



Learning Analytics for Peer Assessment: A Scoping Review

2

Kamila Misiejuk and Barbara Wasson

2.1 Introduction

Learning analytics (LA) is a research field that focuses on analysing educational data, with the goal of understanding and/or improving learning. LA is identified as having the potential to change assessment practices and support “the holistic process of learning” (Ferguson et al., 2016). Knight (2020) argues that LA can be used to move the focus from the summative assessment of products produced to facilitate more process-oriented assessments. Similarly, Archer and Prinsloo (2020) write that LA supports the assessment of and for learning and can help in understanding student learning, analysing learning behavior, predicting student learning needs, and prescribing interventions that may promote more effective teaching and learning; however, the ethics of student surveillance and privacy issues must be considered.

Some LA researchers note that assessment data are not commonly considered “an integral part of the analytics data cycle,” but, rather, as an outcome measurement, which leads to assessment analytics being “still under-explored and largely under-developed” (Saqr, 2017, 1). Other reasons for not including assessment data in LA datasets are related to the strong emphasis on behavioural data rather than traditional assessment data, which may be more meaningful; the fact that LA is not led by pedagogy; and the fact that assessment data are not granular enough to allow

This research is supported by a PhD research grant from the University of Bergen, Norway.

K. Misiejuk (✉) · B. Wasson
Centre for the Science of Learning and Technology (SLATE), University of Bergen, Bergen,
Norway
e-mail: kamila.misiejuk@uib.no

Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

a detailed analysis of student behaviour (Ellis, 2013). There are also some concerns about implementing LA around removing human mentors from the feedback loop and students gaming the analytics (Buckingham Shum & Ferguson, 2012). On the other hand, the inclusion of assessment data, especially feedback data, has the potential to close the gap between data and education, increase LA usefulness, and broaden LA's scope (Ellis, 2013; Pardo, 2018; Saqr, 2017). Assessment data also have the advantage of being relatively easy to capture because students expect to be assessed based on their performance (Ellis, 2013). Knight (2020) highlights the fact that the “development of assessments based on novel process-based data is challenging (...) Thus, this development is likely to be time-consuming, expensive, and require systemic changes” (p. 133), and he also argues that the data should be used to support, not supplant, humans in their assessment practices.

Cope and Kalantzis (2016) mapped new assessment models that emerged alongside the increased prevalence of educational big data, including embedding assessment in learning, an increased focus on formative assessment, and a new conceptualization of summative assessment as a *progress view* rather than an *end view* of learning. Knight (2020) described three ways of transforming formative assessment with the help of LA: (1) developing new assessment techniques, (2) automating existing assessment techniques, or (3) augmenting existing assessment techniques. Moreover, he presented some potential augmentation scenarios, such as using LA to automatically allocate peers or automate feedback on the quality of the student feedback provided (backward evaluation).

One form of formative assessment is *peer assessment* (PA). Liu and Carless (2006) distinguish between *peer assessment* as “students grading the work or performance of their peers using relevant criteria” (p. 280) and *peer feedback* as “a communication process through which learners enter into dialogues related to performance and standards” (p. 280). In this chapter, we use PA as an umbrella term for all forms of PA, including peer feedback, peer grading, and peer review. Early LA research identified the potential of using LA techniques for constructionist learning activities, such as PA (Berland et al., 2014). Some potential practical implementations of LA in PA included feedback classification, a text analysis of rubric answers, combining peer and automated assessment, predicting the accuracy of peer raters, text analysis to monitor feedback quality and appropriateness, and clustering and visualisation techniques to optimise the feedback process (Ryan et al., 2019; Wahid et al., 2016).

2.2 Purpose of the Present Study

As the LA field is heading toward maturity, there is a need to examine how LA has been implemented in the field of PA. To date, there have been no literature reviews conducted on the broad topic of using LA in PA research, although there are reviews of LA and formative feedback (see Banihashem et al., 2022) and one review on an aspect of LA and PA (see Fig. 2.1). In a systematic literature review that included 28 papers, Nyland (2018) identified tools and techniques for

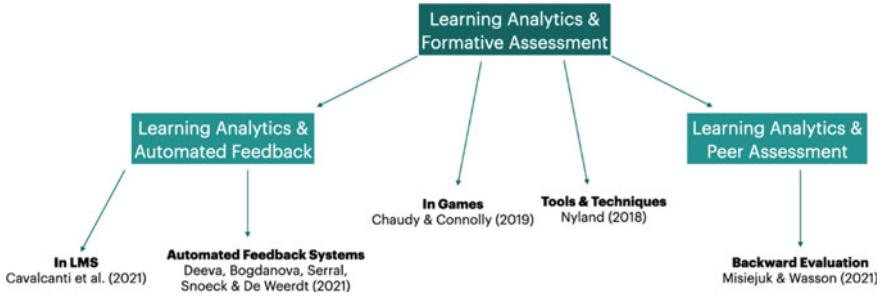


Fig. 2.1 Review studies of learning analytics and formative assessment

data-enabled formative assessment. Cavalcanti et al. (2021) conducted a systematic literature review that included 63 papers on automatic feedback generation in learning management systems. Chaudy and Connolly (2019) explored 34 relevant studies to identify the various approaches to integrating assessment in educational games and their associated empirical evidence. Deeva et al. (2021) classified and described 109 automated feedback systems. Misiejuk and Wasson (2021) focused on backward evaluation in PA, which is students receiving feedback on the quality of the feedback that they have given.

To fill this gap and help understand how LA is being used in PA, this chapter reports on a scoping review that focused on three research questions:

- (1) Where in the peer assessment process are the analytics employed? What is the role of learning analytics in peer assessment research?
- (2) What are the reported peer assessment challenges the research addressed with learning analytics? And how are they addressed?
- (3) What insights into peer assessment can we gain from learning analytics?

2.3 Methodology

2.3.1 Scoping Review

As no studies analysing the broad use of LA in PA research have been conducted, a scoping review exploring the “the breadth and depth of a field” is an appropriate method with which to close this gap (Levac et al., 2010, 1). In this study, the scoping review approach, as described by Levac et al. (2010) was used. This included discussions between two researchers on the inclusion/exclusion of some of the papers, an iterative process of refining the coding criteria and research questions, and a report on the methodological details of the scoping review process.

2.3.2 Search

The search was conducted in December 2021 and resulted in 1534 papers (duplicates removed), which were screened for inclusion over three rounds (see search details in Fig. 2.2). Papers not written in English, those that were not peer-reviewed, and those published before 2011—the year of the first Learning Analytics and Knowledge (LAK) Conference—were excluded. Due to a large number of papers found during the search, the first screening focused on a detection of the phrase “learning analytics” in the abstract, title, or keywords and full text of the papers; in this way, “learning analytics” served as a proxy for authors centering themselves in the field of LA. If “learning analytics” was only found in the references, the paper was excluded. Papers published at the LAK Conference or in the Journal of Learning Analytics were allowed to bypass this rule, with the assumption that publishing in these places automatically establishes a link to LA. After the first round, 598 papers remained. The second screening tackled the “peer assessment” aspect of the review by checking whether some form of PA was described in the methods section of the article. After two rounds of screenings, the full text of 166 papers was examined for their relevance to the research questions.

After the exclusion of the non-relevant papers, the final review included 27 papers: fourteen journal articles and 13 conference papers. While most had an

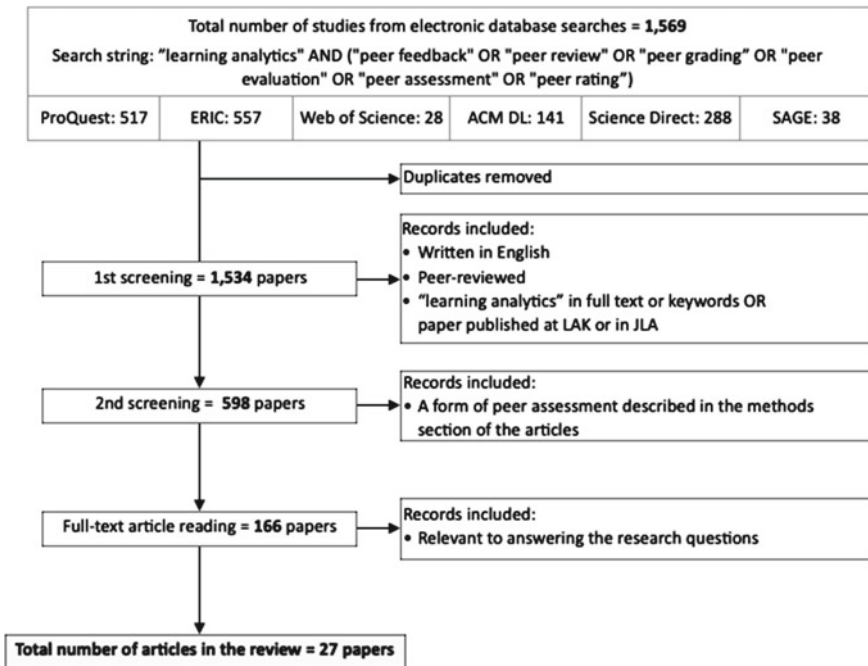


Fig. 2.2 Inclusion/exclusion process

overall focus on PA, for some, PA was secondary. For example, some papers used *PA data* for LA and delivered new insights into PA research, although the focus of the paper was not on PA. Most papers (22 of 27) conducted their studies in the context of higher education, except for Misiejuk et al. (2021), whose dataset included data from both higher education and K-12; Koh et al. (2016), Mørch et al. (2017), who used K-12 data; Hunt et al. (2021), who focused on professional development; and Babik et al. (2019), who simulated a dataset. Table 2.1 provides an overview of the 27 included papers.

2.4 Results

RQ1: Where in the peer assessment process are the analytics employed? What is the role of learning analytics in peer assessment research?

In the application of LA, eleven papers used LA to improve PA activity (*LA for PA*), while 15 papers used LA to analyse PA data (*LA on PA data*). We identified three main roles on the part of LA in improving PA: *tools*, *automated feedback*, and *visualizations*. For the papers that used LA to analyse PA data, four main application areas were mapped: *student interaction*, *feedback characteristics*, *comparison*, and *design*. Although some papers apply LA in more than one role, the paper categorization discussed below focused on the main use of LA in PA research described in the paper. Only one paper included both. Cheng and Lei (2021) both analysed PA data and developed visualizations for PA that showed students the social networks of their blogging and PA activities. Then, they examined the visualisation's influence on their engagement and group cohesion.

Tools. Four papers presented or developed *tools* with LA to help in facilitating PA. Using a novel quantitative approach, Nalli et al. (2021) developed and validated a Moodle plugin to facilitate the creation of heterogeneous groups for a PA activity based on Moodle activity data. Chaparro-Peláez et al. (2020) developed a Moodle application, Workshop Data EXtractor (MWDEX), that can be used to extract, process, analyse, and visualize PA data in Moodle Workshops, and they conducted a short survey with instructors to validate the tool and inquire into how they implement PA. Vozniuk et al. (2014) presented an extension to a social media platform, GRAASP, that facilitates rating-based PA. The extension was evaluated in two analyses: (1) the validity of the PA in relation to the instructor's grade was calculated, and (2) the level of agreement between a group of children who cannot read and a group of university students was compared. Balderas et al. (2018) introduced a scalable framework for conducting qualitative assessments of collaborative Wiki assignments using AssessMediaWiki (AMW), a tool to facilitate PA in Wikis, and StatMediaWiki (SMW), a monitoring tool for Wikis. Both tools provide the instructor with fine-grained assessment information about student's collaborative work.

Table 2.1 An overview of selected papers

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Andriamiseza et al. (2021)	Multiple	616 students 21 instructors	Elaastic	LA for PA	Tools PA design	Scaling	PA design Student learning Validity and reliability
Babik et al. (2019)	–	–	–	LA of PA data	Comparison	PA evaluation	PA design
Balderas et al. (2018)	Computer science	43 students	MediaWiki, SMW, AMW	LA for PA	Tools PA design	Scaling	PA design
Bjælde and Lindberg (2018)	Science teaching	Not reported	–	LA for PA	PA design	–	PA design
Bridges et al. (2020)	Multiple	13 students	LAMS	LA of PA data	Interaction	–	Student interaction
Chaparro-Peláez et al. (2020)	Multiple	23 instructors	Moodle Workshops; MWDEX	LA for PA	Tools	Lack of tools Scaling	PA design
Cheng and Lei (2021)	Educational technology	24 students	Blackboard	LA for PA LA of PA data	Visualizations Interaction	Facilitating interaction	PA design Student interaction
Chiu et al. (2019)	Surgical education	50 students	dVSS	LA of PA data	Interaction	–	Student learning
Choi et al. (2019)	Data science	456 students (2 studies)	–	LA of PA data	Feedback characteristics	Feedback perception	PA design Feedback perception

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Divjak and Maretić (2015)	Entrepreneurship	62 students (2 studies)	Moodle Workshop	LA of PA data	Feedback characteristics	Scaling	Feedback perception Validity and reliability
Djelil et al. (2021)	Multiple	422 students (7 courses)	Sqily	LA of PA data	Interaction	Facilitating interaction	PA design Student interaction
Er et al. (2021a)	Technology enhanced distance learning	14 students	Synergy	LA for PA	Tools Visualizations	Scaling	PA design
Er et al. (2021b)	Primary education	30 students	Synergy	LA of PA data	Interaction	Facilitating interaction	PA design Student interaction
Gunmarsson and Alterman (2014)	Internet and society, human-computer interaction	157 students (2 studies)	Blogosphere by Brandeis University	LA of PA data	Feedback characteristics	Scaling	PA design Student learning Validity and reliability
Huang et al. (2019)	Information management	96 students	Moodle	LA of PA data	Feedback characteristics Interaction	-	PA design Student interaction
Hunt et al. (2021)	Education	76 teachers (2 studies)	EPASS	LA for PA	Automated assessment Visualizations	PA evaluation	PA design Feedback perception

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Khosravi et al. (2020)	Relational databases	384 students	RiPPLE	LA of PA data	Feedback characteristics	Lack of tools	PA design Validity and reliability PA comparison
Koh et al. (2016)	Teamwork competency awareness	272 students	ChatAdmin	LA for PA	Visualizations	–	PA design Feedback perception
Lárusson and White (2012)	Computer science	22 students	–	LA for PA	Automated assessment	–	Student learning
Lin (2019)	Language learning	57 students	IKnow; Zuvio	LA of PA data	Comparison	PA evaluation	PA design Student learning Student interaction
Misiejuk et al. (2021)	Multiple	7660 records	Peergrade	LA of PA data	Feedback characteristics	Feedback perception	Feedback perception
Mørch et al. (2017)	Language learning	125 students	EssayCritic	LA of PA data	Comparison	PA evaluation	Student learning Student interaction
Nalli et al. (2021)	Physics	54 students	Moodle	LA for PA	Tools	Feedback perception	–
Sedrakyan et al. (2014)	Management information systems	86 students	JMermaid tool	LA of PA data	Interaction	–	Student interaction

(continued)

Table 2.1 (continued)

Paper	Discipline/subject	Sample size	Learning tools and platforms*	Application of LA	LA role	PA challenges	PA insights
Shibani et al. (2019)	Law, accounting	293 students (2 studies)	AcaWriter	LA for PA	Automated assessment	-	Student learning Feedback perception
Vogelsang and Ruppertz (2015)	Cyber security and human rights	467 students	iversity	LA of PA data	Comparison	Scaling	Validity and reliability
Vozniuk et al. (2014)	Social media	100 participants (2 studies)	GRAASP	LA for PA	Tools	Lack of tools Scaling	Validity and reliability

*Only learning tools and platforms reported by name in a paper listed

Automated Feedback. Three papers either compared PA with automated feedback or augmented a PA activity with automated feedback. Hunt et al. (2021) conducted a PA activity with teachers who were divided into two groups that used either an e-portfolio without LA or an e-portfolio enhanced with automated feedback and an activity dashboard. The analysis focused on feedback perceptions among feedback receivers and feedback providers. Lárusson and White (2012) developed a tool with which to automatically measure and visualize an originality score (*Point of Originality*) for students' contributions to the teacher to help with monitoring and evaluation in a student co-blogging activity that included PA. The score was validated in the study. Shibani et al. (2019) showcased an implementation of the *Contextualizable Learning Analytics Design* (CLAD) model with the help of an automated feedback tool, AcaWriter, in two contexts: law essay writing and business report writing. In both contexts, the students engaged in a PA activity and were divided groups that either received additional automated feedback from AcaWriter and did not receive automated feedback. An additional usefulness survey was conducted to compare both groups.

Visualization. Three papers focused on data visualization. Koh et al. (2016) presented a *Team and Self Diagnostic Learning* (TSDL) framework aimed at the teamwork competencies and collaboration skills of students. The framework was implemented during a PA activity in which students rated themselves and other team members in an online survey. The similarity scores between self- and peer-ratings were calculated. The results were visualized as student micro-profiles in a radar chart and shown to the students and teachers for their reflection. Er et al. (2021a) presented an open-source platform, Synergy, designed to support PA based on a *Theoretical Framework of Collaborative Peer Feedback*. One of the platform's features is the visualization of students' activity data for the instructor.

Student interaction. Six papers used LA to analyse PA data and explore topics such as student interaction and engagement. Bridges et al. (2020) combined PA data with video and discourse analyses to examine interprofessional team-based learning. Chiu et al. (2019) used peer observation and assessment data as a proxy for active engagement and evaluated their effects on student progress in surgical training using the da Vinci Skills Simulator (dVSS) platform. Djelil et al. (2021) analysed student interaction data from the learning platform Sqily, which included PA, to detect their engagement patterns, roles, and temporal dynamics. Huang et al. (2019) focused on the effects of gamification and quantity- and quality-based badges on peer feedback quality and student engagement in an online discussion forum. The gamification design was based on the Theory-driven Gamification model (GAFCC: *Goal, Access, Feedback, Challenge, Collaboration*), while the PA data were analysed using content analysis and social network analysis. Er et al. (2021b) applied process mining to identify and interpret engagement patterns in data from the PA platform Synergy. Sedrakyan et al. (2014) examined group interaction data during a conceptual modeling process that included PA.

Feedback characteristics. Five papers focused on peer feedback characteristics, such as perception and quality. Gunnarsson and Alterman (2014) conducted a study on peer promotion, a type of PA, in which students assessed other students work by liking other students' posts or awarding badges. Moreover, students were required to engage weekly in more traditional PA assignments by giving feedback using a 3-point scale on a questionnaire form and commenting on two posts. Khosravi et al. (2020) presented an adaptive platform, RiPPLE, that aims to support evaluative judgement skills and conducted a study in which students created multiple choice questions (MCQs) and gave each other peer feedback on the platform. Both the validity of the peer feedback and the development of peer feedback quality over time were explored in this study. Misiejuk et al. (2021) used a variety of LA methods to analyse the backward-evaluation big data to gain new insights into student perceptions of feedback and its relationship to rubrics. Choi et al. (2019) used natural learning processing to code and analyse the PA text data to determine the influence of the social economic status of students on the perceptions of the PA. Divjak and Maretić (2015) developed and tested a novel method via which to measure PA and self-assessment reliability using modified Manhattan metrics.

Comparison. Four papers compared different types of PA. Vogelsang and Ruppertz (2015) analysed MOOC data derived from the innovative integration of teaching assistants into assessment activities to determine student performance and the validity of this method in relation to PA, automated assessment, and instructor grading. Lin (2019) compared online and paper-based PA to explore the differences in learning achievement, learning involvement (measured using log data from a learning management system), learning autonomy, and student learning reflections. Mørch et al. (2017) generated automated feedback in the EssayCritic system for one group in a language learning scenario and compared their learning performance and writing process with a group that engaged in PA without EssayCritic. Babik et al. (2019) simulated datasets using LA methods to compare ranking-based and rating-based PA with a focus on structural effects.

Design. Two papers focused on designing PA. Bjælde and Lindberg (2018) reported on course design examples incorporating continuous feedback, including PA, and LA. Andriamiseza et al. (2021) explored a two-votes-based process, a form of peer instruction with embedded PA. The results of a learning activity that was conducted on the web platform Elaastic were analysed and presented to the instructors to inform their practice. This study not only provided instructors with recommendations for orchestration but also system designers with recommendations when designing a formative assessment system.

RQ2: What are the reported peer assessment challenges the research addressed with learning analytics? And how are they addressed?

Only 18 papers reported on challenges facing PA *that may be mitigated through LA*, while three papers reported on more than one issue. We identified five main

challenges: *scaling*, *PA evaluation*, *lack of tools*, *feedback perception*, and *facilitating interaction*. In this section, we describe the challenges and their potential mitigation.

Scaling. The scaling of PA was the challenge LA had the most potential to help, as reported in eight papers. As noted by Andriamiseza et al. (2021), the scaling of assessment activities generates rich datasets that may be used to help inform instructor practice. In their study, the data from a two-votes-based process with embedded PA was analysed to inform classroom orchestration. Chaparro-Peláez et al. (2020) noted the need to support MOOCs with efficient student-centered assessment methods, such as PA, which can be made scalable by using LA. To encourage the adoption of PA as a scalable assessment solution for large courses, Vozniuk et al. (2014) used LA to validate PA use on a social media platform, GRAASP, which can be used to set up a PA activity. A PA platform, Synergy, with integrated LA, was presented to facilitate the scaling of dialogic peer feedback in Er et al. (2021a). Gunnarsson and Alterman (2014) noted that students' content production in blogging environments may overwhelm instructors and lead them to not being able to identify and highlight high-quality contributions to the class. This was mitigated by the implementation of peer promotion, a type of PA that uses likes and badges. Although Wikis provide rich data that may be used to evaluate various skills, the assessment of Wikis is very complex and difficult to scale. To address this, Balderas et al. (2018) gave teachers information from qualitative and quantitative LA-supported assessment during a PA activity using Wikis. Divjak and Maretić (2015) described the need to use LA data to explore PA and assess the reliability and validity of PA, especially in large classrooms. For example, they noted that LA could help with equalizing in a PA activity—such as students giving all their peers the same marks—by discovering assessment patterns. A second example addresses students' lack of the metacognitive skills needed to perform PA, which may be mitigated by using LA to calculate PA reliability, which would enable teachers to identify students who needs help.

PA evaluation. Four papers identified the challenge of evaluating PA as an activity. Because online PA has the potential to facilitate higher-order thinking, such as improving writing abilities in language learning, and can be used as an effective flipped-classroom strategy, Lin (2019) studied the differences between online and paper-based PA. LA data from a learning management system were used as a proxy for student's learning involvement in both scenarios. Mørch et al. (2017) noted that LA-generated automated feedback may be as accurate and reliable as PA. At the same time, these systems could lead to conformity and less creativity in writing. To explore these issues, a study was conducted that compared the learning performance and writing processes of students who received automated feedback with those of students who only received feedback from their peers. Babik et al. (2019) observed that comparing different PA methods using real-life assessment data may conflate the analysis with cognitive and behavioural effects. To mitigate this phenomenon and focus on structural effects, a simulation model of PA was developed using a Monte-Carlo simulation, and network typology and aggregation

methods were used to compare ranking-based and rating-based PA. Hunt et al. (2021) report a potential advantage derived from adding LA to e-portfolios used in a PA activity: providing more tailored and timely feedback.

Lack of tools. Four papers described the lack of PA tools. Nalli et al. (2021) described a lack of tools that support the formation of heterogeneous groups of students for PA. To address this, a variety of clustering algorithms using Moodle activity data were evaluated, and a Moodle plugin for group formation was developed and validated. Chaparro-Peláez et al. (2020) reported that there are few software tools to support PA. Moreover, the current Moodle Workshops version has many limitations in terms of data visualization, extraction, and exporting. As a solution, a new tool with LA functionalities, Moodle Workshop Data Extractor (MWDEX), was presented. The development of a PA extension for the social media platform GRAASP by Vozniuk et al. (2014) was motivated by the lack of ready-to-use PA platforms and PA validity issues. Many tools do not enable data harvesting, so the impact of implemented strategies cannot be evaluated. Khosravi et al. (2020) presented an adaptive tool, RiPPLE, that enables data extraction and fosters evaluative judgements. In an empirical study focusing on PA validity, students developed and peer-assessed multiple-choice questions (MCQs).

Feedback perception. Three papers recognized improving peer feedback perceptions as an important PA challenge. The current application of the Moodle Workshop randomly forms student groups for a PA activity, which negatively influences student satisfaction with the assessment activity. This motivated Nalli et al. (2021) to propose a sophisticated quantitative LA method and a Moodle plugin to form heterogeneous groups, the implementation of which may lead to more positive perceptions of PA and higher success rates for all students in a class. Misiejuk et al. (2021) identified a challenge in understanding student perceptions of the feedback they received with regards to being able to use it effectively. To address this challenge, an extensive study that used a large dataset and applied a variety of LA methods (ENA, regression, and other methods) was conducted. Choi et al. (2019) described a need to understand the impact of socio-economic status on how student feedback is perceived. As a part of their analysis intended to gain more insights into this problem, they used automated text classification, an LA technique, to detect feedback characteristics.

Facilitating interaction. Three papers identify facilitating student interaction in PA as a problem and suggest that LA could help. Cheng and Lei (2021) identified the need to facilitate student interactions in blogging activities that include PA. Social network analysis (SNA), an LA technique, was used to analyse and visualize student engagement and group cohesion. The SNA graphs were shown to students, and their effect on student behaviour was explored. Djelil et al. (2021) noted that engaging students in PA is difficult and that PA itself is prone to biases. To gain more insights into student interactions, social network analysis, specifically a graphlet-based method, and clustering were used to analyse the PA data. Er et al. (2021b) noticed a challenge in understanding student engagement patterns

that could be used to improve PA. To identify these patterns, log data from a PA platform, Synergy, was analysed using process mining.

RQ3: What insights into peer assessment can we gain from learning analytics?

Only one paper did not report any insights into PA. We found five types of PA insights, which were *PA design*, *student learning*, *PA validity and reliability*, *student interaction*, and *feedback perception*.

PA Design. Most papers contributed new or improved designs for a PA activity with the help of LA, or their insights could inform more effective PA designs. The adaptive platform RiPPLE, presented by Khosravi et al. (2020), provides a learning environment that supports evaluative judgement and PA. Moreover, the tool enables the measurement and evaluation of such interventions. The theory-oriented design of the PA platform Synergy, presented in Er et al. (2021a), was evaluated positively by a group of students. While comparing online and paper-based PA, Lin (2019) noted the students' frustration with small screens in the online PA group when engaged with PA on mobile devices. An evaluation of the Workshop Data EXtractor (MWDEX), developed by Chaparro-Peláez et al. (2020), indicated that instructors typically do not use any software tool to facilitate PA. Moreover, although most instructors use Moodle in their day-to-day practice, they choose Blackboard's PA application rather than Moodle Workshops when they decide to use software to support PA, which may indicate their dissatisfaction with the Moodle Workshop module for PA.

A radar chart visualizing the similarity scores between self- and peer-ratings in a team awareness activity, presented in Koh et al. (2016), was perceived positively as a visualization tool, although the students had difficulties interpreting the similarity scores. The need for a more user-friendly dashboard was emphasized, and because some students and teachers found the PA ratings dishonest in the team awareness study, more training in PA was recommended. Cheng and Lei (2021) found that, when an interaction graph of within-group interactions was shown to students after the first PA activity, this had the undesired effect of generating fewer cross-group comments in the following cycles. This indicates that a clearer explanation of performance expectations is needed to help students interpret the visual analytics of their behaviour.

The finding that PA rating scales outperformed PA ranking scales according to a study conducted by Babik et al. (2019) can be used to design PA activities and systems because choosing either scale must be considered together with other design choices that they may influence, positively or negatively, either scale's validity and reliability.

Bjælde and Lindberg (2018) presented a course design that integrated PA and LA to facilitate assessment as learning and continuous feedback as an early intervention method. Student feedback perceptions after the PA activity guided the future course design. A scalable qualitative assessment framework that uses LA, developed by Balderas et al. (2018), can help teachers with the large-scale assessment, including PA, of collaborative Wiki contributions. Andriamiseza et al. (2021)

recommended that formative assessment systems based on a two-votes-based process show teachers the proportion of correct answers at the first vote, as well as the correlation between the correctness of a student's rating and their confidence level. In addition, they recommend that PA activities not include self-ratings, and that the system should be flexible in terms of how many peers assess one another.

Gamification is cost effective and relatively easy to implement and likely increases PA engagement in online discussion forums, as shown in Huang et al. (2019). Peer promotion using badges and likes may be considered as an addition to traditional PA to reduce instructors' workload, as described in Gunnarsson and Alterman (2014).

Er et al. (2021b) found that high-performing students had many bidirectional transitions between self-regulated learning and socially regulated learning, as well as between self-regulated learning and co-regulated learning on a PA platform, Synergy. One implication of this behaviour that may lead to better student performance is that additional support for engaging students in self-regulated learning, socially regulated learning, and co-regulated learning should be provided. Hunt et al. (2021) compared a group using an e-portfolio with LA visualizations and a group using an e-portfolio without LA. Both groups indicated a need for a face-to-face discussion as a part of the feedback process. Teachers in the e-portfolio with the LA group indicated that they need more support in dealing with analytics due to a lack of digital skills. Furthermore, they expressed a need to have more control over the visual analytics of their activities because the teachers felt overwhelmed at times. Djelil et al. (2021) used social network analysis (a form of LA) with data from the learning platform Sqily and found that teacher presence was significant across courses and crucial to initiating initial PA activities, suggesting that students may need support and direct guidance from a teacher to begin interacting with peers. The finding by Choi et al. (2019) that students reacted differently to feedback provided by students with different socioeconomic statuses (i.e., based on the nationality of the peer feedback provider) has design implications. Instructors must pay attention to which information about learners is visible to others, including indirect information that may indicate socioeconomic status, such as a name or profile picture. On the other hand, socioeconomic information may help instructors pair students with different socioeconomic statuses and thus ensure exposure to different perspectives.

Student learning. Six papers reported insights into student learning. Lin (2019) found no learning performance difference between online and paper-based PA in a flipped language-learning class. However, the online PA group expressed more ideas in their work, expressed more interest in the flipped learning environment, and showed higher learner autonomy during previewing before the class. Mørch et al. (2017) found no significant difference in learning performance between a group using automated feedback and a group engaged in PA. The group that used automated feedback, however, used significantly more subthemes and showed more ideas inspired by the automated feedback in their writing. It was difficult for students in the PA group to give content-oriented feedback, and they preferred to

comment on the essay structure. A group with additional automated feedback used significantly more rhetorical moves in their essays in the first context in Shibani et al. (2019). Furthermore, PA helped students with sense-making regarding the automated feedback.

Lárusson and White (2012) found a statistically significant positive correlation between the number of contributions that included comments on other students' blogs and students' final performance and the originality of their contributions. Chiu et al. (2019) found that implementing peer observation with PA during surgical student practice on a da Vinci Skills Simulator (dVSS) facilitated the improved performance of intermediate-level surgical tasks but not basic or advanced tasks. In a study of the two-votes-based PA process, Andriamiseza et al. (2021) found that benefits of formative assessment sequences increased when (1) the proportion of correct answers is close to 50% during the first vote or (2) the written rationales from students who gave correct answers are better rated than those from students with incorrect answers. However, the number of peer ratings made no significant difference in terms of the benefits of the formative sequences' benefits.

Reliability and Validity. Four papers described findings about PA reliability and validity. Vogelsang and Ruppertz (2015) found peer and teaching assistants' grading invalid as compared with the expert's grading. However, peer grading was valid, assuming that the teaching assistants' grading was accurate. Andriamiseza et al. (2021) established that peer ratings were consistent when correct learners were more confident than incorrect ones, while self-ratings were inconsistent in the peer rating context. Khosravi et al. (2020) established a strong and positive correlation between student and domain expert ratings on multiple choice questions (MCQs) on an adaptive platform, RiPPLE. Furthermore, the difference between the domain expert ratings and peer ratings decreased with time and practice. Gunnarsson and Alterman (2014) found that peer promotion helped identify higher quality posts. Some students could be identified as more reliable in evaluating post quality than others. Moreover, badges given before or after the traditional PA activity were found to be more reliable than those given during the PA activity. The evaluation of the GRAASP extension developed by Vozniuk et al. (2014) showed a strong agreement between the grades assigned by students and instructors in rating-based PA. To confirm that students did not grade the reports based on appearance, a second experiment was conducted with children, who rated the reports only based on their appearance, without reading the reports' content. Little agreement between the grades assigned by the students and children was found, and this result confirmed that students engaged with the content of the reports before grading them.

Divjak and Maretić (2015) developed a reliability measurement based on the modified Manhattan metrics (based on taxicab norm) indicating that reliable peer grading should be within 2 points (i.e., less than or equal to 2), while peer grading would be unreliable if it exceeded 2 points. In their case study, the PA grades were reliable.

Student interactions. Eight papers provided new insights into student interactions and behaviour. Students in the online PA group demonstrated higher learning involvement during flipped learning than the paper-based PA group in the study by Lin (2019). Mørch et al. (2017) noted that students in the automated feedback group were more motivated and worked harder on their essays than the PA group. The gamification-based group posted more, engaged more in PA, and gave higher-quality peer feedback in an online discussion forum than the control group in the study reported in Huang et al. (2019). A larger group of students in the gamification-based group provided feedback in comparison to the control group. After showing students their social network graphs on their intra-group blogging and PA behaviour, in the study reported in Cheng and Lei (2021), the interactions within the same group increased and the exploration of outside-group blogs decreased. This resulted in a clear subgroup structure.

Sedrakyan et al. (2014) found that both the best- and the worst-performing students were more engaged in their modeling activities just before the activity deadlines, including the PA deadline. However, the best performing groups were also very active between the deadlines. Moreover, the best-performing groups implemented more peer feedback in their models in comparison to the worst-performing groups during the conceptual modeling activity. Er et al. (2021b) found that high-performing students were more likely to engage as described in the theoretical framework for collaborative peer feedback, while medium-performing students deviated from the theory.

Djelil et al. (2021) found a positive trend in terms of learners engaging in PA activities on the learning platform Sqily. Furthermore, it was found that students may need some time to feel comfortable providing feedback to new peers. Bridges et al. (2020) compared the video, discourse, and PA data of two groups during an interprofessional team-based learning activity. According to an analysis of the PA data, the first group did not identify a leader, and their physical orientation was spatially and interactionally cohesive. The second group identified a strong leader, both in their PA and in their spatial composition.

Feedback perception. Five papers reported findings on feedback perception. Feedback providers, reported on in Hunt et al. (2021), found that giving feedback to others helped them reflect on their own work. At the same time, they felt uncomfortable being critical toward their colleagues. This influenced feedback receivers and their perception of feedback providers as not always being honest. Another finding in this study was that the group using an e-portfolio with LA had significantly more positive perceptions of the entire feedback experience than the group using an e-portfolio without LA. However, there were no significant differences in the perceptions of the quantity, quality, and use of feedback between the two groups. The PA activity in Divjak and Maretić (2015) was overwhelmingly perceived as motivating. In the team awareness study of Koh et al. (2016), some students and teachers disagreed with PA ratings and perceived them as dishonest. Students who perceive feedback as useful acknowledged their errors, expressed the intention to revise their text, and/or gave praise regarding the feedback in their

backward-evaluation comments, as shown in a study by Misiejuk et al. (2021). Students who evaluated the feedback that they received as not useful or showed confusion about the feedback were critical toward it and/or disagreed with it in their backward-evaluation comments. In general, students wanted feedback to be more specific, just, and constructive, rather than kind. No significant relationship between backward-evaluation and the structure of the PA rubric was found. Two studies by Shibani et al. (2019) found that the students in both studies perceived the writing activity with added automated feedback to be more useful than only PA without automated feedback. High-socioeconomic status students reacted differently to the feedback from medium- and low-socioeconomic status students in terms of feedback agreement and formality when status information was disclosed in a study reported in Choi et al. (2019).

2.5 Discussions and Conclusions

This chapter presents the first scoping review mapping of the LA applications in PA research. The review included 27 very diverse papers, which made reporting on the results challenging. Our research questions focused on the PA challenges that the papers identified and how were they addressed using LA, the role of the LA application in the PA activity, and, finally, the kinds of PA insights reported. We found two main areas in which learning analytics was used for PA: using LA to improve PA activity and using LA to analyse PA data.

We found that most research focused on addressing the challenge of scaling PA, developing new PA tools enhanced by analytics, or attempting to inform PA theory by evaluating different types of PA. Many insights from the research reported in the included papers may inspire new PA designs or improve existing ones by paying heed to reports of successful and unsuccessful implementations of LA in PA activities. In addition to the traditional PA research focus, such as validity and reliability and student learning, interesting studies were conducted on student interaction in PA, in which self-regulation, group building, and student interaction are analysed. Moreover, rich data from gamification-enhanced PA and collaborative writing in blogs or Wikis are utilised to gain more dynamic insights into students' development of feedback skills and learning.

This study has certain limitations. First, the inclusion/exclusion process was difficult, and perhaps, some papers that should have been included were excluded. It was challenging to define which papers were actually using LA because some papers used LA methods without the authors describing them in their papers. Thus, instead of evaluating the "LA-ness" of the papers, we used a proxy that defined a paper as being about using LA in PA research if that paper described LA or was published at the LAK conference or in the Journal of Learning Analytics. Furthermore, a significant group of papers were excluded because they focussed on insights into LA, rather than PA. For example, PA data contributed to the final grade, which was a part of a dataset analysed using LA to identify patterns of episodic emotions in MOOCs (Han et al., 2021) or to predict time-on-task estimation

strategies (Kovanović et al., 2015). Future reviews might focus on PA data as a part of big datasets and how are they explored using LA; this topic was outside the scope of this chapter, but we found many papers addressing this issue.

Second, PA may be a part of many learning activities, such as student interactions in a discussion forum, but it is not always conceptualised as PA or analysed as such. We tried our best to include a variation of PA implementations, but with the large number of papers found in the search, this was not a trivial task. Finally, the diversity of the papers made the analysis challenging because it was difficult to identify the same issue across them, which may have led to some simplifications in our analysis of the papers and their insights into PA.

Several areas for further research were identified in this review. First, more work is needed to use insights from LA to improve the PA activity before once again using LA to see if there has been improvement (*cf.* Clow, 2012). Some of the papers in our review included two studies, however, the results from first study which gave insights into PA did not lead to a second study that used those insight to improve PA. Second, the automatization of aspects of PA (e.g., feedback classification; Wahid et al., 2016; Ryan et al., 2019) was identified as a potential application for LA, and though there are papers in our review that attempted to automate, the examples are few. Moreover, it seems that the automated methods, such as automated assessment, were used to compare with PA without an automated method, rather than using it as part of a PA process. Thus, more research is needed to improve the automation of aspects of PA through, for example, additional text quality measurements or group formation, as well as empirical studies of their implementations in teaching practice. Third, we found that the focus on either analysing PA data to gain insights into PA or trying to improve PA in practice is limiting, although necessary in some cases. Future research should investigate combining the two and using the LA insights directly to improve PA activity and tools. Fourth, as found in this review, showing analytics to students influences their behaviour, which in turn may be used as a powerful pedagogical tool within PA; however, more work in this area is needed both to understand the effect of analytics on students and teachers/instructors and how the analytics could be integrated into a PA activity. Finally, the analytics used to analyse data in the studies reported in the papers are significantly more advanced than the analytics currently available in the PA tools. A sensible integration of advanced analytics in the PA tools is another promising research area that would include not only technical aspects, but also would include examining the perception and understanding of the analytics by both students and instructors.

This review has shown that LA has the potential better understand and improve PA activity through new insights into student behaviour and the artefacts that they produce, interpersonal and intergroup interactions, or tool improvement. However, the research is still emerging and scattered. LA gives access to hidden data and finding patterns and insights in data from PA activity that is not easily accessible to humans. We hope that this review will act as a starting point for future work on using learning analytics to improve peer assessment activity.

References¹

- Andriamizeza, R., Silvestre, F., Parmentier, J. F., & Broisin, J. (2021). Recommendations for orchestration of formative assessment sequences: A data-driven approach. In *Proceedings of the EC-TEL'21* (pp. 245–259).*
- Archer, E., & Prinsloo, P. (2020). Speaking the unspoken in learning analytics: Troubling the defaults. *Assessment and Evaluation in Higher Education*, 45(6), 888–900.
- Babik, D., Stevens, S., & Waters, A. E. (2019). Comparison of ranking and rating scales in online peer assessment: Simulation approach. In *Proceedings of the LAK'19* (pp. 205–209).*
- Balderas, A., Palomo-Duarte, M., Doderó, J. M., Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2018). Scalable authentic assessment of collaborative work assignments in wikis. *International Journal of Educational Technology in Higher Education*, 15(1), 1–21.*
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/j.edurev.2022.100489>.
- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19(1–2), 205–220.
- Bjælde, O. E., & Lindberg, A. B. (2018). Using continuous assessment with feedback loops to generate useful data for learning analytics. In *Proceedings of the ASCILITE'18* (pp. 53–62).*
- Bridges, S. M., Chan, L. K., Chen, J. Y., Tsang, J. P., & Ganotice, F. A. (2020). Learning environments for interprofessional education: A micro-ethnography of sociomaterial assemblages in team-based learning. *Nurse Education Today*, 94.*
- Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology and Society*, 15(3), 3–26.
- Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2.
- Chaparro-Peláez, J., Iglesias-Pradas, S., Rodríguez-Sedano, F. J., & Acquila-Natale, E. (2020). Extraction, processing and visualization of peer assessment data in Moodle. *Applied Sciences*, 10(1), 163.*
- Chaudy, Y., & Connolly, T. (2019). Specification and evaluation of an assessment engine for educational games: Integrating learning analytics and providing an assessment authoring tool. *Entertainment Computing*, 30.
- Cheng, J., & Lei, J. (2021). A description of students' commenting behaviours in an online blogging activity. *E-Learning and Digital Media*, 18(2), 209–225.*
- Chiu, H. Y., Kang, Y. N., Wang, W. L., Chen, C. C., Hsu, W., Tseng, M. F., & Wei, P. L. (2019). The role of active engagement of peer observation in the acquisition of surgical skills in virtual reality tasks for novices. *Journal of Surgical Education*, 76(6), 1655–1662.*
- Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In *Proceedings of the LAK'12* (pp. 134–138).
- Choi, H., Dowell, N., Brooks, C., & Teasley, S. (2019). Social comparison in MOOCs: Perceived SES, opinion, and message formality. In *Proceedings of the LAK'19* (pp. 160–169).*
- Cope, B., & Kalantzis, M. (2016). Big data comes to school: Implications for learning, assessment, and research. *Open*, 2(2).
- Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerd, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers and Education*, 162.

¹ *indicates papers included in the scoping review.

- Divjak, B., & Maretić, M. (2015). Learning analytics for e-assessment: The state of the art and one case study. In *Proceedings of the CECIIS'15*.*
- Djelil, F., Brisson, L., Charbey, R., Bothorel, C., Gilliot, J. M., & Ruffieux, P. (2021). Analysing peer assessment interactions and their temporal dynamics using a graphlet-based method. In *Proceedings of the EC-TEL'21* (pp. 82–95).*
- Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology*, 44(4), 662–664.
- Er, E., Dimitriadis, Y., & Gašević, D. (2021a). Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions. *Assessment and Evaluation in Higher Education*, 46(2), 169–190.*
- Er, E., Villa-Torrano, C., Dimitriadis, Y., Gašević, D., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Gómez-Sánchez, E. & Martínez Monés, A. (2021b). Theory-based learning analytics to explore student engagement patterns in a peer review activity. In *Proceedings of the LAK'21* (pp. 196–206).*
- Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education policy*. Joint Research Centre.*
- Gunnarsson, B. L., & Alterman, R. (2014). Peer promotions as a method to identify quality content. *Journal of Learning Analytics*, 1(2), 126–150.*
- Han, Z. M., Huang, C. Q., Yu, J. H., & Tsai, C. C. (2021). Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. *Computers in Human Behavior*, 122.
- Huang, B., Hwang, G. J., Hew, K. F., & Warning, P. (2019). Effects of gamification on students' online interactive patterns and peer-feedback. *Distance Education*, 40(3), 350–379.*
- Hunt, P., Leijen, Ä., & van der Schaaf, M. (2021). Automated feedback is nice and human presence makes it better: Teachers' perceptions of feedback by means of an e-portfolio enhanced with learning analytics. *Education Sciences*, 11(6).*
- Khosravi, H., Gyamfi, G., Hanna, B. E., & Lodge, J. (2020). Fostering and supporting empirical research on evaluative judgement via a crowdsourced adaptive learning system. In *Proceedings of the LAK'20* (pp. 83–88).*
- Knight, S. (2020). Augmenting assessment with learning analytics. In *Re-imagining University Assessment in a Digital World* (pp. 129–145).*
- Koh, E., Shibani, A., Tan, J. P. L., & Hong, H. (2016). A pedagogical framework for learning analytics in collaborative inquiry tasks: An example from a teamwork competency awareness program. In *Proceedings of the LAK'16* (pp. 74–83).*
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *Proceedings of the LAK'15* (pp. 184–193).
- Lárusson, J. A., & White, B. (2012). Monitoring student progress through their written "point of originality." In *Proceedings of the LAK'12* (pp. 212–221).*
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(69), 1–9.
- Lin, C. J. (2019). An online peer assessment approach to supporting mind-mapping flipped learning activities for college English writing courses. *Journal of Computers in Education*, 6(3), 385–415.
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–290.
- Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers and Education*, 175.
- Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior*, 117.*
- Mørch, A. I., Engeness, I., Cheng, V. C., Cheung, W. K., & Wong, K. C. (2017). EssayCritic: Writing to learn with a knowledge-based design critiquing system. *Journal of Educational Technology and Society*, 20(2), 213–223.*

- Nalli, G., Amendola, D., Perali, A., & Mostarda, L. (2021). Comparative analysis of clustering algorithms and Moodle plugin for creation of student heterogeneous groups in online university courses. *Applied Sciences*, *11*(13).*
- Nyland, R. (2018). A review of tools and techniques for data-enabled formative assessment. *Journal of Educational Technology Systems*, *46*(4), 505–526.
- Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment and Evaluation in Higher Education*, *43*(3), 428–438.
- Ryan, T., Gašević, D., & Henderson, M. (2019). Identifying the impact of feedback over time and at scale: Opportunities for learning analytics. In *The impact of feedback in higher education* (pp. 207–223).
- Saqr, M. (2017). Assessment analytics: The missing step. *International Journal of Health Sciences*, *11*(1).
- Sedrakyan, G., Snoeck, M., & De Weerd, J. (2014). Process mining analysis of conceptual modeling behavior of novices—empirical study using JMermaid modeling and experimental logging environment. *Computers in Human Behavior*, *41*, 486–503.*
- Shibani, A., Knight, S., & Buckingham Shum, S. (2019). Contextualizable learning analytics design: A generic model and writing analytics evaluations. In *Proceedings of the LAK'19* (pp. 210–219).*
- Vogelsang, T., & Ruppertz, L. (2015). On the validity of peer grading and a cloud teaching assistant system. In *Proceedings of the LAK'15* (pp. 41–50).*
- Vozniuk, A., Holzer, A., & Gillet, D. (2014). Peer assessment based on ratings in a social media course. In *Proceedings of the LAK'14* (pp. 133–137).*
- Wahid, U., Chatti, M. A., & Schroeder, U. (2016). Improving peer assessment by using learning analytics. In *DeLFI Workshops* (pp. 52–54).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

