley Interdiscip Rev. Data Min Knowl Discov* **2020**, *10* (3), e1355 DOI: 10.1002/widm.1355.
(59) Verbert, K.; Manouselis, N.; Drachsler, H.; Duval, E. Dataset-Driven Research to Support Learning and Knowledge Analytics. *J. Educ Techno Soc.* **2012**, *15* (3), 133–148.
(60) Murchan, D.; Siddiq, F. A Call to Action: A Systematic Review of Ethical and Regulatory Issues in Using Process Data in Educational Assessment. *Large Scale Assess Educ* **2021**, *9* (1), 25 DOI: 10.1186/s40536-021-00115-3.
(61) Roski, M.; Hoppe, A.; Nehring, A. Supporting Inclusive Science Learning through Machine Learning: The AIISE Framework. In *Uses of Artificial Intelligence in STEM Education*; Zhai, X.; Krajcik, J., Eds; Oxford University Press, **2024**.
(62) Ntoutsi, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdl, W.; Vidal, M. E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; Kompatsiaris, I.; Kinder-Kurlanda, K.; Wagner, C.; Karimi, F.; Fernandez, M.; Alani, H.; Berendt, B.; Kruegel, T.; Heinze, C.; Broelemann, K.; Kasneci, G.; Tiropanis, T.; Staab, S. Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey. *Wiley Interdiscip Rev. Data Min Knowl Discov* **2020**, *10* (3), e1356 DOI: 10.1002/widm.1356.

https://doi.org/10.1021/acs.jchemed.3c00794
J. Chem. Educ. 2024, 101, 930–940