Collaborative Correction in Mass Education as a Social Media Application

Niels Heller, François Bry
Ludwig Maximilian University of Munich, Germany
niels.heller@ifi.lmu.de
bry@lmu.de

Abstract. Social media hold great promises in tertiary mass education because they foster the creation and sharing of information among communities. With educational social media, students can exchange on courses' contents and discuss homework solutions while being supervised by teachers. However, all too often social media are used in education as mere organisational tools, delivery tools for course material, or collection tools for homework assignments. This article reports on a versatile social medium which, among others, exploits collaborative correction, a form of human computation, for tracking systematic errors of STEM students. Learning analytics generated from both human-produced and system-collected data and the results of a real-life evaluation are presented. The results point to a positive reception by students and teachers alike and to the effectiveness of the medium for higher STEM education.

Keywords: Conceptual Change, Case Study, Technology Enhanced Learning, Social Media, Systematic Error Identification

1. Introduction

Social media hold great promises for higher education. With social media, students can share ideas on course content, discuss different homework solutions and even reach a consensus on correct answers while being supervised by the teacher whose role shifts from transmitting knowledge to facilitating learning. The reality is, in the authors' experience, however far from this. All too often, social media are used as mere organisational tools, for scheduling activities, for distributing course materials, or for collecting students' homework. Posts shared by students on social media used in such manners mostly refer to organizational matters and rarely to the course contents.

This article addresses problems of teachers and students alike: Teachers often face the problem of providing feedback on student homework, which, while being time-consuming to compile, should be of high quality and delivered in a timely manner. Students, on the other hand, need to "make sense" of that feedback, which often means overcoming misconceptions they have or correcting errors they made.

The research presented in this article is centred at the following hypothesis: By using social media to collect and report descriptions of systematic errors commonly made by students, teaching processes, in this case homework corrections, and learning processes, in this case overcoming misconceptions and correcting errors, are improved.

To the authors' best knowledge, this article presents the first attempt to use social media in such a manner. The approach is a form of human computation: An orchestration and incorporation of human contributions to a computer-controlled process with the goal of generating a collaborative outcome of a higher value than each single human contribution.

The contribution of this article is twofold. First, a software system is described which aims at systematically gathering teacher assessments of students' homework while easing the (time-consuming) process of reviewing students' homework. The system achieves its goals by facilitating for the collection and collaborative revision of descriptions of systematic errors, that is, fallacies independently committed by different students performing the same or similar tasks. The system eases teachers' corrections by making it possible to refer to already collected systematic errors, hence to *a posteriori* provide students with error descriptions (of errors they had committed). This collaborative form of correction is beneficial to the students' learning because it *a priori* gives students access to systematic error descriptions related to the tasks they currently are working on or have recently been working on.

A second contribution of this article is a report on an evaluation of the aforementioned approach. The evaluation is based on data gathered in several venues of a computer science course and is focused at the following research questions:

1. Do students better avoid systematic errors when they have access to descriptions of systematic errors commonly made by students carrying out that same task?
2. Does the engagement of students negatively correlate with the numbers of systematic errors they make?

1. What are the teachers' and students' (subjective) attitudes towards the novel approach? In particular, do teachers think the approach eases their homework correction and do students find the *a priori* access to systematic error descriptions helpful for learning?

The evaluation presented in this article positively answers research question one and two. Teachers using the medium indicated that it was helpful and reduced their workload and students found the medium helpful for learning and for carrying out exercises. These findings positively answer research question 3.
This article builds upon previous work (Heller & Bry 2018) which proposed the collaborative approach to systematic errors detection and analysed part of the data this article refers to. The priming of students with previously gathered error descriptions and the report on an evaluation of this priming in a real-life university course is a novel contribution.
This article is structured as follows: Section 1 is this introduction. Section 2 discusses related work. Section 3 introduces the scientific methods. Section 4 and 5 respectively present and discuss the evaluation results. Section 6 concludes the article with perspectives for future work.

2. Related Work

The work reported about in this article ties together a number of research fields, namely social media and their use in education, human computation, and pedagogical research regarding systematic errors and conceptual change. In the following, those aspects of these fields essential for understanding the results presented in this article are briefly addressed.

Social media in education. The use of social media in education has been advocated since it promises to empower and engage students (Lee & McLoughlin 2017, Rankin 2009) and since it allows students to define personal learning spaces (Dabbagh & Kitsantas 2012).
By examining a group of pre-service teachers, Sadaf et al. found that 51% of their subjects were committed to use web 2.0 technology in their future classrooms since it promises to increase student engagement, yet expressed uncertainty on *how* to incorporate such technologies in their teaching (Sadaf, Newby & Ertmer 2012). In a study conducted in 2010 Brady et al. found that the majority of students think that the use of social media facilitates communication and increases communication's effectiveness (Brady & Holcomb & Smith 2010).
Rambe found by examining posts on facebook that most social media use in education was administrative and formal (Rambe 2012), a finding backed by Ophus and Abbitt who found that "schedule views" and "access to notes and materials" were among the most anticipated interactions on a social medium used for education (Ophus & Abbitt 2009). In a more recent survey conducted with 333 university teachers, content sharing (both by students and teachers), discussing, and organizing were found to be the most common social media usages (Gruzd et al. 2018).
Students' perceptions of social media usage for education was explored by Hamid et al. in focus group discussions (Hamid et al. 2015). Students indicated, among others, that social media were beneficial for learning because they allow to easily assess the learning progress, to promote critical thinking and to foster self-directed learning. These properties are also often referred to in the literature on heutagogy (see Blascke 2016 for instance), a student-centred learning and teaching approach especially suited for social media use (Blascke 2016).
The use of established, typically commercial, social media systems such as facebook or twitter is often mentioned in the literature: See (Tess 2013) for an extensive literature survey on the subject. Drawbacks of commercial social media which were not designed for educational purposes is discussed in (Friesen & Lowe 2010). Note that the social medium the case study reported in this article was conducted with an experimental platform designed for case studies, not with an established platform, that the students would have been familiar with.

Human Computation. Von Ahn defines Human Computation as "a paradigm for utilizing human processing power to solve problems that computers cannot yet solve" (Von Ahn 2008, page 3). Human Computation applications address problems from a variety of fields (Quinn & Bederson 2011), yet to the authors' best knowledge, no use of human computation in education has been so far reported in the literature.
Using the taxonomy of Quinn and Bederson (Quinn & Bederson 2011), the application presented in this article can be categorized as both a human computation and a "social computing" application. Social computing has been defined as "applications and services that facilitate collective action and social interaction online with

rich exchange of multimedia information and evolution of aggregate knowledge" (Parameswaran & Whinston 2007).
The tasks performed by humans this research reports about are two fold: A first task is the collaborative collecting of systematic errors students make in solving exercises. A second task is detecting already collected systematic errors in students' submissions. The first task is a wiki-like activity which might not fully fit with the definition of human computation (Quinn & Bederson 2011) because it is a mere collecting, the added value of which could be questioned. The second task, recognizing systematic errors in submissions, on the other hand, clearly is a human computation task: Indeed, tagging artefacts (like images, tunes, or in this case student submissions) is a typical form of human computation (Wieser et al. 2013, Law, Von Ahn 2009, Von Ahn & Dabbish 2004).

Systematic Errors and Misconceptions. Student misconceptions (or preconcepts or naïve concepts) and systematic errors have been thoroughly researched in the literature: See (Confrey 1990) for an extensive literature review. In this review, Confrey observes that the research on student concepts and misconceptions typically stems from within the constructivist tradition, often examining science and mathematics education, while "systematic errors", being more related to the cognitivist tradition, are typically studied in mathematics and computer science education research. She further states that misconceptions relate to systematic errors, as "they result from non-random applications of rules based on certain beliefs" (Confrey 1990, page 33). In the following, no distinction between systematic errors and misconceptions is made. As the case study presented in this article is conducted in a computer science course, the term "systematic error" can be used to refer to both phenomena.
The research presented in this article reports on a social media application, which allows teachers to collaboratively collect, report and elaborate on descriptions of systematic errors. Certain authors have similarly attempted to find systematic errors in student works and to classify these errors in more general schemes: See for example (Radatz 1979) and (Movshovitz & Hadar 1987) for error classifications in secondary education mathematics. Similar generalisations are not attempted in the present article: A description of a systematic error collected in the case study reported about in the following is only applicable for a specific (type of) exercises.
In the case study reported about below, descriptions of systematic errors are presented to the students in the hope that the students do not make these errors. This approach is related to the conceptual change model first proposed by Posner et al. (Posner et al. 1982) which states that learners exchange their, possibly erroneous, conceptions when engaging in a learning activity for a new, hopefully less erroneous, conception if certain criteria are met. Texts that are designed to provoke conceptual change (such as the systematic error descriptions in the case study reported about below) are referred to in the literature as "refutation texts". Refutation texts are widely accepted as effective for learning (see (Tippett 2010) for an extensive literature review) which motivates the approach of the case study reported about in this article.

3. Methods

Course, Participants, and Course Organization. The course this case study reports about was an introduction to theoretical computer science offered every year to computer science bachelor students. The course lasted 14 weeks from April to July 2018 and was attended by 433 students (113 female and 315 male). The students were tasked with 11 weekly homework assignments, each consisting of 3 to 4 exercises. The course's teaching staff consisted of a professor, a research associate and five student tutors.

Data Collection and Datasets. Three datasets were examined for this article. A *base dataset* was gathered in a course venue of 2016: One teacher gathered systematic errors in homework submissions of 82 students in 30 exercises. Previous work of the same authors relies on this dataset and establishes the validity of the error classification by measuring an inter rater reliability of sufficient quality with a Fleiss-κ of 0.78 (Heller, Bry 2018).
A *validation dataset* was gathered in a course venue in 2017: Two teachers identified systematic errors in homework submissions of 89 students in 28 exercises. This dataset was created to determine whether systematic errors would re-occur in different venues of the same course and to compare systematic error frequencies. The evaluation of the *base dataset* and the *validation dataset* can be seen as a longitudinal study design, as the teaching process was identical in both venues.
A *primed students dataset* was gathered in a course venue of 2018. In this course venue, error descriptions that had been gathered and validated in the aforementioned previous venues were presented to students together with their homework assignments as refutation texts (Tippett 2010) aiming at priming the students.

As the teaching methodology was not the same for this course venue as for the previous venues, this dataset cannot be seen as part of a longitudinal study. 4 teachers identified systematic errors in homework submissions of 433 students in 28 exercises. After the course venue, students and teachers were asked to complete a questionnaire assessing their attitudes towards the approach.

The topics addressed in all three course venues were identical, the exercises were almost identical.

Note that the error identification was performed by the teachers in the three cases, yet partly for different purposes: The *base dataset* and the *validation dataset* were gathered purely for academic purposes after the courses had ended. The task of error identification while gathering the *primed students dataset* was incorporated in the teaching routine: Teachers could reduce their workload while correcting student homework by referencing to already collected error descriptions.

Software used. The social medium used in the case study is a document centred project-based learning platform. On that platform, projects are hosted which involve users (students and teachers) and in which documents are aggregated. Exercises (in the sense of "problem definitions"), homework submissions of students, teaching material and supplementary material, and descriptions of systematic errors (referred to on the platform as "labels") are represented on the platform as documents belonging to a project. Teachers can attach error descriptions to submissions if they find the corresponding errors in submissions. All documents can be commented: By students for example to ask questions regarding an exercise or an error description, and by teachers for example to give feedback on a submission. Figure 1 shows the platform's interface the teachers use while reviewing student homework submissions.

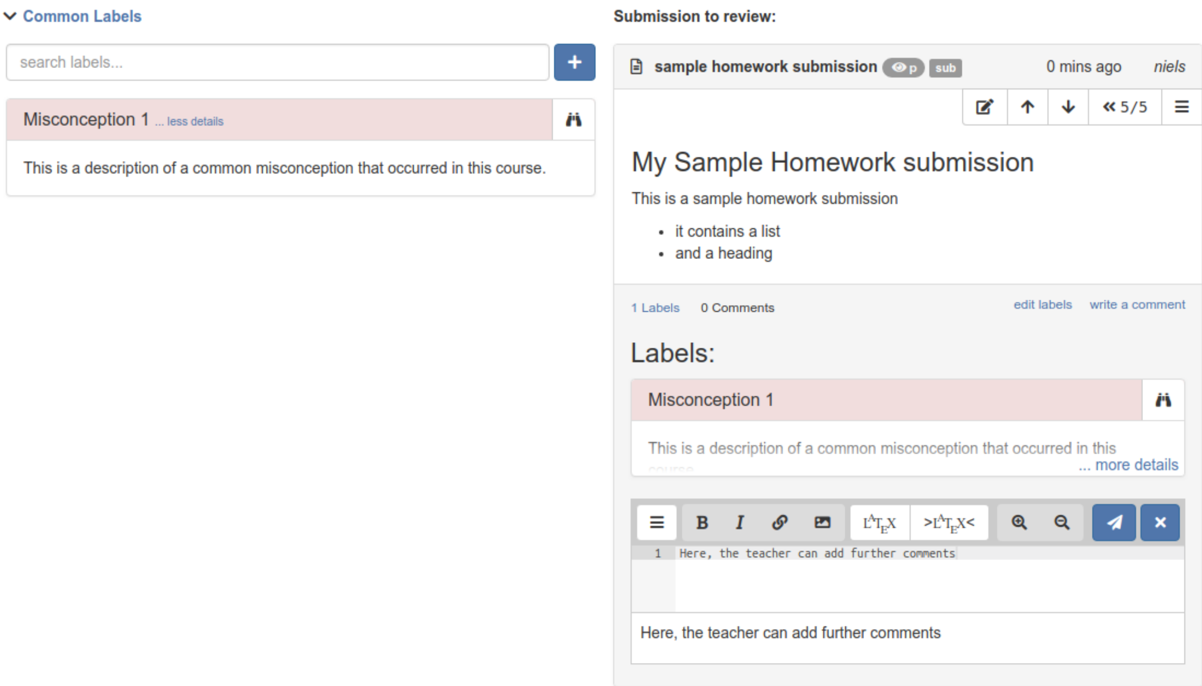

Figure 1: Teacher Interface for reviewing student homework: A list of common error descriptions on the left (containing only "Misconception 1"), which is already attached to the student submission on the right.

Data Analysis. The absolute frequencies of systematic error occurrences in the *base dataset* were compared to the absolute frequencies of systematic error occurrences in the *validation dataset* and in the *primed students* dataset using the Pearson correlation coefficient to determine whether systematic errors re-occur with similar frequencies in later course venues. Relative frequencies of error occurrences per exercise were compared for the *base dataset* and the *primed students dataset* using the exact Fisher test to determine whether students in the case study better avoided errors than in the previous course venues during which they had not been primed with error descriptions.

For each student participating in the course venue of 2018 an *activity index* was computed. The *activity index* is defined as the number of hours in which an action of the student was registered on the platform. Students were then grouped by their activity indices in four sets of equal size (quartile one to four) and average counts of occurrences of systematic errors per exercise were compared between the four groups.

The regularity of a students activity was measured by computing the standard error of time interval lengths between active hours, referred to as the *irregularity index*. A linear model estimating the average occurrence of systematic errors by the irregularity index was fitted to reveal the influences of work irregularity on error occurrences.

4. Results

In this section, the results of the case study are presented. First, data on behaviour changes, then data on the students' and teachers' attitudes are reported.

Indications for behaviour changes. The frequencies of error occurrences correlated positively for the *base dataset* and the *validation dataset* (k = 0.49), a smaller correlation was found when comparing the *base dataset* to the *primed students dataset* (k = 0.21) which indicates that the same systematic errors re-occurred with similar frequencies in the first two venues and that these frequencies decreased in the case study.
The relative occurrence frequency of errors significantly differed between the *base dataset*, where the average number of systematic errors per exercise was 35%, and the *primed students* dataset, where that number dropped to 19%.
A student's activity on the platform had a significant impact on her error occurrences. In the *primed students* dataset, students in the lowest quartile of activity made nearly as many systematic errors per exercise as in the *base dataset* (31% compared to 35%), while this proportion was only 14% for students in the highest activity quartile. Figure 2 Illustrates the decrease of systematic error rates with increasing activity.

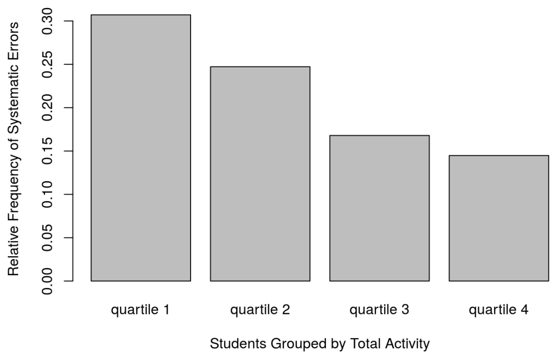

Figure 2: Relative frequency of systematic errors per exercise for different quartiles of activity.

To evaluate the impact of the regularity index (defined in the previous section) on error occurrence rates, a linear model was fitted. The model exhibited a small, yet significant, impact of the irregularity index on error occurrences ($p = 0.01$), predicting that reducing the irregularity index by 10 hours would reduce systematic error occurrence by 1%.

Student and teacher attitudes. Students were asked to indicate on a 6 point Likert scale ranging from "not at all" (1) to "absolutely" (6) whether the descriptions of systematic errors they had *a priori* been provided with had been helpful for their learning, had been helpful when received *a posteriori* as feedback, and if they would (in future courses) be interested in knowing the numbers of errors committed by their peers.
The results show a general perceived helpfulness of being provided with error descriptions before working on exercises (*a priori*), while, when received as personal feedback (*a posteriori*), they are perceived as as less helpful (Figure 3 right). This finding is supported by survey data comparing the *a posteriori* use of error descriptions with personalised comments: 35% of the students perceived error descriptions as less helpful than personalized comments, 11% as more helpful. The majority of students (54%) found both measures equally helpful (Figure 3 left).
57% of students reported to have consulted the provided systematic error descriptions before or while carrying out a homework exercise. Most students (54%) who reported never to have used systematic error descriptions while working on their homework rated these descriptions as less helpful than personalized feedback.

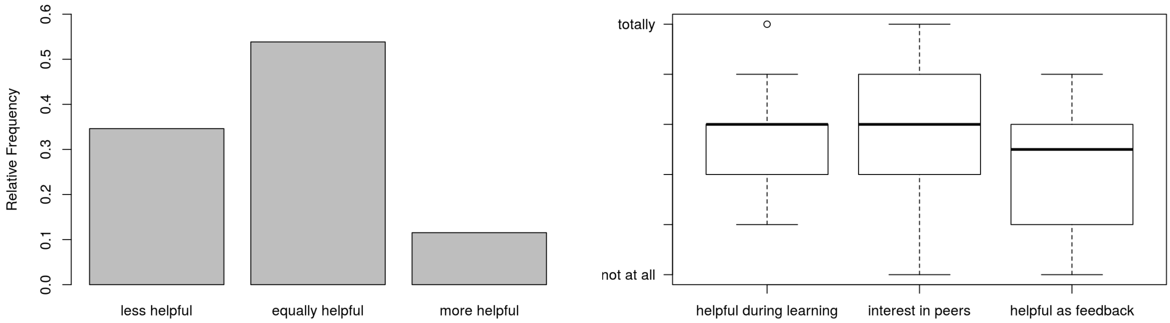

Figure 3: Student perceptions of systematic error descriptions. Left: compared with personalized comments, right: general perceptions.

Three of the four teachers who used the platform for correcting homework submissions participated in the survey. They found the platform's services for collecting and providing formerly collected systematic error descriptions helpful in their homework corrections (all three teachers reported a value of 5 on a 6 point Likert scale ranging from "not at all" (1) to "absolutely" (6)) and also reported that these services reduced their workload (values 4, 5, and 5).

The evaluation also revealed that the platform's usability should be improved ("the component is easy to use" yielded values 3, 4 and 3) with one teacher stating: "The implementation of the labelling function was a bit too simple and as a consequence removing an already entered label was a real pain."

The teachers also expressed that they would like to use in the future a software suggesting systematic error descriptions (values 4, 6, and 5) and that they would like to share the identification and collecting of systematic errors (collective agreement on value 6).

5. Discussion

The results presented in the previous section clearly point to the effectiveness at improving learning and easing teaching of using social media for collecting systematic errors and for sharing the collection of systematic errors between teachers and course venues. The students who were primed with descriptions of systematic errors were significantly more likely not to commit these systematic errors. The strong correlations of total activity and activity regularity with error avoidance are of special interest: The students with the lowest activity levels showed error occurrence rates similar to those of the students in previous course venues in which no error descriptions were provided. This suggests that students primed with systematic error descriptions learned from these error descriptions.

The students' attitudes towards the novel approach were mixed. The survey results clearly identify a perceived difference between "*a posteriori* receiving an error description as personal feedback" (which half of the students identified as helpful) and "*a priori* accessing an error description as supplementary material" (which more than half of the students perceived as helpful). This difference may result from the fact that the majority of students who did not use error descriptions while solving exercises preferred personal feedback: These students maybe did not draw much information from error description provided *a priori*. Interestingly, while every document on the platform could be commented by students (for example to ask clarifying questions), error descriptions did not receive any comments from students.

The teachers on the other hand seemed to have genuinely appreciated the novel approach (while pointing out improvements to the software): They all reported it to reduce their workload and to be helpful.

Limitations. The results presented above suffer from those limitations that are typical of case studies referring to consecutive venues of a same course.

First, the venues examined in this article were not completely identical: Parts of the teaching staff (the student tutors) changed over the years and a few exercises were slightly modified.

Secondly, the lower error occurrence rates that have been observed with the novel approach might result from the teachers having collected systematic errors for different purposes. For the *base dataset* and the *validation dataset*, the teachers collected systematic error description for academic purposes. For the primed students dataset, the teachers collected systematic error descriptions so as to enrich their written feedback. It

is possible that the teachers were in that last case less sensitive, or less willing to integrate systematic error descriptions into their teaching routine. The data collected do not allow to discard this possibility.
Lastly, it is worth stressing that the smaller rate of systematic errors observed during the third course venue – the venue in which the novel approach was used – does not necessarily imply that more students learned better than in the previous two course venues. Indeed, the quality of homework submissions has not been assessed. Yet, interestingly, none of the teachers added a novel systematic error description during the third course venue. This suggests that the collaborative collecting of systematic errors by a group of teachers using social media is effective not only at reducing the teachers' workload, but also at gathering comprehensive collections of systematic errors.

6. Conclusion and Perspectives For Future Work

This article has reported on a case study conducted in a university level computer science course. The case study investigated the use of social media specifically developed for a use in higher STEM education. The social medium made a novel approach to both learning and teaching possible. The aim of the approach was to exploit the potential of social media for reducing the occurrence of systematic errors in STEM and the workload of the teacher. An evaluation has demonstrated the effectiveness of the novel approach to higher STEM education. Limitations of the evaluation have been pointed out.
The results presented in this article are promising since, under the novel approach and using specific social media, the systematic errors committed by the students could be reduced by 16% and the workload of the teachers was reduced. In the future, personalisation in the form of suggestions which errors could be important to a specific student could be incorporated in the medium to further reduce occurrences of systematic errors: A previous evaluation has shown that the systematic errors a given student is likely to commit can, to some extent, be predicted (Heller, Bry 2018). Such predictions could be used for regularly sending personalized hints at those errors. Note that similar messages, including personalized predictions, have been shown to have a positive effect on learning (Arnold & Pistilli 2012).
In the current prototypical implementation of the social medium, systematic error descriptions take the form of so-called „refutation texts", that is, explanations and refutations of common fallacies. In future versions of the social medium, texts referring to systematic errors could encompass supplementary exercises and examples. Such additions could be made little by little by the teaching staff of future venues of similar courses using the same social medium, thus avoiding a hardly tractable one-time high workload. It is worth stressing that using social media in teaching not only eases the collaboration between teachers of a course but also between different course venues.
The use of social media described in this article seems specific to STEM education (where the terms "systematic error" and "systematic misconception" are most commonly used), yet, it could be applied to other fields without any changes. Language teachers, for instance, typically use error categories such as "wrong word order" or "wrong tense" when giving written feedback. These categories can be represented as systematic error descriptions on the platform. Furthermore, students could also use the system to actively learn. When learning any form of categorization (such as associating a painting with an epoch it was produced in or a body of text with a rhetorical fallacy present in that text), students could take the role of the "reviewer", associate documents with the category they see fit, and discuss their choices.

References

Arnold, K. E. & Pistilli, M. D., 2012, April. Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270). ACM.
Blaschke, L. M. and Hase, S., 2016. Heutagogy: a holistic framework for creating twenty-first-century self-determined learners. In *The future of ubiquitous learning* (pp. 25-40). Springer, Berlin, Heidelberg.
Brady, K. P., Holcomb, L.B. and Smith, B.V., 2010. The use of alternative social networking sites in higher educational settings: A case study of the e-learning benefits of Ning in education. *Journal of Interactive Online Learning*, *9*(2).
Confrey, J., 1990. Chapter 1: A Review of the Research on Student Conceptions in Mathematics, Science, and Programming. *Review of research in education*, *16*(1), pp.3-56.
Dabbagh, N. and Kitsantas, A., 2012. Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and higher education*, *15*(1), pp.3-8.

Gruzd, A., Haythornthwaite, C., Paulin, D., Gilbert, S. and del Valle, M.E., 2018. Uses and gratifications factors for social media use in teaching: Instructors' perspectives. *New Media & Society*, *20*(2), pp.475-494.
Hamid, S., Waycott, J., Kurnia, S. and Chang, S., 2015. Understanding students' perceptions of the benefits of online social networking use for teaching and learning. *The Internet and Higher Education*, *26*, pp.1-9.
Heller, N. and Bry, F., 2018, June. Predicting Learners' Behaviours to Get It Wrong. In *International Conference in Methodologies and intelligent Systems for Technology Enhanced Learning* (pp. 12-19). Springer, Cham.
Law, E. and Von Ahn, L., 2009, April. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1197-1206). ACM.
Lee, M.J. and McLoughlin, C., 2007. Teaching and learning in the Web 2.0 era: Empowering students through learner-generated content. *International journal of instructional technology and distance learning*, *4*(10), pp.21-34.
Movshovitz-Hadar, N., Zaslavsky, O. and Inbar, S., 1987. An empirical classification model for errors in high school mathematics. *Journal for research in mathematics Education*, pp.3-14.
Ophus, J. D. and Abbitt, J.T., 2009. Exploring the potential perceptions of social networking systems in university courses. *Journal of Online Learning and Teaching*, *5*(4), pp.639-648.
Parameswaran, M. and Whinston, A.B., 2007. Social computing: An overview. *Communications of the Association for Information Systems*, *19*(1), p.37.
Posner, G. J., Strike, K.A., Hewson, P.W. and Gertzog, W.A., 1982. Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education*, *66*(2), pp.211-227.
Quinn, A. J. and Bederson, B.B., 2011, May. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1403-1412). ACM.
Radatz, H., 1979. Error analysis in mathematics education. *Journal for Research in mathematics Education*, pp.163-172.
Rambe, P., 2012. Critical discourse analysis of collaborative engagement in Facebook postings. *Australasian Journal of Educational Technology*, *28*(2).
Rankin, M., 2009. Some general comments on the 'Twitter experiment.'. *University of Texas at Dallas*.
Sadaf, A., Newby, T. J. and Ertmer, P.A., 2012. Exploring pre-service teachers' beliefs about using Web 2.0 technologies in K-12 classroom. *Computers & Education*, *59*(3), pp.937-945.
Tess, P. A., 2013. The role of social media in higher education classes (real and virtual)–A literature review. *Computers in Human Behavior*, *29*(5), pp.A60-A68.
Tippett, C. D., 2010. Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, *8*(6), pp.951-970.
Verbert, K., Duval, E., Klerkx, J., Govaerts, S. and Santos, J.L., 2013. Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), pp.1500-1509.
Von Ahn, L., & Dabbish, L. (2004, April). Labeling images with a computer game. In *Proceedings of the SIGCHI*
Von Ahn, L., 2008, April. Human computation. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* (pp. 1-2). IEEE Computer Society.
Wieser, C., Bry, F., Bérard, A. and Lagrange, R., 2013, November. Artigo: building an artwork search engine with games and higher-order latent semantic analysis. In *First AAAI Conference on Human Computation and Crowdsourcing*.