A review of automated feedback systems for learners: classification framework, challenges and opportunities

Galina Deeva, Daria Bogdanova, Estefanía Serral, Monique Snoeck, Jochen De Weerdt

Research Centre for Information Systems Engineering (LIRIS), KU Leuven, Belgium

Abstract

Teacher feedback provided to learners in real-time is a crucial factor for their knowledge and skills acquisition. However, providing real-time feedback at an individual level is often infeasible, considering limited teaching resources. Fortunately, recent technological advancements have allowed for developing of various computer tutoring systems, which can support learners at any place and time by generating personalized feedback automatically. Such systems have emerged in various domains, tackle different educational tasks, and often are designed in very distinctive ways. Consequently, the knowledge in the field of automated feedback systems is rather scattered across different domains and applications, as we illustrate in this study. This paper aims to outline the state-of-the-art of recently developed systems for delivering automated feedback, and thus serves as a source of systematic information for educators, researchers in the educational domain and system developers. More specifically, the contribution of this study is twofold. Firstly, we offer an extensive literature review of the field. As a result of a rigorous selection process, consisting of 4 phases, a total of 109 automated feedback systems is selected for a detailed review and thoroughly classified against numerous dimensions. Secondly, based on a detailed analysis of the recent literature sources and following the design science research approach, a classification framework for automated feedback systems is developed. This framework is used to classify the selected systems, and thus give a detailed overview of the predominantly available educational technologies, the educational settings in which they are applied, the properties of automated feed-

---

Email address: `galina.deeva@kuleuven.be` (Galina Deeva)

back they deliver, and the approaches for their design and evaluation. Based on this analysis, several important observations and recommendations are put forward as an outcome of the study. In particular, our study outlines the current fragmentation of the field, discusses a need for a common reference framework, and calls for more personalized, data-driven and student-centered solutions to exploit a larger set of opportunities offered in this age of data.

A review of automated feedback systems for learners: classification framework, challenges and opportunities

Abstract

Teacher feedback provided to learners in real-time is a crucial factor for their knowledge and skills acquisition. However, providing real-time feedback at an individual level is often infeasible, considering limited teaching resources. Fortunately, recent technological advancements have allowed for developing of various computer tutoring systems, which can support learners at any place and time by generating personalized feedback automatically. Such systems have emerged in various domains, tackle different educational tasks, and often are designed in very distinctive ways. Consequently, the knowledge in the field of automated feedback systems is rather scattered across different domains and applications, as we illustrate in this study. This paper aims to outline the state-of-the-art of recently developed systems for delivering automated feedback, and thus serves as a source of systematic information for educators, researchers in the educational domain and system developers. More specifically, the contribution of this study is twofold. Firstly, we offer an extensive literature review of the field. As a result of a rigorous selection process, consisting of 4 phases, a total of 109 automated feedback systems is selected for a detailed review and thoroughly classified against numerous dimensions. Secondly, based on a detailed analysis of the recent literature sources and following the design science research approach, a classification framework for automated feedback systems is developed. This framework is used to classify the selected systems, and thus give a detailed overview of the predominantly available educational technologies, the educational settings in which they are applied, the properties of automated feedback they deliver, and the approaches for their design and evaluation. Based on this analysis, several important observations and recommendations are put forward as an outcome of the study. In particular, our study outlines the current fragmentation of the field, discusses a need for a common reference framework, and calls for more personalized, data-driven and student-centered solutions to exploit a larger set of opportunities offered in this age of data.

---

1. Introduction

Providing feedback is a critical factor for improving knowledge and acquiring skills. If such feedback is provided when learners need it, it can considerably improve the quality and speed of learning processes (Economides, 2005; Hattie and Timperley, 2007). However, students report serious deficiencies in the amount and quality of feedback they receive (Ferguson, 2011; Boud and Molloy, 2013). Delivering real-time feedback at an individual level is often infeasible, considering the limited teaching resources and the heterogeneous nature of the students' profiles (Boud and Molloy, 2013; Pardo et al., 2019).

Recent technological advancements have caused a widespread uptake of various technologies in the field of education, e.g. Intelligent Tutoring Systems (ITS), which are intended to produce personalized feedback in an automated way (Pardo et al., 2019). Such systems support learning processes at any time, given that electronic devices and internet connection are broadly accessible to a large number of people (Hwang, 2014). Throughout the history of computer tutoring, it was common to express certain concerns regarding its effectiveness in general, as well as in comparison to human tutoring. For example, Bloom (1984) argued that human tutors provide a larger difference in the learning gains compared to computer tutors. However, as the remarkable advancements of computer technologies in the last decades allow for developing better and more capable computer tutors, more recent works confirm that the effect of intelligent tutoring systems providing automated feedback is comparable to that of human tutoring (VanLehn, 2011). Furthermore, as Singh et al. (2013) highlight, technologies can offer quality education to numerous learners around the world. These advantages of technology-enhanced learning have led to more and more digital learning environments being developed and implemented.

Considering this wide variety of digital learning environments, there is a need to outline and analyze the current state of the field. Previously, there have been some studies that aimed to review the literature that tackles the topic of automated feedback technologies/systems (the terms "technology"

and "system" are used interchangeably throughout this paper and address the same concept). For instance, Normadhi et al. (2019) presented a systematic literature review on the most common personal traits used to create learner models within e-learning systems, as well as methods for identification of such traits, thus focusing on a specific aspect of automated feedback system architecture. From a theoretical point of view, a conceptual framework for feedback automation in smart learning environments, providing various classification categories, was proposed by Serral and Snoeck (2016) (see also a later work in Serral Asensio et al. (2019)). In another literature review, Bimba et al. (2017) presented an overview of adaptive feedback in computer-based learning environments, including its target, means, goal and strategy. Similarly, in a systematic literature review by Martin et al. (2020), adaptive strategies and adaptive learning technologies are analyzed, emphasizing the content, instructional models and learner models used in the literature. Furthermore, given that automated feedback is commonly used in the field of programming education, partially due to the fact that many aspects of programming can be formalized and thus assessed automatically, there is a substantial number of studies that provide a review of the systems designed for automatic grading (Caiza and Del Alamo, 2013), or automated feedback in programming exercises (Le et al., 2013; Keuning et al., 2018). Similarly, other literature reviews have been focused on tutoring systems within a particular domain or educational task, such as teaching computer graphics (Suselo et al., 2019), science education (Zacharia et al., 2015) and foreign language learning (Golonka et al., 2014).

These reviews provide useful information on different aspects of automated feedback systems. However, as they are focused either on particular educational topics or on specific domains, a global perspective is missing. Having a global perspective is important for anyone involved in the development of an automated feedback system. Such systems are usually initially developed or piloted in a specific application domain, tackle different educational tasks, and often are designed in very distinctive ways. A review that is tailored to a specific application domain (e.g. programming), or focuses only on a limited set of properties of such systems holds the risk of missing interesting insights from systems developed in other domains and for other applications. This could lead to reinventing the wheel. Moreover, bringing together the knowledge on designing automated feedback systems that is currently scattered across different domains and applications could be beneficial for system developers, and might be a source of inspiration and

cross-fertilization. For that reason, this study aims to approach a review of automated feedback systems in a more general way without focusing on a particular domain and looking into a large number of different aspects, ranging from the educational context to the evaluation of the system effectiveness, and thus develop a general classification framework for these systems. The importance and utility of such a framework is summarized as follows. First, by looking at systems across different domains, system developers could potentially get inspiration in using new methods and approaches typical for a certain field and less common in another. Moreover, certain domains might have their own standards when describing an automated feedback system, and, as such, a more general standard for reporting such systems would improve systems' comparison.

As such, the core contributions of this study are:

1) the development of a classification framework of automated feedback technologies;

2) a systematic literature review of automated feedback technologies in the period from 2008 to 2019;

3) a detailed overview of the predominantly available automated feedback technologies, the educational settings in which they are applied, the properties of automated feedback they deliver, and the approaches for their design and evaluation.

The findings of this study will be relevant for educators, researchers in the educational domain, as well as developers of new automated feedback systems, both as a source of inspiration for the systematic development of new systems and as an introduction to the field, especially for interdisciplinary researchers and educational technologists.

The remainder of this paper is organized as follows. In Section 2, the survey methodology is presented, including the research questions, the selection criteria and the search process. Next, Section 3 introduces the developed classification framework of automated feedback technologies. Consequently, Section 4 provides a classification of the selected papers according to the proposed framework. Finally, Sections 5 and 6 discuss the main findings, implications and limitations of the study, as well as provide general recommendations.

2. Research methodology

The methodology of this paper is based on the guidelines for performing systematic literature reviews proposed by Kitchenham (2004). Following these guidelines, in this section we describe the designed review protocol, which includes the research questions, the search process, the selection criteria and the selection process. Subsequently, the process of framework development based on design science research is discussed in detail. Finally, the stage of framework validation is presented.

2.1. Research questions

The literature review was guided by the following research questions:

- **Research Question 1:** What kind of automated feedback systems were predominantly developed in the field of education over the last decade?

- **Research Question 2:** Which dimensions are most relevant and important for developing a classification framework of such systems?

- **Research Question 3:** For each of those dimensions, which common categories of systems can be defined?

It is important to note that Research Question 1 implies that this review is focused exclusively on the papers describing fully functional and already implemented systems providing automated feedback rather than prototypes, concepts or theoretical frameworks. This restriction allows us to focus on a narrower set of systems, and discover a decent and substantial selection of practical works that show complete implementations of automated feedback systems and serve as a source of inspiration for further research and developments in the area. The eligibility criteria are described in subsection 2.3.

2.2. Search process

To search for scientific papers in a systematic way, we used Scopus[1], which is considered the largest abstract and citation database of peer-reviewed literature and contains over 75 million records (updated November 2020). Scopus

---

[1]https://www.elsevier.com/solutions/scopus

covers a large majority of specific databases, such as IEEE Explore, ACM database, Springerlink, Elsevier, but also meta-databases such as Web of Science. Several independent searches in other large databases, such as Google Scholar and Web of Science, were performed as well for a comparison before the final decision was made. The results have shown that Scopus indeed is capable to find the largest set of papers, which also to a great extent includes the results found by other databases. Hence, in the remainder of the paper we report the results of the Scopus search. The additional reason for using Scopus for the search is that it allowed to construct a flexible and detailed search query, as further discussed in the paper.

To construct a search query that can yield the most relevant and complete list of papers, we performed several search iterations and extracted various synonyms denoting technologies associated with feedback delivery in education, e.g. Intelligent Tutoring Systems and Smart Learning Environments. Since the names of some educational systems consist of similar word collocations, e.g. Digital Learning Environment and Interactive Learning Environment both contain "learning environment", we shortened such collocations to make the search more general. Next, we collected frequently used synonyms of "education" to restrict the scope of our search to educational systems. After several possible combinations of the keywords were examined, they were finally combined in the following way, which we believe to provide an optimal trade-off between finding all the relevant and not too many irrelevant studies:

1. The word "feedback" should be present in the title, since we focus on systems that are capable of providing feedback to learners. This restriction allows us to focus our search on the systems that specifically tackle the topic of automated feedback. The word "feedback" was chosen as it is a well-established and commonly accepted term across the fields involved in the educational studies. Additionally, we analyzed the possibility of including to the search query synonyms of the word "feedback", such as "hint", "recommendation", "help", "suggestion" and "advice". However, this analysis revealed that their potential inclusion either in the title or in other searchable fields would drastically increase the number of retrieved studies. Furthermore, despite the higher number of papers, the large majority of newly retrieved papers are totally irrelevant with respect to the scope of our research, and the number of papers worthwhile to test against the exclusion criteria is

6

deemed insignificant.

2. An educational technology-related keyword should be present in either the abstract, title or keywords.

3. A synonym of the word "education" must appear in one of the searchable fields (title, abstract, references, keywords, issue, volume, publisher, etc.).

The list of the keywords and the search query are presented in Appendix in Tables A.9-A.10. Although the query could have been generalized even further to return more papers that match our criteria, this adjustment would also have yielded a large number of irrelevant results, as has been shown by several search iterations with different combinations of terms in the query that were performed at the preparatory stage of the study. We believe this query is capable of finding a sufficient number of relevant papers for identifying the state of the art of the proposed topic. By "sufficient" we mean that adding papers would not lead to 1) a substantial refinement of the framework or 2) different conclusions on the current state-of-the-art.

2.3. Inclusion and exclusion criteria

The review process included four phases, as indicated in Figure 1. In the first phase (performed in 2018), the papers were preselected for a detailed review if they met the following *inclusion criteria*:

1. The paper is published in the period from 2008 to 2018.
2. The paper is published in English.
3. The paper presents a system that is able to provide feedback to learners.

Next, in the second phase, the preselected papers were reviewed in detail, and the final selection of the papers was made. In the selection process, we have opted to include the papers that present fully implemented, functioning e-learning platforms that are capable for providing feedback in an automated way. The systems that are not fully implemented (e.g. theoretical methods for providing feedback) and the systems with unclear architecture (i.e. that are not described in sufficient details to understand the way they work) have been excluded. Finally, the methods for providing overly simplistic feedback were excluded as well due to their very simple architecture or a lack of architecture and/or implementation; e.g. prototypes used to correct students' answers to multiple-choice questions and inform them whether their answer is right or wrong.

As such, the *exclusion criteria* are summarized as follows:

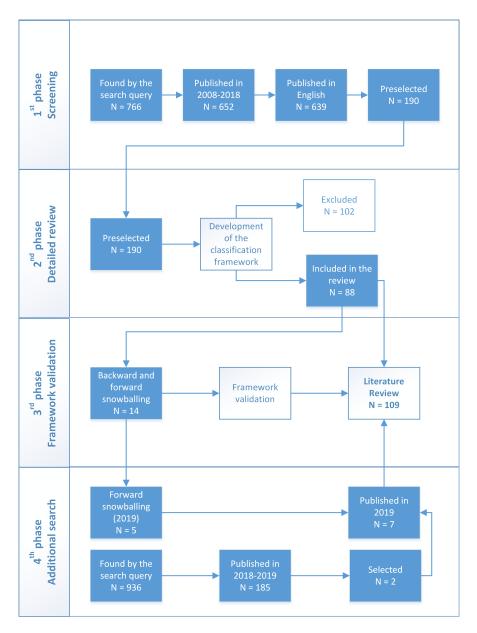Figure 1: An overview of the review process, consisting of four phases: screening, detailed review, framework validation and additional search. During the third phase snowballing procedure, 17 new papers were found, from which 14 depicted new systems and 3 papers presented one of the 88 systems. The additional search was appended to update the set of papers with the most recent works published in 2019.

1. When there are several papers describing the same system, we selected the most recent or the most complete report, i.e. the one providing the most details regarding the system architecture. Consequently, our study presents one paper per system. Potentially, in situations when a system architecture has evolved significantly over time that it can be viewed as a different system from the design perspective, two papers per system could be included as well.
2. The proposed approach for providing automated feedback is solely theoretical, has not been fully implemented, or only describes an initial prototype of a system that requires significant improvements before it can be used by learners.
3. The study does not provide sufficient details to understand the system architecture and design, e.g. solely mentions a system or describes a research in which the system is used, without specifying the way the system functions.
4. The feedback provided by the system is not fully automated, i.e. it requires teacher or peer contribution.
5. The paper reports on methods for providing overly simplistic feedback, such as prototypes used to correct students' answers to multiple-choice questions and informing them whether their answer is right or wrong.

The selection phases and the corresponding number of studies selected after applying the eligibility criteria are summarized in Figure 1. First, a total of 766 papers was retrieved by the search query. Next, in the screening phase, the titles and abstracts of these 766 papers were read and the inclusion criteria were applied, which resulted in 190 papers being preselected for a detailed review. The selection process was performed by one of the authors after an in-depth discussion and agreement with the other co-authors on the exact selection criteria. The fact that the preselection was performed by the same author ensured consistency of the selection process. Next, an in-depth review of the common categories presented in the preselected papers served a basis for development and refinement of the classification framework. This stage of the framework development was conducted with involvement and in close collaboration with all co-authors. Finally, after the detailed review of the preselected studies and applying the exclusion criteria, 88 papers were selected. To ensure the reliability of the second selection stage, all the preselected papers, as well as their categorization, were examined by two co-authors. In case of disagreement between the two co-authors, other

co-authors were consulted.

2.4. Development of the Classification Framework

In this subsection we describe the design methodology for developing TAF-ClaF (Technologies for Automated Feedback - Classification Framework).

It is important to note that various existing frameworks for feedback classification exist, aiming at different properties of feedback. For example, Fleming and Levie (1978) focused on the purpose (function) of feedback, Hattie and Timperley (2007) proposed a model based on levels, such as task, process, self-regulation and self levels, and Voerman et al. (2012) proposed to distinguish between progress and discrepancy feedback. However, the majority of existing classifications were developed for educational feedback provided by real teachers (human tutoring) rather than by automated feedback systems (computer tutoring). Some authors proposed feedback categories specific to adaptive automated feedback (e.g. Economides (2005), Narciss (2008), Serral and Snoeck (2016)), but these frameworks do not explicitly focus on the technology aspect nor on architectures of the systems. Moreover, when trying to apply the existing frameworks to the systems described in the selected papers, we faced certain difficulties, such as systems that are not publicly available, which would make the classification by content unreliable, and the fact that some categories of feedback were specific to the classroom setting and not to e-learning systems. Ultimately, no classification framework listed above was focusing on aspects that are specifically important for system design and for computer tutoring as opposed to human tutoring. Therefore, following the cases where certain dimensions were missing, TAF-ClaF has been developed.

TAF-ClaF was developed in two iterations using the design science research approach introduced by Hevner et al. (2004). Conforming to the design science research approach (Hevner et al., 2004) and checklist (Hevner and Chatterjee, 2010), three cycles of activities were performed, as shown in Figure 2, namely the relevance cycle, the rigor cycle, and the design cycle. The relevance cycle comprises the initialization of research, i.e. the problem that needs to be addressed and the means of solution evaluation. Next, the rigor cycle embodies the existing methods or frameworks helpful to construct and evaluate the solution. Finally, the design cycle is the key component of design science research, representing the process of solution development and its evaluation.

In the first iteration, we developed the preliminary version of TAF-ClaF using the matrix-approach as suggested by Webster and Watson (2002); in particular, a literature review should be concept-based rather than author-based, and thereto papers are mapped against important concepts in a matrix. To find the concepts, a grounded theory approach can be used. Rather than performing the qualitative analysis of the papers using a fully "open coding" approach, we started from the high-level (and very generic) coding scheme proposed by Bandara et al. (2011). In this particular case, we focused on the proposed codes "Characteristics" (Architecture, Feedback), "Research Methods" (Evaluation) and "Contexts" (Educational Context). Those basic codes were gradually refined as papers were added to the study. As advised by Bandara et al. (2011), each category was further analysed for deeper insights and/ or when more than 5 categories were created in the coding scheme. Following the advises of the same authors, a two-level coding was preferred, the second level of analysis allowing to derive deeper meaning. The diversity of terminology used by different authors makes finding good code names more difficult. For the characteristics of "Feedback" we based the code names on 1) the highly cited Hattie and Timperly framework (Hattie and Timperley, 2007) and 2) a work dedicated to a detailed categorization of educational feedback by Serral and Snoeck (2016). The categories used in the existing frameworks were reviewed for their applicability to automated feedback technologies, and thus an initial set of categories was derived and extended with the ideas of the authors. Based on the content of the selected studies and the properties/categories identified in the two aforementioned frameworks, the following feedback properties were retained: its degree of adaptiveness (degree of personalization), the timing, the degree of learner control over feedback and the purpose of feedback. When reading the papers, concepts from the papers were coded according to the categories above if possible, and relevant concepts that could not be matched, gave rise to new categories. Resulting categories were grouped into more general concepts until a first version of the framework was obtained.

The first version of the classification framework was applied to 190 studies preselected in the first phase of the screening. Despite its relative usefulness for paper analysis and categorization, some of the dimensions were found not to be very informative, and some important aspects were missing. In the second iteration, the classification framework was further improved by adding the missing dimensions and excluding those that were not informative. As a result, the following changes were introduced:
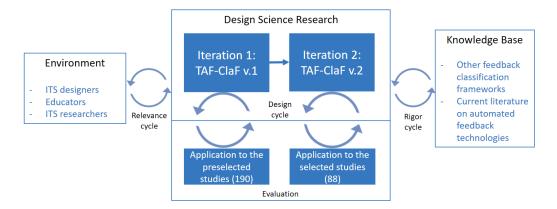
Figure 2: Design science research methodology for developing the classification framework

1. Excluded dimensions: level of feedback (according to Hattie and Timperley (2007)), nature of feedback (positive/negative), interactive device, feedback delivery (was substituted by *timing*), and educational task (was combined with *domain*).
2. Added dimensions: evaluation method, group size, statistical testing, and student data.

The new version of the classification framework was applied to the final selection of papers from phase two (88 papers) and revealed that saturation was obtained: the coding of the 88 papers did not give rise to new categories, not to modification of the categories or the framework.

2.5. Snowballing and framework validation

Following the first two phases of the review process as shown in Figure 1, a third phase was executed, consisting of snowballing and framework validation. The snowballing was performed based on the guidelines described by Wohlin (2014). The purpose of the snowballing within our study is twofold. First, it aims at finding those papers that were for one or another reason not extracted using the search query, and thus avoiding a bias introduced by the way the query was constructed and by the choice of the database. Thus, it ensures that the set of the reviewed papers is close to a complete set of the papers published on this topic. The second purpose of the snowballing is to validate the designed framework by showing that no dimensions needed to be added after additional papers were reviewed.

The numerical results of the snowballing process are summarized in Table 1. A total of 4589 papers were retrieved as either cited by the selected papers (backward snowballing) or citing the selected papers (forward snowballing). After reviewing the titles, and when needed also abstracts of these 4589 references, 134 papers were found potentially relevant to our study, 54 of which were already processed in the previous stages of the review process (49 included and 5 excluded). Out of the rest of the papers, only 17 were considered relevant, and in 3 of these the reported systems were already included in our list. Since all 3 were fuller versions, they substituted the previous references. As can be derived from the results, the majority of the systems that were eligible for our study during snowballing were already on our list (79%), and only 14 new papers needed to be added, which amounted to a 16% increase in the number of papers. This indirectly validates the efficacy of the search query that was used in this study, as it successfully retrieved the majority of the relevant papers.

As shown in Figure 1, the 14 newly added studies (and 3 substitutes) have been added to the selected papers and classified using the developed framework, resulting in 102 papers being selected for the period of 2008-2018. During the classification of the 14 papers, no additional dimensions needed to be added, which serves as a validation of the framework.

Table 1: The results of the snowballing process

	Backward	**Forward**	**Total**
Number of references	2936	1653	4589
Relevant	61	73	134
Previously excluded	5	0	5
Previously selected	21	28	49
Excluded	25	38	63
Substitutes	1	2	3
Included	9	5	14

2.6. Additional search

Since the original search was performed in 2018, in 2020 we updated the selected set of papers with the most recent works by conducting an additional search of papers published in 2018-2019. This is illustrated in Figure 1 as the fourth phase of the review process. The search query retrieved 936 papers, including those 766 papers discovered in the first phase. Of these 936 papers, 185 were published in 2018-2019. After an initial screening, 2 of these papers were added to the review. Additionally, as a result of forward snowballing

conducted in the third phase, 5 relevant papers published in 2019 were revealed. Thus, a total of 7 new papers from 2019 were added to our review. As such, the final selection consists of 109 papers.

3. Technologies for Automated Feedback - Classification Framework (TAF-ClaF)

In this section we discuss the classification framework (TAF-ClaF) developed for the purpose of this study, and thus address the Research Questions 2 and 3.

The classification framework is shown in Figure 3. It consists of four main components: architecture, feedback, educational context and evaluation, each characterized by a number of dimensions. The first component, **architecture**, comprises essential elements of a technical implementation of a system with capability of providing feedback, such as algorithms for feedback generation and required input data. Second, the **feedback** component covers the properties of automated feedback that are important to take into account during development of an automated feedback technology. Third, as the nature of the educational task and domain has potential influence on the way the feedback needs to be provided, we look into a spectrum of potential domains in the **educational context** component. Finally, after developing an automated feedback technology, it is crucial to validate its effectiveness using some sound evaluation methods. Therefore, the fourth component, **evaluation**, describes the typical methods (both quantitative and qualitative) for validating the quality aspects of an automated feedback system, the group size and, when relevant, the methods used for validating the statistical significance of the results. Each of these components are discussed in detail below.

3.1. Architecture

After the analysis of a large set of available technologies, we derived the typical architecture of an automated feedback system illustrated in Figure 4. Most systems follow a scheme where feedback is being deduced from three main elements: *domain model*, *expert knowledge* and *student data*. For automated feedback to be provided, several questions must be addressed, namely: "What?" kind of feedback to provide, "How and to whom?" it should be provided, and "When?" we provide it. The question of "What?" is addressed by the *domain model*, which is a structured representation of
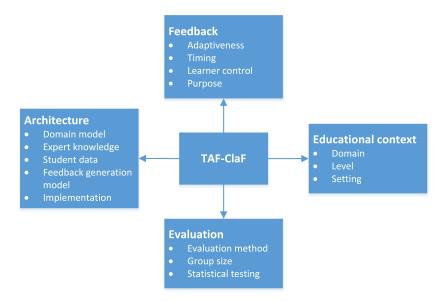
Figure 3: The Technologies for Automated Feedback classification framework (TAF-ClaF), consisting of four components, each characterized by a number of dimensions

domain knowledge. It can come in the form of a simple knowledge base, or be modeled in a more sophisticated way, e.g. as learning items categorized using Bloom's taxonomy.

The questions of "How and to whom?" are tackled by the *feedback generation model*, i.e. the feedback engine, that can be data-driven and/or expert-driven. In case of a data-driven feedback engine, the rules for feedback delivery are derived from *student data* by means of some data mining algorithm, which we call a "technique". In case of an expert-driven feedback engine, the rules for providing feedback are exclusively dictated by experts in the field and/or educational theory, which together are viewed as *expert knowledge*. For the *mixed* feedback engines, the rules are defined based on a combination of expert knowledge and insights derived from student data.

With the domain model and the feedback generation model it is possible to know what type of feedback should be provided in a certain situation. However, we are still in need to address the "When?" question, or, in other words, to know the situation of the student that triggers a need for feedback. This input, i.e. *student data*, can come in the form of a student answer, behavior or even as certain personality traits.

Finally, the ways automated feedback technologies are developed and
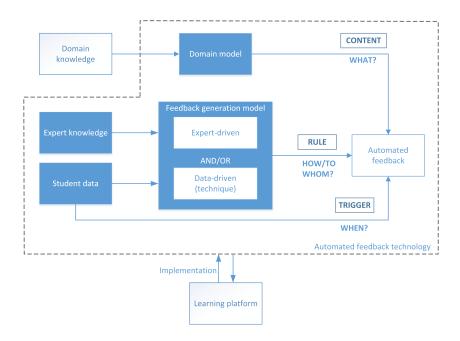
15

Figure 4: An abstract representation of a typical automated feedback technology

integrated into learning platforms are discussed under the *implementation* dimension. All of these dimensions are elaborated below.

3.1.1. Domain model

A domain model is a structured representation of domain knowledge, i.e. knowledge about the subject being taught, and specifically the subset of domain knowledge used to address the question of "What kind of feedback to provide?" in this specific system. For instance, a system that is used to teach mathematics to high school students will only use a certain subset of the mathematics domain which is deemed to be necessary for these students to know. Moreover, this knowledge might be further structured according to the program of the course or according to the view of the teacher on the sequential order and amount of information required.

3.1.2. Expert knowledge

It is easy to confuse expert knowledge with domain knowledge, especially since some systems might use different terminology for these elements. In our framework, expert knowledge is viewed as certain information coming from teachers and/or educational theories that either explicitly provide a set

of rules for feedback generation or are used for constructing such rules. In other words, it is defined as information used for "mapping" domain knowledge with potential student inputs. Thus, a math teacher that is developing the aforementioned system for teaching mathematics to high school students might follow his experience and define a list of typical student mistakes, which would be an example of expert knowledge used for expert-driven feedback generation. This expert knowledge will be potentially transformed to a set of rules (a feedback generation model), such as "if a student's answer is 5/10 instead of 1/2, the system should give a suggestion to repeat the section 17 in the textbook with an explanation about reducing equivalent fractions to the lowest terms".

3.1.3. Student data

In a typical automated feedback system, student data can be used for two different purposes. First, it can serve as an input for building a data-driven (or mixed) feedback generation model. In this case, rather than being based on the rules defined by the experts, a feedback generation model is based on learning analytics, i.e. the rules are derived from the data. For instance, if a sequential pattern analysis has indicated that successful students tend to solve mathematical problems in a certain order, it might be desirable to provide a confused student with a suggestion to follow such an order. This usage of student data only occurs in data-driven and mixed systems.

The second purpose is using student data as an indicator of a current situation that triggers feedback, which is also described above as an answer to the "When?" question. For instance, in case of the error-based feedback, the system needs to receive a wrong student answer as an input in order to know which feedback rule to trigger. Similarly, if a feedback generation model states that "after a student creates 10 new objects in a UML diagram, the system should remind him/her to simulate the created diagram", we need to take into account behavior data of students to know how many objects they created in a certain moment. Independently from the type of feedback, it is always required to have some student information. It is important to note that the student data used for the purpose of building a feedback generation model may often be different from the data used as a feedback trigger.

Types of student data vary depending on the type of learning task and the overall architecture of the system. Nevertheless, certain types of student data can be identified in the existing literature:

- *Predefined answer*, e.g. answers to multiple choice questions or fill-in the blank tasks.

- *Open answer* includes textual answers to open questions and tasks (e.g. programming, essay tasks, etc), as well as graphical answers, such as presentations or pictures.

- *In-system behavior*, i.e. log data: student activity (clicks, choice of tasks, log-ins) within the e-learning system.

- *Real-life behavior*: movements/audio/video datastream from the student.

- *Student characteristics*: skills level, student model, learning style or affective state pre-assessment data gathered by means of a survey.

3.1.4. Feedback generation model

A feedback generation model comprises a set of rules or a model for feedback delivery. Depending on the way this model is built, we distinguish between the following feedback generation engines:

- *Data-driven*, in which the model is derived purely from student data.

- *Expert-driven*, for which only expert knowledge serves as a source of feedback delivery rules.

- *Mixed*, when both expert knowledge and insights derived from student data determine the rules for providing feedback.

Moreover, in our analysis we also look into a *"technique"*, which is a specific data analysis algorithm that is used in data-driven and mixed systems, as indicated in Figure 4. In Section 4 we give an overview of the common techniques used in the literature.

3.1.5. Implementation

In this dimension we look into three aspects of systems' implementation: *language/tool* used for implementation, *web-based* vs *offline* implementation, and a system's *connection with the learning platform*. Some automated feedback technologies are not designed as standalone applications, but rather are built on top of existing learning platforms or other tools. For instance,

some authors develop a system which gives feedback while students are using the Eclipse environment to program with Java. Another example is a system that provides feedback to essays written in MS Word. Such systems are often implemented as plug-ins to existing software, which is illustrated in Figure 4 by the fact that a learning platform is shown as being separate from an automated feedback technology. To address such a distinction, we subdivide automated feedback technologies into two categories depending on the type of interaction with a learning platform:

- *Standalone*, if the feedback generation technology was developed together with the learning platform and is its undivided part.

- *Plug-in*, if the technology is developed to be used with an existing learning platform or software, and is developed by different authors and separately from such software.

3.2. Feedback

The range of educational tasks presented in the literature is wide, encompassing tasks of different levels of complexity - from solving multiple choice quizzes or basic programming exercises to dancing, performing surgery or manipulating a robotic arm. Such a variety of tasks and contexts implies a certain variety in types of feedback properties.

3.2.1. Adaptiveness

By the degree of adaptiveness of automated feedback, the presented systems can be classified into the following categories:

- *Non-adaptive*: the feedback system is considered to have no adaptiveness if before providing feedback it only takes into account the current discrete action of the student without analyzing the previous actions, or, in other words, if two students with a different background, preferences and history of actions receive exactly the same feedback on any action they perform.

- *Task adaptive (adaptive to student solution)*: these systems are characterized by the condition that the learning task for which the feedback should be provided has an infinite set of correct solutions (for example, a written essay task) that depend on the individual properties of a student. For example, a dance teaching application is an example of a system adaptive to student solution.

- *Student adaptive (adaptive to student characteristics)*: the feedback in such systems is given taking into account individual characteristics of the student - either determined through a questionnaire, or derived from the logged behavior.

3.2.2. Timing

The timing of feedback was subdivided into three main categories:

- *Immediate* or *action-based* feedback includes feedback provided right after certain actions or a set of actions, or after the system detected an error in the student's answer.

- *On request* feedback is the feedback asked by the student, which means that the student has to make some effort to get the feedback (for example, to click on a hint button).

- *End-of-task* feedback is provided after the learning task has been finished and submitted.

Many systems had multiple types of feedback in terms of timing, thus they were categorized by several terms at once.

3.2.3. Learner control over feedback

Some of the aspects of feedback can potentially be changed by the learner. The "learner control over feedback" dimension indicates how fully the learners can control, or in other words, change the feedback they receive. The following feedback aspects can be potentially controlled: amount (e.g. receive more feedback, request extra feedback, ask for less feedback if already mastered the concept), type (switch between audio, text, picture, video feedback), timing (receive feedback immediately or at the end of task), appearance (e.g. font, the way a talking avatar who gives advice looks like, color), and frequency (how often the feedback is received). All the aforementioned aspects might somewhat intersect with one another, e.g. requesting extra feedback could be considered as changing its amount, its timing, or its frequency, while changing the type of the feedback from visual to audio feedback might be considered as changing its appearance or its content.

Depending on the availability of such modifications to learners, the degree of learner control over feedback can range from no control to high level of control:

- *No control* over feedback: a student cannot choose whether to get feedback or not, and cannot set any properties of the feedback.

- *Mild control* over feedback: a student can either only request (i.e. explicitly ask for by pressing a certain button) feedback, or only adjust the feedback appearance, without any wider possibilities regarding the change in the feedback type or the amount of feedback received.

- *High control* over feedback: a student can determine whether to receive feedback and decide on the appearance/type of such feedback.

The "learner control over feedback" dimension is a meta-dimension that describes whether and to what extent a student can control different aspects of feedback, including e.g. feedback timing. Therefore, some types of feedback timing would be correlated with certain degrees of learner control, but these two dimensions certainly are not redundant and are not always possible to derive one from another. For instance, on-request timing will not be encountered in systems with no learner control. However, mild control over feedback does not automatically imply on-request timing, as it is possible that other aspects of feedback are the subject to student control, such as feedback appearance or frequency, and not the timing. On the contrary, on -request timing would imply some degree of learner control, but it is not immediately clear whether it is mild or high just by looking at the timing.

Having a higher degree of control over feedback can have potential positive influences on learner's performance (Perlmuter and Monty, 1977; Pridemore and Klein, 1991). Even very subtle adjustments, such as changing the font, might represent a very important change for the learner. For example, the learner might have vision problems and therefore require font adjustments. Another option is that the learner is simply annoyed by the current representation of the feedback, and a simple color/contrast/font change can significantly improve the learner perception (and thus also reception) of the feedback. One can imagine possible examples of more sophisticated feedback changes that might be crucial for certain learners. For example, in e-learning systems that use a talking avatar that gives direct hints to the learner, changing the appearance of such an avatar could significantly facilitate the leaners engagement, while changing quantity of feedback to the exact amount the learner wants could transform the experience of using an e-learning system from annoying to enjoyable.

3.2.4. Purpose of feedback

The classification of feedback by purpose was partly adapted from Serral and Snoeck (2016), and consisted of the four general types of feedback that could be found in the analyzed systems: informative, corrective, suggestive and motivational. To avoid misunderstandings of the terms, as different studies provide different definitions of the types of feedback, we will provide short definitions relevant for this particular analysis.

- *Corrective* feedback implies that it directly informs the student on correctness or incorrectness of his answer or course of actions.

- *Suggestive* feedback comprises hints and leading questions, as well as suggestions on particular course materials that have to be revised.

- *Informative* feedback implies non-suggestive and non-corrective messages that provide additional information on the performed learning task or the studied subject. For example, feedback in the format of "fun facts" or quotations from the textbook is considered informative. The difference between suggestive and informative feedback is the orientation of such feedback towards a particular task and the presence of call for action. Suggestive feedback implies that the student would perform certain action as a result of receiving such feedback. Informative feedback does not imply that any action would be performed upon receiving it: it is introduced to facilitate students deeper involvement in the task or subject.

- *Motivational* feedback is a type of feedback aimed at motivating the student to making more effort or to continue working on a problem, by means of praising or reassuring messages.

It is important to note that feedback can be of multiple types, e.g. both corrective and informative (i.e. explanatory). A multiple choice question that provides explanation next to the "True/False" of the given answer is an example of both corrective and informative feedback.

3.3. Educational context

One of the key features characterizing learning environments implementing automated feedback is educational context. For this study, we consider that educational context includes: educational domain (the subject or field

for which the system is used), educational setting ranging from in-class learning to pure online learning, and the targeted educational level.

3.3.1. Domain

As more and more domains are employing automated feedback for education, the list of domains may change and augment. Based on the analysis of the preselected literature, the following commonly occurring general educational domains were identified:

- *STEM*: Science, Technology, Engineering and Mathematics, including computer science and programming.

- *Languages*: all disciplines related to language learning and foreign language learning in particular.

- *Medicine*: all disciplines related to medical studies, both theoretical and practical.

- *Art*: performing arts (music, theater, dance) and visual arts (painting, photography, etc).

- *Special education*: disciplines related to tackling individual needs of students with disabilities, for example, sign language learning.

- *Other skills*: various skills that could not be categorized in any of the above categories.

- *Non-domain-specific*: systems not aimed at any particular domain, and can be used universally.

3.3.2. Level

For the classification of the systems along the addressed levels of education, we utilized the following categories:

- *Primary* education: primary or elementary school level, with approximate age of children between six and twelve years old.

- *Secondary* education: the second stage of school education, with approximate age of children between twelve and eighteen years old.

- *Tertiary* education: postsecondary education, including universities and other institutions of higher education.

- *Adult* education: additional training of adults outside of the university context, such as workplace education and training.

- In addition, *non-level-specific* educational systems could be identified.

3.3.3. Setting

Three types of educational settings were specified:

- *Online*: purely online educational settings without presence of any in-class activities related to the taught discipline.

- *Blended*: a mixture between online and in-class activities, that may vary in proportion.

- *In-class*: purely in-class use of the e-learning system, with an opportunity to get some support from a teacher.

- *Not specified* or *non-setting-specific*: systems for which the educational setting was not specified or that can be used in various settings.

3.4. Evaluation

Following the stage of development of an automated feedback technology, it is crucial to validate its effectiveness using some sound evaluation methods. Therefore, we find evaluation to be an important component of the framework. Specifically, we look into typical evaluation methods, the number of students on which the systems were evaluated, and the usage of statistical testing.

3.4.1. Evaluation method

The most common evaluation methods found in the literature are described below. In the majority of cases, these approaches are used in certain combinations to complement one another, or simply several evaluation techniques are applied. For example, for a control group experiment, it is typical to either compare a performance of a treatment and control groups with a pre-test/post-test scenario or with student questionnaires.

- *Student surveys*: surveys or sometimes interviews, e.g. based on a Likert scale, in which students evaluate their perception of the system and/or the improvement of their knowledge after using the system. Sometimes more specific aspects are evaluated as well, such as changes in frustration levels.

- *Control group experiment*, in which the students are separated in two groups, and the performance of a group without treatment (control group) is compared with the performance of the students that received automated feedback.

- *Comparison of pre-test and post-test results.* The students make a test before and after receiving automated feedback, after which the results of these tests are compared.

- *Analysis of student behavior.* Analysis of student reactions to feedback through real-life observations or log files analysis. The influences of feedback on student behavior can be assessed in different ways, the most common of which are 1) analysis of student logs, both quantitatively and qualitatively; 2) analysis of video recordings of students, which is especially common for the tasks that contain some physical aspects, e.g. public speaking or dancing.

- *Expert evaluation or comparison with expert solutions.* In this category some field experts, often the teachers of the subject, assess automated feedback in a certain way. Sometimes teacher feedback is compared to automated feedback, in which case similarities are considered to be a sign of a successful feedback approach.

- *Comparison of repeated performance measures.* Analysis of student performance measures in multiple tasks throughout the semester, which were not explicitly compared to a control group results or didn't include pre-test/post-test comparison.

- *Technical assessment of the algorithm's performance*, for example, reports on accuracy, contingency tables or other algorithms' performance metrics. This method is feasible in data-driven systems.

3.4.2. Group size

Group size represents the number of participants in the evaluation experiment. The categories of such groups are:

- *Small*: 0-30 participants.

- *Medium*: 31-100 participants.

- *Large*: 101-399 participants.

- *Massive*: more than 400 participants.

3.4.3. Statistical testing

Statistical testing is often used to validate statistical significance of the results. They are typically used in combination with other methods, e.g. with a control group experiment, to ensure statistical significance of the comparison. This dimension describes most common statistical approaches, such as t-test, ANOVA and chi-square test, which are presented in the results section.

4. Classification of the research on automated feedback technologies

This section provides a broad overview of the predominantly available systems in the field and their main properties, and thus answers the Research Question 1.

4.1. Demographic data

An overview of the selected papers are shown in Tables 2-3. The full classification based on the TAF-ClaF framework can be found in Appendix C.13-C.26. In what follows, references to certain systems are given as examples of particular classification dimensions and are not always exhaustive.

Figure 5 shows the distribution of the selected studies over time. Overall, one can see a significant increasing trend, with the number of papers devoted to an e-learning system with automated feedback rising from 7 in 2008 to a maximum of 19 in 2017. In the last two years, there is a decrease in the number of papers in comparison with 2017, but globally the number of papers published in recent years still noticeable exceeds the numbers of papers presented in the earliest years of the analyzed time period. A chi-squared test indicated statistical significance of the observed differences with a p-value below 0.01. Overall figures from the search (cfr. Section 2) indicate that the number of papers in general on the topic is not declining.

Even though a few rather popular conferences and journals can be distinguished, the range of available outlets is rather broad, as can be observed in Table 4. In particular, the selected papers are published at as many as 78 different venues, of which 44 and 34 are journals and conferences, respectively. All these outlets are dedicated to a quite diverse range of domains.

Table 2: An overview of the selected papers - Part 1

ID	Study	Year	Tool name	ID	Study	Year	Tool name
N1	Roselli et al.	2008	VaNTH ERC	N41	Akçapınar	2015	-
N2	Liu et al.	2008	-	N42	Hull and du Boulay	2015	SQL-Tutor*
N3	Lu et al.	2008	-	N43	Sullins et al.	2015	AutoTutor
N4	Kelly et al.	2008	-	N44	Barmaki and Hughes	2015	TeachLivE
N5	Parvez and Blank	2008	DesignFirst-ITS	N45	de Vries et al.	2015	GREET
N6	Stranieri and Yearwood	2008	-	N46	Fung et al.	2015	ROC Speak
N7	Di Eugenio et al.	2008	DIAG	N47	Grigoriou et al.	2015	-
N8	Gimeno Sanz and De-Sequeira	2009	InGenio	N48	Shaalan et al.	2015	-
N9	Calvo and Ellis	2010	Glosser	N49	Höhn and Ras	2016	EAGLE
N10	Dominguez et al.	2010	-	N50	Schneider et al.	2016	Presentation trainer
N11	Barker	2010	-	N51	Wilson and Czik	2016	PEG Writing
N12	Fossati et al.	2015	iList	N52	Choi	2016	iTutor
N13	Frutos-Morales et al.	2010	OeLE	N53	Frenoy et al.	2016	-
N14	Wong et al.	2010	-	N54	Belcadhi	2016	-
N15	Inventado et al.	2010	POOLE III	N55	Lukáč and Sekerák	2015	Geogebra
N16	Kazi et al.	2012	METEOR	N56	Mirzaeian et al.	2016	-
N17	Rahati and Kabanza	2010	-	N57	Sung et al.	2016	-
N18	Watson et al.	2011	-	N58	Sedrakyan et al.	2017	JMermaid
N19	Roll et al.	2011	The Help Tutor	N59	Grawemeyer et al.	2017	iTalk2Learn
N20	Gutierrez and Atkinson	2011	-	N60	Correia et al.	2017	Kora
N21	Marcus et al.	2011	-	N61	Hodgkinson et al.	2016	glGetFeedback
N22	Jylhä and Erkut	2011	iPalmas	N62	Liu et al.	2017	-
N23	Ng	2011	i-Maestro	N63	Wang et al.	2017	iGrader
N24	Nguyen et al.	2012	-	N64	Cheng	2017	-
N25	Karavirta et al.	2012	MobileParsons	N65	Dias Pereira dos Santos et al.	2017	Forró Trainer
N26	Fournier-Viger et al.	2012	CanadarmTutor	N66	Cutumisu et al.	2017	Posterlet
N27	del Mar Sánchez-Vera et al.	2012	OeLE	N67	McGuire et al.	2017	ASSISTments
N28	Johnson and Zaiane	2012	Shufti	N68	Perikos et al.	2017	NLtoFOL
N29	Rivers and Koedinger	2013	-	N69	Gerdes et al.	2017	Ask-Elle
N30	Roscoe et al.	2013	W-Pal	N70	Karkalas et al.	2017	AuthELO
N31	Chung et al.	2013	-	N71	Schaffer et al.	2017	-
N32	Jordan et al.	2013	Rimac	N72	Ai	2017	-
N33	Kennedy et al.	2013	-	N73	Zhu et al.	2017	-
N34	Mich et al.	2013	LODE	N74	Jiménez et al.	2018	TIPOO
N35	Mitrovic et al.	2013	SQL-Tutor	N75	Benotti et al.	2018	Mumuki
N36	Shekhar et al.	2014	JFLAP	N76	Chu et al.	2018	-
N37	Zinn	2014	-	N77	Demaidi et al.	2018	OntoPeFeGe
N38	Dzikovska et al.	2014	BEETLE II	N78	Fujita et al.	2018	-
N39	Narciss et al.	2014	ActiveMath	N79	Hautala et al.	2018	Mindstars Books
N40	Kão and Niitsoo	2014	MatchMySound	N80	Ho et al.	2018	Knowledge Maps

Table 3: An overview of the selected papers - Part 2

| ID	Study	Year	Tool name	ID	Study	Year	Tool name
N81	Monahan et al.	2018	CRA	N96	Price et al.	2016	Hint Factory
N82	Naismith and Lajoie	2018	BioWorld	N97	Liu et al.	2016	EQClinic
N83	Ochoa et al.	2018	RAP	N98	Baneres and Saíz	2016	VerilUOC
N84	Rajendran et al.	2018	Mindspark	N99	Basu et al.	2017	CTSiM OELE
N85	Cabestrero et al.	2018	HINTS	N100	Arevalillo-Herráez et al.	2017	-
N86	Hülsmann et al.	2018	-	N101	VanLehn et al.	2016	Dragoon
N87	Sun et al.	2018	-	N102	Tiam-Lee and Sumi	2018	-
N88	Holland et al.	2009	J-Latte	N103	Bjorn et al.	2019	-
N89	Rhienmora et al.	2011	-	N104	Lee et al.	2019	HASbot
N90	Chan et al.	2010	-	N105	Li and Hsieh	2019	CNC-Tutor
N91	Ward et al.	2013	My Science Tutor	N106	Qian and Lehman	2019	Mulberry
N92	Arevalillo-Herráez et al.	2013	-	N107	Romano et al.	2019	Dancing Coach
N93	Wong et al.	2014	COVE TS	N108	Silva et al.	2019	-
N94	Taele et al.	2015	Maestoso	N109	Wünsche et al.	2019	CodeRunnerGL
N95	Kyan et al.	2015	CAVE				
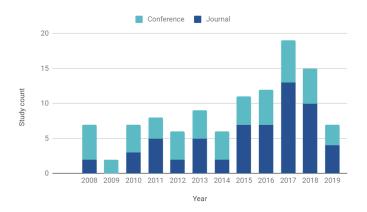
Figure 5: Number of studies published in conferences and journals each year, $\chi^2(p < 0.01)$

Table 4: The list of the most popular outlets wherein the selected automated feedback systems are reported. The outlets under the category of "other" comprise only one publication per outlet within the selected papers

Journal	N
Computers & Education	6
International Journal of Artificial Intelligence in Education	5
IEEE Transactions on Learning Technologies	5
User modeling and user-adapted interaction	2
Computer Assisted Language Learning	2
Knowledge-Based Systems	2
Other	38
Total	60
Conference	**N**
International Conference on Intelligent Tutoring Systems	7
International Conference on Artificial Intelligence in Education	4
International Conference on Computers in Education	4
International Conference on Web-Based Learning	3
ACM Conference on User Modeling, Adaptation and Personalization	2
Other	29
Total	49

Table 5: Distribution of the publication outlets across 3 dimensions: education, domain and technology, and their combinations. Most of the outlets tackle the topics on a combination of these dimensions

Outlet category	N
Education-related	8
Domain-specific	0
Technology-oriented	31
Education & Technology	22
Education & Domain	11
Domain & Technology	4
Education & Domain & Technology	2
Total	78

Based on their titles, we categorized them along 3 dimensions: education-related, domain-specific and technology-oriented (the full classification can be found in Appendix B.11-B.12). Most of the outlets are either technology-oriented or belong to more than one of the categories, e.g. both technology and education-oriented or tackle education in a specific domain. The distribution of the outlets according to this classification is shown in Table 5. Eight journals and conferences are only education or learning-oriented, 31 outlets are strictly technology-related, and no outlets are strictly domain-specific. Many outlets are interdisciplinary, i.e. include two or three fields: 22 outlets are related to both education and technology, 11 outlets are related to education in a specific domain, four outlets are describing technologies in a specific domain and two outlets are related to learning and technology for
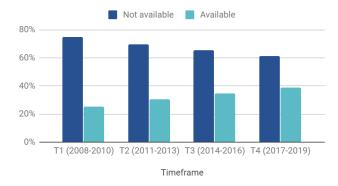
Figure 6: Comparison of the percentage of publicly available systems during different time periods

a specific domain.

Another aspect related to the publications is the presence or absence of public availability of the reported systems. Within the set of the analyzed papers, only 34% of the reported tools are publicly available, i.e. either the study includes a link to a demo of the system or provides an access to the implementation code, or a search for the name of the tool allows to access its corresponding website. In the remaining 66% of the papers, the authors do not provide any links to their systems or demos. Nevertheless, as shown in Figure 6, the percentage of publicly available systems noticeably grew over time from 2008 to 2019 from 25% to almost 40%.

4.2. Architecture

4.2.1. Domain model

Given that domain knowledge is by nature very diverse and domain-dependent, it is hard to distinguish the most common ways to structure it. In this section we give some representative examples of domain models found in the literature, which however do not cover an exhaustive list of possible options.

Liu et al. (2008), Barker (2010) and Demaidi et al. (2018) constructed a domain ontology of educational objectives using Bloom's taxonomy (Anderson et al., 2001), which is a widely used tool for classifying learning tasks in the educational domain. Parvez and Blank (2008) develop a Curriculum Information Network (CIN), in which the concepts are linked together through measures of their relationship, e.g. prerequisite and equivalence, and are also arranged according to their learning difficulty. This means that, for example,

in order to master a certain concept, the student should first understand all of its prerequisites.

It is relatively common in the literature to represent domain knowledge as a relational database, both as a means of assigning relationships between concepts and as a convenient way of implementation and extraction of such knowledge. For example, Di Eugenio et al. (2008) presented a complex system for teaching troubleshooting, where knowledge connections are stored in a database through notions of hierarchy, equivalence, reference and correlation. Furthermore, in an automated feedback system for improving reading comprehension of deaf children by Mich et al. (2013), the interactive stories are also stored in a database, allowing the system to use a web-based client-server architecture, while representing some connection between those stories might not be the main goal of such a domain model.

Other examples of structuring domain knowledge are a tree structure, adapted by Dominguez et al. (2010) and Schaffer et al. (2017), in which the topics and sections are arranged according to their relevance and associations with each other; and constraint-based modeling used by Mitrovic et al. (2013) and Hull and du Boulay (2015), where the theory is represented as a set of constraints. Many studies propose their own ways of expressing hierarchy and relations between items of domain knowledge, which are often task-specific and not easily generalizable.

4.2.2. Expert knowledge

Similarly to domain knowledge, expert knowledge can be very specific to the system it is applied to, and thus we did not aim to extract an exhaustive list of its categories. However, some general approaches of using expert knowledge can be observed, namely:

- Expert solutions to problems and libraries of correct answers.

- Sets of typical errors in the discipline.

- Feedback templates/rules.

As an example of the first approach, Correia et al. (2017) described an automated assessment and feedback generation system for teaching diagrammatic language in computer science, in which a set of correct solutions (expert knowledge) is mapped with student diagrams (student data), and the differences between those solutions trigger different feedback. Next, Ochoa

et al. (2018) developed a system for giving automated feedback on oral presentations. Categorized as having a mixed feedback generation model, i.e. both expert and data-driven, this system exploits a data-driven algorithm for classifying student videos into "good" or "bad". However, the features that are used for this classification (such as failing to maintain eye-contact, filled pauses and close posture) are developed and provided by experts, and thus are an example of a set of typical errors used as expert knowledge.

The third and the most general description of expert knowledge, feedback templates, is frequently based on some specific educational models or theories. For example, Cabestrero et al. (2018) utilized the rules proposed in the conceptual model of engagement proposed by O'Brien and Toms (2008), which aim to maintain the level of student engagement with a specifically designed order of feedback messages. Similarly, in a system by Parvez and Blank (2008), the pedagogical framework incorporates the Felder-Silverman learning style model and thus the set of feedback rules takes into account the student learning styles. In some systems it is also possible that expert knowledge is not based on an educational model, but rather is implicitly used when defining answer-feedback pairs. Additionally, although there is always some teacher reasoning involved while developing an automated feedback system, in our model we assume that purely data-driven systems don't require expert knowledge.

4.2.3. Student data

Figure 7 shows an overview of the common categories of student data used for providing feedback. It is important to note that sometimes several types of student data are used within the same system. The most common category was open answers, which were used in roughly half of the studies (50%). For example, in del Mar Sánchez-Vera et al. (2012), the system receives free text answers to open questions, which are consequently processed by natural language processing techniques. More specifically, the semantic similarity between the answer to a question and the predefined annotations for this question are calculated, generating a similarity matrix. Similarly, in a system by Wilson and Czik (2016), student essays are analyzed using semantic analysis and natural language processing. Another typical example of student data represented as open answers is a programming assignment (Watson et al., 2011; Karavirta et al., 2012), consequently analyzed by various debugging algorithms. Finally, sometimes open answer data come in a form of graphical solutions, such as stylus-written Chinese characters in an ITS for

learning Chinese language (Chu et al., 2018), tackled by sketch recognition techniques.

The next most common type was in-system behavior or log data, used by 37 systems (34%). An example of this kind of student data can be found in Rahati and Kabanza (2010), where an ITS for teaching medical diagnosis skills provides feedback in a form of dialogues between the system and a student. Consequently, not only the current message of a student is being analyzed, but also the previous steps/answers. Roselli et al. (2008) offer a system that supports teachers (who in this case play a role of learners) in creating student assignments by providing feedback on their actions within the platform. As such, input data for this system comes in a form of learner's clicks on certain buttons in the system, e.g. "create a new problem", etc.

Predefined (non-open) answers were used in 32 systems (29%). For example, Narciss et al. (2014) presented a system for learning fractions, in which student input is received as either answers to multiple choice questions or a numeric answer. The difference of a non-open answer from an open answer is a limited set of possible student responses, which makes it easier to potentially analyze such a response without a data-driven approach. In case of the system reported by Narciss et al. (2014), it is an expert-driven system that uses a set of expert rules to provide feedback.

22 studies (20%) reported the use of the real-life behavior of students, such as physical movement (Nguyen et al., 2012) or music performance (Kão and Niitsoo, 2014), and 12 studies (11%) involved built-in questionnaires or other similar tools to obtain additional student characteristics. Finally, one study (Inventado et al., 2010) applied EEG (electronic encephalogram) to obtain additional student data, which was classified as "other" type of data.

The usage of the data is to some extent determined by the educational task. For example, a system for training public speaking skills will most likely use streams of video images of students, i.e. real-life behavior, while the system for developing programming skills will be more prone to use open answers or, in some cases, predefined answers. Nevertheless, some types of data, namely log data (in-system behavior) and student characteristics, are more task-independent and can be implemented in most of the systems along with more task-specific data formats, in order to enrich the system functionality and adaptiveness.
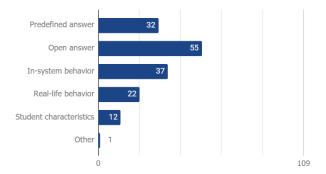
Figure 7: Classification of the selected systems across different categories of student data used
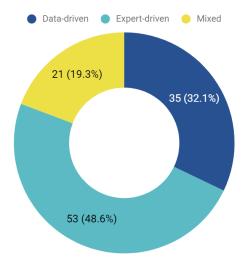

Figure 8: Distribution of the data-driven, expert-driven and mixed feedback generation models in the selected papers. The number of papers and their corresponding percentage are depicted within the diagram.

4.2.4. Feedback generation model

An overview of feedback generation models (data-driven vs. expert-driven vs. mixed systems) is provided in Figure 8. Almost half of the analyzed systems are exclusively expert-driven (Rahati and Kabanza, 2010; Jordan et al., 2013; Shekhar et al., 2014; Monahan et al., 2018), while purely data-driven systems account for around 32% of the studies (Chu et al., 2018; Hülsmann et al., 2018; Sung et al., 2016). Mixed feedback generation models, i.e. systems that utilize rules provided by experts in combination with those derived by data mining algorithms, are encountered in roughly 19% of cases (Schneider et al., 2016; Liu et al., 2017; Rajendran et al., 2018).

Expert-driven systems commonly use predefined answers as student data, such as multiple-choice questions or true-false questions (such as a system by Belcadhi (2016)), although this is not always the case. For example, a tool presented in Wünsche et al. (2019) gets a part of programming assignment, i.e. open answer, as student data. Next, the student answer in expert-driven systems is typically being evaluated by comparing it to a sample solution, a set of typical errors or a set of predefined answers. Then, a simple rule-based algorithm is used to make a decision regarding which feedback is best to offer to the student.

In case of data-driven systems, the decision for giving a certain type of feedback is to a large extent linked to a result of an algorithm trained on some historical data. For example, in a system proposed by Lee et al. (2019) for teaching scientific argumentation, a natural language processing algorithm was trained on student open answer data, combined with a support vector regression algorithm, which is capable of predicting a specific score to be given to the student answer. The results of the regression algorithm are combined with a set of expert-driven scoring rules, and the system is therefore classified by us as mixed.

As such, a presence of a machine learning or statistical algorithm *trained on student solutions* (or sometimes even on solutions provided by experts) is a sign of a system being either data-driven or mixed. It is nevertheless important to note that in purely expert-driven systems, algorithms can as well be used to implement the rules proposed by experts. Such algorithms are typically not machine learning algorithms, and they often describe procedures of comparing a student answer with a set of templates predefined by experts (sometimes such procedures are rather simple, but this is not always the case) and then making a decision based on the result of the comparison. We

did not label such algorithms, i.e. those that are used for the purpose of implementing experts rules, as the *technique* dimension. In other cases, a system may combine a machine learning model along with expert rules, e.g. a large piece of text entered by a student may be analyzed by an intelligent text mining algorithm. Such a machine learning model is however typically trained on some data: for example, a collection of student essays labeled by experts is used to train a text mining algorithm. This would be classified as a data-driven component of a mixed feedback generation model.

Table 6 summarizes the most common techniques used in data-driven and mixed systems. The percentages provided in the table is relative to the total number of data-driven and mixed feedback generation engines (encountered in 56 systems). For most common techniques, we briefly review the purpose and methodology behind them. For more detailed information regarding the techniques we refer the interested reader to Friedman et al. (2001).

Audio and video processing techniques, such as natural language processing, speech recognition, etc., are used in 30% of the data-driven systems (Taele et al., 2015; Barmaki and Hughes, 2015). Such techniques are more common in systems that give feedback on audio or video data of learners, such as systems for learning public speaking (Schneider et al., 2016), music (Ng, 2011), etc. Similarly, classification techniques are employed in 30% (Cabestrero et al., 2018) of cases. Unlike audio and video processing techniques, classification algorithms can be used in practically any system. More specifically, classification is a category of supervised machine learning algorithms that are trained on labeled input to be able to classify new data into several categories. For example, if a classification algorithm is trained on a collection of student essays labeled as "passed" as "failed", it will be able to assign these two categories to a new essay. Typical classification algorithms include random forest, logistic regression, decision trees, and support vector machines, among others.

Text mining is used in 21% of the systems. In our classification, we categorize as text mining most algorithms for working with textual data, including algorithms for detecting similarities between two pieces of text, for example, cosine similarity algorithm used in (Akçapınar, 2015). Optimization techniques, such as gradient descent and Monte Carlo methods, are utilized in 11% of the systems (Rajendran et al., 2018) and can be used for a rather large set of applications. For instance, greedy algorithm in a system for teaching Chinese characters (Chu et al., 2018) is used for detecting strokes that are close to optimal and thus to create a one-to-one correspondence between the

input and template strokes.

Other commonly used techniques include statistical modeling (Roscoe et al., 2013), clustering (Dominguez et al., 2010), regression (Inventado et al., 2010), sequence mining (Sun et al., 2018), time series analysis (Hülsmann et al., 2018) and association rule learning (Lu et al., 2008). We classified the techniques that are reported by no more than one study into the "other" category, which account to 7% of the studies (4 systems). These relatively rare techniques were, e.g., reinforcement learning (Johnson and Zaiane, 2012) and self-organizing maps (Kyan et al., 2015).

Table 6: An overview of the commonly used techniques for deriving insights from student data in data-driven and mixed systems

Technique	N	%	Examples
Audio and video processing	17	30%	Natural language processing (del Mar Sánchez-Vera et al., 2012), speech recognition (de Vries et al., 2015), gesture recognition (Barmaki and Hughes, 2015)
Classification	17	30%	Naive Bayes (Fung et al., 2015), SVM (Cabestrero et al., 2018), random forest (Ma et al., 2017), nearest-neighbor algorithm (Nguyen et al., 2012), logistic regression (Frenoy et al., 2016)
Text mining	13	23%	Latent semantic analysis (Fung et al., 2015), weighted keywords matching (Sullins et al., 2015)
Optimization	6	11%	Gradient descent (Rajendran et al., 2018), Monte Carlo methods (Johnson and Zaiane, 2012)
Statistical modeling	5	9%	Bayesian network (Inventado et al., 2010), CRF models (Gutierrez and Atkinson, 2011)
Regression	4	7%	Linear regression (Roscoe et al., 2013), support vector regression (Zhu et al., 2017)
Clustering	3	5%	k-means clustering (Ng, 2011)
Hidden Markov models	3	5%	Tiam-Lee and Sumi (2018); Rhienmora et al. (2011); Gutierrez and Atkinson (2011)
Sequence mining	2	4%	Apriori (Kennedy et al., 2013)
Time series analysis	2	4%	Dynamic time warping (Kão and Niitsoo, 2014)
Association rule learning	2	4%	Lu et al. (2008)
Other	4	7%	

4.2.5. Implementation

The programming languages most commonly used for implementation of automated feedback are shown in Table 7. For slightly more than a half (56%) of the studies, no implementation information is specified by the authors, including languages, platform or software. The most commonly used languages are JavaScript (Karavirta et al., 2012), Java (Mirzaeian et al., 2016; Di Eugenio et al., 2008), C++ or C# languages (Kelly et al., 2008; Akçapınar, 2015), and Python (Roselli et al., 2008).
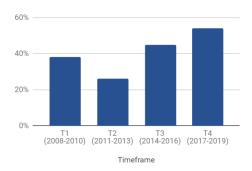
Figure 9: Percentage of the web-based systems during different time periods. The percentage of the web-based system increases over time.

The systems that work with gesture or movements data tend to use Microsoft Kinect (Nguyen et al., 2012; Barmaki and Hughes, 2015). Other common tools that were used in more than one system are Web Ontology Language (Frutos-Morales et al., 2010), SQL (Ai, 2017) and Rapid Miner (Inventado et al., 2010). A relatively large amount of systems (12 systems or 11%) use unique tools that are encountered only once in the selected papers.

Table 7: An overview of the programming languages used for implementation

Language/tool	N	%
Not specified	61	56%
JavaScript	10	9%
Java	9	8%
C++/C#	7	6%
Python	5	5%

Another implementation dimension is whether the systems are web-based or not. According to our analysis, 42% of the reported systems were web-based, while 58% of the systems were offline/non-web-based. The amount of web-based systems grew over time from less than 40% to around 55%, as depicted in Figure 9.

Finally, in the third implementation dimension, the systems' connection to the learning platform was analyzed. Most of the systems (73 studies or 67 %) are classified as standalone, e.g. Cutumisu et al. (2017); Sullins et al. (2015); Belcadhi (2016). Plug-in systems providing automated feedback to the users of previously existing platforms account for 28 % (Roll et al., 2011; Correia et al., 2017). For 5% of studies (6 studies) it is not clear whether they can be classified as standalone or plug-ins.
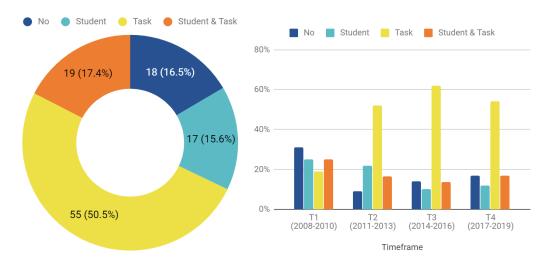
38

4.3. Feedback

4.3.1. Adaptiveness

Figure 10: Distribution of adaptiveness categories in the papers (left) and its changes over time (right)

The distribution of feedback adaptiveness levels is shown in Figure 10 (left). 50.5% of the studies employed task adaptiveness (Kazi et al., 2010; Nguyen et al., 2012; Mirzaeian et al., 2016). In comparison, only 15.6% of the systems reported to have student adaptiveness (Schaffer et al., 2017; Roselli et al., 2008; Chung et al., 2013). A comparable number of studies (17.4%) utilized both student and task adaptiveness (Gutierrez and Atkinson, 2011; Grawemeyer et al., 2017). Finally, another 18 (16.5%) of the reported systems showed no adaptiveness (McGuire et al., 2017; Chu et al., 2018).

Analyzing the feedback adaptiveness reported in the different time periods (see Figure 10, right), we can see that the percentage of systems with student adaptiveness and with combined student and task adaptiveness notably decreases over time from 2008 to 2016, with a slight increase in 2017-2019. In contrast, the percentage of task-adaptive system grows drastically from 19% in 2008-2010 to over 60% in 2014-2016. The proportion of systems with no adaptiveness ranges from 9% in 2011-2013 to over 30% in 2008-2010; overall, there is a slight counterintuitive tendency for the proportion of such systems to be growing, but the fluctuations are not significant enough to make such conclusions.
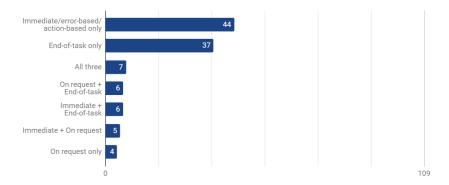
Figure 11: Distribution of the selected papers according to feedback timing they utilize

4.3.2. Timing

Figure 11 shows the distribution of feedback timing for the reported systems. Since some systems used several types of feedback in terms of timing, the combinations of different types within one system are shown as well. Given the distribution, two prominent categories of systems can be distinguished: the ones that provide immediate feedback based on student actions or errors (40%), and systems that provide feedback after the task is completed (end-of-task, 34%). On request feedback was less common, used in 20% systems and, most frequently, used in combinations with the other types of feedback timing. Just a few systems (12% in total) combined end-of-task and immediate feedback, from which more than half employed all three types of feedback simultaneously (6%).

4.3.3. Learner control over feedback

The distribution of levels of learner control over feedback can be seen in Figure 12 (left). The majority of the reported systems (71.6%) did not involve any learner control over feedback (Karavirta et al., 2012; Akçapınar, 2015). 23.9% of the systems provided students with mild control (Perikos et al., 2017), while only a few (4.6%) of the studies described systems where students had extensive control over feedback type, timing and/or appearance (Shekhar et al., 2014).

As illustrated by Figure 12 (right), the distribution of the learner control stays quite stable over time. Despite potential expectations of an increase in user-friendliness of automated feedback technologies, it is consistent that the majority of them do not offer a possibility to influence the provided feedback in any way. The number of systems with mild degree of user control changes

Figure 12: Distribution of learner control categories in the selected papers (left) and its changes over time (right)

only slightly throughout years, and the number of systems with high degree of user control do not show a tendency to increase in the last years and can even be observed to be decreasing.

4.3.4. Purpose

The number of papers for each combination of feedback types based on their purpose is illustrated in Figure 13. The majority of the reported systems use feedback of two or more types. Only a few systems provide just one type of feedback in terms of purpose. Overall, the most common feedback type is corrective, with 98 out of 109 (90%) studies utilizing it, and typically used in combinations with informative and suggestive feedback. For example, an intelligent tool for learning music theory developed by Taele et al. (2015) provides students with corrective feedback on the elements of their music theory assignments indicating correct and incorrect parts. In addition to that, the tool gives suggestive feedback, which does not indicate the exact error but rather gives a suggestion for improvement, and also informative feedback as part of individual quiz questions, in which the students are provided with additional learning material in a form of small text messages.

The combinations of other types of feedback are slightly less frequent, such as informative with suggestive, e.g. in Roscoe et al. (2013)), where students receive targeted suggestive feedback for improving their essays (suggestive feedback), along with general tips on essay writing (informative feed-
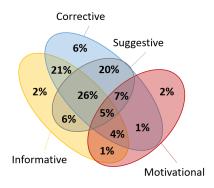
Figure 13: Distribution of papers for each category of feedback purpose

back). Besides, in some systems motivational is combined with one of the other types of feedback. Generally, motivational feedback is the least represented type, with only 19% studies mentioning it. One of the interesting uses of such feedback is demonstrated by Rajendran et al. (2018): in their system, supportive motivational messages are used to tackle student frustration, when it is detected based on the log data.

Some systems (5%) employ all types of feedback at the same time. For example, the system for teaching basic electricity and electronics reported by Dzikovska et al. (2014) uses a rich combination of various types of feedback: simultaneously with performing a task in a circuit simulator, a student is engaged in a dialog with a virtual tutor, which asks suggestive questions, provides corrective and informative feedback textual messages, as well as supports the student with motivational messages such as "Keep trying!".

4.4. Educational context

4.4.1. Domain

The summary of the addressed domains is shown in Figure 14. As the figure shows, the STEM (Science, Technology, Engineering and Mathematics) domain is the most represented one, with 52 reported technologies (48%) used for teaching STEM subjects. The distribution of domains within STEM is shown in Figure 15. It is important to note that programming was included in the "Computer science" category, and accounts for one third of all STEM subjects addressed by the studies (in total the "Computer science" category accounted for about a half of all STEM subjects). "Science" category included physics (Jordan et al., 2013; Dzikovska et al., 2014), biology (Hautala et al., 2018) and general "science" subject (Basu et al., 2017). The
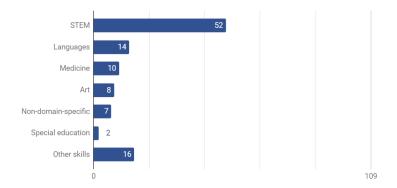
Figure 14: Distribution of systems based on the educational domain they address
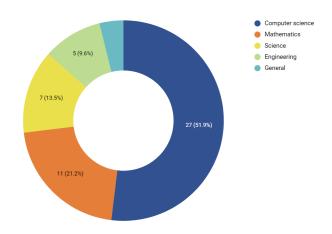
Figure 15: Subdivision of domains within STEM

next most common domain was language learning, with 14 papers (13%). Medicine-related systems were presented in 10 papers (9%), while 7 studies (6%) were related to non-domain-specific systems. The least represented distinct domains were art (dance and music - e.g. Kão and Niitsoo (2014); Dias Pereira dos Santos et al. (2017)) and special education (Kelly et al., 2008; Mich et al., 2013). The other 16 studies (15%) addressed other specific skills, such as public speaking (Schneider et al., 2016), teaching (Barmaki and Hughes, 2015) and negotiation skills (Monahan et al., 2018).

4.4.2. Level

Regarding the levels of education at which the described systems or automated feedback technologies were aimed (Figure 16), tertiary education
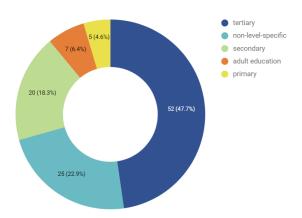
Figure 16: The proportion of studies aimed at different educational levels

(higher education) was the most represented with 47.7%. Non-level-specific systems accounted for 22.9% of the studies. School education, comprising primary and secondary educational levels, was addressed in 18.3% of the selected studies, with primary education being the least represented level (about 4.6% of all studies). The adult education category included professional education, vocational training and other types of training provided to adults outside the context of tertiary education and accounted for 6.4% of all studies.

4.4.3. Setting

The proportion of papers that used different categories of educational settings are shown in Figure 17 (left). The larger share of the analyzed systems aimed at an in-class educational setting rather than online or blended learning: in-class implementations of the systems with automated feedback outweigh all the other types, accounting for more than half of the studies (54.1%). A blended learning setting was presented in 11.9% of the papers, while a purely online setting for automated feedback systems was described in 13.8% of the studies. Finally, many papers (about 20%) did not specify the educational setting or declared that the described system was non-setting-specific.

Despite the fact that the systems used purely online are outnumbered by those used in-class, a certain growth trend in the number of studies addressing purely online settings as well as blended learning settings can be observed from 2008 to 2016 (Figure 17, right). The percentage of the studies for which

the educational setting was not specified (or the described system could be applicable in any settings) has been growing steadily from 2008, while the percentage of the systems employed in in-class settings dropped dramatically from 2008 to 2016, and then showed a slight increase in 2017-2019.
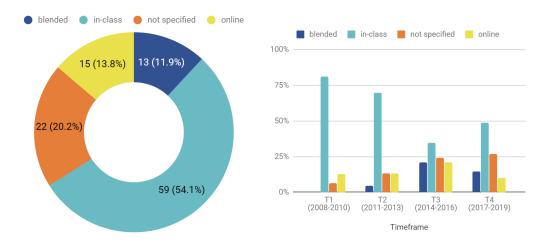
Figure 17: Distribution of the categories of educational setting in the selected papers (left) and its changes over time (right)

4.5. Evaluation

It can be considered a positive trend that the majority of the studies (86%) evaluate their technology in some way. Furthermore, the vast majority of these evaluations (91% of the studies that provide some evaluation) reported some evidence of positive influence of automated feedback. The particular evaluation approaches, the sizes of the groups of student participated in the evaluation experiments, and statistical testing methods used in the evaluations are discussed below.

4.5.1. Evaluation method

Frequently used evaluation approaches are summarized in Figure 18. It is important to note that sometimes several evaluation methods are combined within one study; however, those combinations do not imply that each of the evaluation methods cannot be used independently without being combined with other ones. Moreover, given the study eligibility criteria used for selecting papers in this literature review, it could be that for some systems, only one publication out of several reporting the same system was selected for the

review, and therefore, only one of the performed evaluation approaches is reported in our study. Nevertheless, in what follows we aim to give a numerical representation of evaluation methods used in the selected sample of papers, and thus give an idea to the reader regarding relative popularity of each of the methods in the domain.

The most common way to evaluate automated feedback systems within the selected papers is an analysis of student perceptions of the system (or other aspects, such as their perception of achieved learning goal, decreased frustration level, etc.) through students surveys or oral interviews, which is encountered in 64% of evaluations, e.g. in Roselli et al. (2008); Rahati and Kabanza (2010); Zinn (2014); Sedrakyan et al. (2017). The second popular approach is a control group experiment employed in 56% reported evaluations, e.g. in de Vries et al. (2015); Cheng (2017); Karkalas et al. (2016). Moreover, the results of pre- and post-tests are compared in 36% cases. These frequent evaluation methods were followed by assessment of the algorithm performance (25%), expert evaluation (22%), student behavior analysis (20%), and repeated performance measures (19%). Only two (2%) of the evaluations are not listed above, because they were specifically tailored to the study's objectives, and were therefore classified as "other". More specifically, these evaluations are a simulation of student performance based on a probabilistic model performed by Frenoy et al. (2016) and looking at the significant number of students who chose to use the tool in Baneres and Saíz (2016).

Overall, 60% of the reported evaluations include statistical testing for validating the statistical significance of other methods. The combinations of statistical testing with other methods is depicted in Figure 18. Not surprisingly, it can be observed that statistical testing was frequently used with pre-test/post-test analyses, repeated performance measures comparisons and control group experiments, while it was less common to use for assessing the performance of the algorithms and student surveys.

4.5.2. Group size

The distribution of the student group sizes used for evaluation is discussed below. Out of the papers that reported evaluation, only 3 of them (3%) did not use any student group for evaluation (Mirzaeian et al., 2016). Similarly, only a small proportion of studies (5%) applied their tool to a massive (over 400) group of students. For example, Wang et al. (2017) evaluated algorithm performance on a total of 3361 students, and Barker (2010) conducted
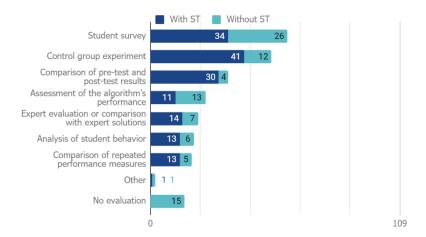
Figure 18: The number of the papers that performed each of the common evaluation approaches. Different colors indicate whether the evaluation included statistical testing (ST)

student surveys with 500 participants. The majority of the studies evaluate their systems using small (34%), medium (35%) and large-sized (22%) student groups. Among these most common group sizes, ranging from small to large or, in other words, from 0 to 400 students, a few tendencies can be observed with regards to the evaluation method used, as shown in Table 8. First, more mathematically demanding and rigorous methods, namely control group experiments, pre-test & post-test approach, and statistical testing, are significantly more common when a sufficient number of students (medium or large group size) is involved. Second, when a smaller number of students is available, it is more common to exploit student questionnaire as an evaluation method in comparison with larger groups. Other evaluation methods are used equally likely with different group sizes.

4.5.3. Statistical testing

The typical statistical tests used in automated feedback research for validating the statistical significance of the evaluation methods are summarized below. The majority of the statistical tests employed some variant of ANOVA (48%), used paired t-test (43%) or calculated correlation (36%).

4.6. Terminology

We analyzed how the selected systems are named when they are being introduced by the authors in abstracts or introduction sections. The purpose

Table 8: The distribution of the group sizes depending on the evaluation method used

Group size	Control group	Student survey	Pre-test & post-test	Statistical testing	Repeated measures	Algorithm performance	Expert evaluation	Behavioral analysis	Other
No student group	0	1	0	0	0	1	0	1	0
Small	14	22	7	11	6	11	7	5	0
Medium	23	17	16	26	6	6	6	6	0
Large	14	16	10	19	6	4	8	7	0
Massive	2	4	1	1	0	2	0	0	1

of this analysis is to give an impression of a wide variety of different terms that are used to describe systems with the same or very similar functionality. In our opinion, such a large set of names that are used within the domain of automated feedback system at times makes it difficult for a reader to grasp the type of system that is actually being presented.

Confirming our expectation, the analysis revealed that the authors of the selected papers use a large variety of different names to describe their systems: more specifically, we found 49 different terms for the 109 selected papers, e.g. Intelligent Tutoring System, teaching platform, online feedback tool, etc. While at times the variety of the systems' names can be justified and clearly represents synonyms and equivalent terms, such as digital learning environment vs. virtual learning environment, sometimes some used terminology might be confusing for the reader, as it is not always easy to grasp what the authors mean by a certain term (e.g. for the terms like virtual agent, interactive multimedia, etc.). Furthermore, our analysis has not revealed any significant correlation between the way the systems are named and their functionality. The variety of the names used is summarized by a word cloud in Figure 19.

The most frequent keywords derived from the systems' names are indicated in Figure 20. The most common full name is by far Intelligent Tutoring System, used by 35% of the authors. Next frequent keyword combination was "learning environment" (15%), which was used together with some adjectives and prefixes such as: virtual (Kennedy et al., 2013), interactive (Karkalas et al., 2016), digital (Lukáč and Sekerák, 2015), intelligent (Sun et al., 2018), training (Hülsmann et al., 2018), web (Correia et al., 2017), e- (Marcus et al.,

Figure 19: A word cloud built from the names of the analyzed systems
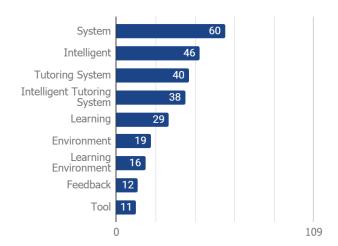
Figure 20: The list of the most frequent keywords and the number of papers that used these keywords to name their systems

2011), online (McGuire et al., 2017), computer (Naismith and Lajoie, 2017), open-ended (Basu et al., 2017) and example-based (Watson et al., 2011). The most frequent single keywords are "system" (55%), "intelligent" (42%), "learning" (27%), "feedback" (11%) and "tool" (10%).

5. Discussion and recommendations

In this paper, we thoroughly explored the field of automated feedback technologies for learners. With regard to Research Question 1, Section 4 provides a broad overview of the predominantly available systems in the field. Moreover, the second and third research questions are addressed in Section 3, in which a classification framework was developed that allows to categorize these systems according to key dimensions. The overall framework is illustrated in Figure 3. The remainder of this section describes a reflection on our study by pinpointing the main observations, recommendations and limitations.

5.1. Main observations

The main findings and trends in the field of educational technologies in the past 10 years can be summarized as follows:

A rather scattered reporting in this field of research. First of all, we observe a clear increase in the number of publications related to automated feedback systems since 2008. Notwithstanding the growing scientific community addressing the development of such systems, our analysis of publication outlets shows that publications are scattered across different fields, with no less than 78 different outlets for 109 papers. This diversity of venues and outlets can be explained by the inherently multi-disciplinary nature of this type of research, spanning educational sciences, computer sciences and disciplinary learning domains. Different parts of the system design, validation, and evaluation can be of interest of different journals. Such a diversity however makes it difficult to obtain an adequate overview of the field and might cause difficulties for researchers to position their work with respect to the state of the art. An entirely uniform structure for describing research results would be obstructive for the targeted reporting of specific points of interest for a given community. However, a common basic set of information would facilitate communication across journals and fields.

The framework laid out in this paper could play an important role in what those minimal reporting requirements might be.

Implementation accessibility. A second rather worrisome trend in the field of automated feedback technologies is that authors do not tend to share the technical details on how their system was created, for instance by detailing tools or languages used for implementation. This trend finds additional confirmation in the fact that a majority of the authors does not make their system publicly available.

Learning theories. Another important takeaway from this analysis of literature is that in most cases the applied learning theories or educational frameworks for automated feedback are not reported, even though this part of the methodology may play an important role in understanding the context in which a system is implemented. Moreover, an interesting point for future research on automated feedback could be the evaluation of applicability of the existing learning theories to contemporary e-learning systems employing automated feedback, as well as making an overview of newly emerged theories, which are applicable exclusively in the context of intelligent tutoring systems.

Data vs. expert-driven. The next important observation relates to the construction of the feedback generation model. Only half of the systems reported in the selected studies can be described as data-driven. The other half of the system still relies on substantial expert knowledge to configure rule-based engines providing students with feedback. While relying on expert knowledge in itself is not necessarily a drawback, this means that the full potential of recent technological advancements is not being fully realized. Moreover, a large number of the data-driven systems are mostly used to process real-life behavior of students such as speech, video recordings or movements, and less often consider personalized student models based on e.g. emotional states, skill levels, etc. This makes the level of personalization of these systems questionable, which is more clearly evidenced from our findings regarding the systems' adaptiveness. This is a cause for concern, given that it is largely discussed in the educational community that personalized feedback is more effective than a one-fit-all approach. For example, Gallien and Oomen-Early (2008) report that students who received online personalized feedback consistently show higher levels of performance and

satisfaction compared to those who received general feedback.

Adaptiveness. In general, our analysis reveals that, while a lot of literature discusses personalized feedback as being an important factor for students' knowledge acquisition (Economides, 2005; Gallien and Oomen-Early, 2008; Belcadhi, 2016), only around one third of the systems (34.3%) adapt the provided feedback to student characteristics and personalities. Almost half of the studies (48%) present the systems that are adaptive only to the task. We argue that this observation might stem from the growing numbers of students and limited teaching resources, which make that an increasing number of automated feedback systems are being developed with the main purpose to ease the burden of correcting multiple student assignments rather than ensuring adaptive learning. This trend is further supported by the naming of some systems (e.g. assessment tool, automatic grader, etc.).

It is important to clarify that customized/personalized/adaptive feedback is not synonymous to elaborative or suggestive feedback in terms of feedback complexity and the level of granularity in which it guides a student. For example, excessive specificity of feedback or "micro-management" is shown to have poor performance for high achievers, especially when developing skills associated with a high level of creativity (Seechaliao, 2017). Personalized feedback is not necessarily more detailed or suggestive than generic/collective feedback; on the contrary, it can actually help to avoid the issues of feedback being overly specific for certain learners. As an example, generic "default" feedback messages with explanations or suggestions can be programmed in an e-learning environment. Without personalization, this feedback will appear no matter what the student' achievements in a previous tasks is. In order to avoid a higher achiever to get bored or irritated with additional support, one can make sure that for those students that performed very well on certain tasks this feedback will not be shown, or another type of feedback, giving additional thought challenge, can be introduced. Thus, adaptability does not coincide with the level of specificity of feedback.

Personalization. A further important observation relates to the timing of feedback and learner control. Overall, we find that automated feedback systems still remain rather rigid in terms of when feedback is provided,

e.g. very few systems offer feedback on request, and whether the learner can self-configure the feedback system with 72.5% of systems offering no learner control. This trend somehow contradicts the core concept of personalized learning systems, being tailored to the characteristics and desires of individual learners.

It could be argued that in certain situations a complete control over feedback delivery from student's side could interfere with providing the feedback in a way already known to be most effective. For instance, in a paper describing guidelines for providing formative feedback, Shute (2008) discusses the literature on effects of different feedback timing on student learning, and, although there appears to be no consistent differences in the efficiency of types of feedback timing overall, certain interactions between learning tasks/learner level and preferable timing could be observed. Undoubtedly, in situations when it is known by the teacher that a certain feedback parameter (e.g. timing) is the most effective for a specific task and/or learner, not allowing the learner to freely change such an effective 'feedback setting' is a rational choice. In this sense, some cautiousness and consideration could be understandable before providing a full control over all possible feedback properties. Nevertheless, in situations when it is not evident that a certain feedback type is superior, allowing the students to adapt the feedback delivery to their own needs and preferences could help them to discover most effective feedback by themselves.

A big potential drawback of computer tutoring when compared to human tutoring is that feedback, in theory, becomes massive and impersonal. This, however, does not have to be the case, given the kind of capabilities that the contemporary technologies possess. Using these capabilities, it is possible to develop more personalized, student-oriented and student-adaptive solutions. A higher degree of learner control allows for additional personalization from the student side, while no control situation could sometimes be a one size fits all approach, which is not always working for all learners. As Perlmuter and Monty (1977) claim, the mere illusion of control significantly improves performance in the variety of situations.

5.2. Recommendations

As a result of our findings, certain recommendations can be given to the future developers and researchers in the field of intelligent tutoring systems

or other systems employing automated feedback.

Common reference framework. It would be helpful for the research community if studies would use a common reference framework when presenting research results. This would facilitate the understanding of the reports for the interdisciplinary reader or educator, and would avoid gaps in reporting information about the developed system. The classification framework developed in this paper might serve as or offer a starting point for developing such a reference framework.

Public availability. For the sake of scientific integrity, it is important that the developed systems (or at least the demo-versions of such systems) would be accessible for other specialists in the field. It is understandable that, in certain cases, making an automated feedback system publicly accessible years after it has been presented to the scientific community might pose some difficulties in practice. Moreover, continuous research and improvement of an e-learning system might mean that the system is constantly updated and an older version of the system is no longer available. Therefore, in situations when it is not possible to provide public access to the system, it could be a useful practice make available videos of its core functionality and/or its specifications in digital documents, containing most relevant information regarding system implementation. Such practice will make it more probable that the system information is still retrievable far into the future. This could improve the quality of research in the field of intelligent tutoring systems drastically and facilitate knowledge exchange.

Integrating evidence-based learning practices. As discussed in our observation regarding learning theories, the field of automated feedback development could benefit from a close collaboration with educational research in order not only to facilitate better learning processes supported by evidence-based practices, but also to allow for novel learning theories to emerge, both in online and offline contexts.

More focus on personalization. Currently, available automated feedback technologies can be considered teacher-oriented rather than student oriented, in the sense that the first focus lies on feedback automation, thus firstly addressing the teacher's workload, rather than primarily

addressing the quality of the feedback given to the student. This conclusion is supported by consistently low levels of learner control, less systems with adaptiveness to student characteristics and more systems with a major goal of automatic assessment grading. While freeing up a teacher's time by automating parts of the feedback process may indirectly benefit to students too, students can nevertheless potentially benefit directly from a shift towards a more student-oriented approach, in which they could influence a system more and in which a system would adapt to their needs, goals, personal characteristics and learning preferences in a better way. Data-driven personalization, i.e. where no teacher intervention is required, has the additional advantage that it can be generated immediately, for not depending on the teacher's availability.

Age of data. Currently, we are experiencing an unprecedented growth of opportunities and possibilities for collecting and analyzing data. This growth continually leads to the emergence of more and better learning analytics techniques that are capable of discovering valuable insights from student data. This possibility to use more data-driven techniques or fine-granular input data formats could potentially enhance many automated feedback technologies. Firstly, logging and analyzing in-system behavior of the students might offer more means to enhance system adaptiveness, understand students' problems and needs, and support them in a better way. Secondly, even if purely data-driven systems might not always be desirable to experts in the field, mixed systems that use both an expert and data-driven approach could be superior compared to purely expert-driven ones. Thus, our recommendation for the future system developers is to exploit a larger set of opportunities offered in this *age of data* and, in particular, through learning analytics.

Supporting complex knowledge acquisition. As can be observed in the overview of the currently available technologies, the contemporary automated feedback technologies can offer a wide spectrum of functionality, such as analyzing and processing scientific essays, programming assignments, psychical movements and musical performances in real time, or using sophisticated algorithms for providing a large variety of feedback types. We hope that the provided range of systems inspires

new system developers to use the opportunities of the age of data and the capabilities of contemporary data processing techniques to provide feedback for more complex tasks, and thus facilitate the development of more advanced automated feedback systems aimed at complex knowledge and practices.

5.3. Limitations

A number of limitations should be taken into account regarding this study. First, the very process of search and the search query, though effective, might not give hundred percent certainty in terms of completeness and absence of bias. However, the large quantity of papers assessed and reviewed, enforces our conviction that our review is revealing an adequate understanding of the state of the art in the field. Nevertheless, it should be noted that the findings of this study, and especially numerical trends, are confined to the selected systems and not on all available systems, which puts certain limits to the generalizability.

Second, only a fraction of the systems were publicly available, so some properties of automated systems were retrieved based on the descriptions in the corresponding paper and might be subject to interpretation. The main cause of this issue relates to term ambiguity in the field, but this is exactly what we try to address with our framework.

Finally, given the decision of only including one most full study per system in rare cases when several studies that represent the same system were discovered, some of the statistics depicting evaluation approaches might not grasp the systems that were evaluated in different ways in different studies. Moreover, the evaluation approach is not a static characteristic of the system, and can be expected to change over time if the system effectiveness is evaluated again. Therefore, we invite our readers to consider the section with evaluation approaches as an enumerated set of evaluation possibilities and numerical trends rather that the exact percentages of evaluations of all possible systems.

6. Conclusion

This study reports on a structured literature review of automated feedback technologies in education. After a four-phase search process, we selected and analyzed 109 papers. Based on this analysis, we identified the most relevant dimensions for classification of automated feedback technologies and

combined them into a general classification framework. Additionally, we classified the selected papers according to the framework and presented an overview of recent trends in the field of educational systems. A complete list of the selected papers and the full classification are available in Appendix. It is argued that our classification framework is a proper instrument for researchers in the field to set out and align their research agendas, for instance as evidenced by the recommendations put forward in this study.

AppendixA. Search query

Table A.9: Keywords and words collocations that were selected to construct a search query

Keywords denoting educational technologies	Keywords denoting education
Tutoring system	Education
Tutorial dialogue system	Teaching
Intelligent computer tutoring	Learning
Educational system	Instruction
Educational software system	Tutoring
Learning environment	
Learning management system	
Technology enhanced learning	
Technology mediated learning	
Automated feedback	
Automatic feedback	

Table A.10: Search query

Search query
TITLE(feedback) AND TITLE-ABS-KEY("tutoring system" OR "tutorial dialogue system" OR "intelligent computer tutoring" OR "educational system" OR "educational software system" OR "learning environment" OR "learning management system" OR "technology enhanced learning" OR "technology mediated learning" OR "automated feedback" OR "automatic feedback") AND ALL("learning" OR "teaching" OR "education" OR "instruction" OR "tutoring")

AppendixB. Classification of the publication outlets

Table B.11: Classification of the outlets in which the selected studies were published, along the 3 dimensions: education-related (E), domain-specific (D) and technology-oriented (T). Most of the outlets belong to several categories. Part 1. Conferences

Publication outlet	Education	Domain	Technology
A Conference on Interaction with Sound			T
ACM Conference on Learning@ Scale	E		
ACM Conference on User Modeling, Adaptation and Personal-ization			T
ACM IC on Multimodal Interaction			T
ACM IC on Pervasive and Ubiquitous Computing			T
ACM Technical Symposium on Computer Science Education	E	D	
American Society for Engineering Education Conference	E	D	
Graphics Interface Conference			T
Conference on Innovative Applications of Artificial Intelligence			T
IEEE IC on Automatic Face & Gesture Recognition			T
IEEE IC on Contemporary Computing			T
IEEE IC on Cognitive Infocommunications			T
IEEE IC on Electronics, Information, and Communication			T
IEEE IC on Emerging eLearning Technologies and Applications	E		T
IEEE IC on Image and Vision Computing New Zealand			T
IEEE IC on Systems, Man, and Cybernetics			T
IEEE Symposium on E-Learning, E-Management and E-Services	E		T
IC on Artificial Intelligence in Education	E		T
IC on Complex, Intelligent, and Software Intensive Systems			T
IC on Computer Supported Education	E		T
IC on Computers in Education	E		T
IC on Educational Data Mining	E		T
IC on Information Technology: New Generations			T
IC on Intelligent Tutoring Systems	E		T
IC on Learning Analytics and Knowledge	E		T
IC on Web-Based Learning	E		T
Interservice/Industry Training, Simulation, and Education Conference	E		T
Joint German/Austrian Conference on Artificial Intelligence			T
Natural Language Generation Conference			T
Pasific Asia Conference on Information Systems			T
Symposium on Languages, Applications and Technologies			T
The First Workshop on AI-supported Education for Computer Science	E		T
World Conference on Educational Sciences	E		
Young computer scientists conference			T

Table B.12: Classification of the outlets in which the selected studies were published. Part 2. Journals

Publication outlet	Education	Domain	Technology
ACM Transactions on Intelligent Systems and Technology			T
Advances in Health Sciences Education	E	D	
Artificial Intelligence in Medicine		D	T
Australasian Journal of Educational Technology	E		T
Behaviour & Information Technology			T
Computer Applications in Engineering Education	E	D	T
Computer Assisted Language Learning	E	D	T
Computer Science Education	E	D	
Computers & Education	E		T
Computers & Graphics			T
Computers in Human Behavior			T
Educational Media International journal	E		T
Educational Psychology	E		
Educational technology research and development	E		T
Engineering Applications of Artificial Intelligence			T
European Association for Computer Assisted Language Learning	E	D	
Expert Systems with Applications			T
Frontiers in Psychology	E		
Frontiers in ICT			T
IEEE Access			T
IEEE Transactions on Learning Technologies	E		T
Information and Software Technology		D	T
Interactive Learning Environments	E		T
International Journal of Artificial Intelligence in Education	E		T
International Journal of Emerging Technologies in Learning	E		T
International Journal of Engineering Education	E	D	
International Journal of Learning Technology	E		T
International Journal of Science Education	E	D	
Journal of Biomedical Informatics		D	T
Journal of Educational Psychology	E		
Journal of Engineering Education	E	D	
Knowledge-Based Systems			T
Learning and Instruction	E		
Mathematics Education (ZDM)	E	D	
Medical Education Online	E	D	
Musical Robots and Interactive Multimodal Systems		D	T
Natural Language Engineering			T
Procedia-Social and behavioral sciences	E		
SAGE Open	E		
Sensors			T
Science Education	E	D	
Technology, Instruction, Cognition and Learning	E		T
The Internet and Higher Education	E		T
User modeling and user-adapted interaction			T

60

AppendixC. Classification of the selected papers

Table C.13: Classification of the selected papers based on TAF-ClaF - Part1-1

ID	Study	Public availability	Terminology	Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other	Feedback generation model	Technique
				Student data							
N1	Roselli et al. (2008)	x	ITS			x				Expert	
N2	Liu et al. (2008)		ITS		x					Expert	
N3	Lu et al. (2008)		ITS		x	x				Data	Classification, association rule learning
N4	Kelly et al. (2008)		VLE				x			Data	Classification, statistical modeling
N5	Parvez and Blank (2008)	x	ITS	x		x		x		Expert	
N6	Stranieri and Yearwood (2008)		Interactive learning environment	x						Expert	
N7	Di Eugenio et al. (2008)		ITS		x					Data	Natural Language Generation
N8	Gimeno Sanz and De-Sequeira (2009)		LMS	x		x				Expert	
N9	Calvo and Ellis (2010)	x	Automated feedback system		x					Data	Text mining
N10	Dominguez et al. (2010)		Online learning system		x	x				Data	Clustering, association rule mining
N11	Barker (2010)		Automated feedback system	x		x				Expert	
N12	Fossati et al. (2015)	x	ITS		x	x				Mixed	Statistical modeling, optimization
N13	Frutos-Morales et al. (2010)	x	Learning platform		x					Data	Text mining
N14	Wong et al. (2010)		Computer-assisted learning system		x					Expert	
N15	Inventado et al. (2010)		ITS		x			x	x	Data	Statistical modeling, regression, classification
N16	Kazi et al. (2012)		ITS		x					Mixed	Text mining

Table C.14: Classification of the selected papers based on TAF-ClaF - Part1-2

ID	Study	Public availability	Terminology	Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other	Feedback generation model	Technique
				Student data							
N17	Rahati and Kabanza (2010)		ITS		x	x				Expert	
N18	Watson et al. (2011)		Example-based learning environment		x					Data	Text mining
N19	Roll et al. (2011)		ITS	x		x		x		Expert	
N20	Gutierrez and Atkinson (2011)		ITS		x					Data	Hidden Markov models, classification
N21	Marcus et al. (2011)		E-learning environment	x		x				Expert	
N22	Jylhä and Erkut (2011)	x	Interactive Tutoring System				x			Expert	
N23	Ng (2011)	x	Interactive Multimedia				x			Data	Audio and video-processing, clustering
N24	Nguyen et al. (2012)		Online feedback system				x			Data	Classification
N25	Karavirta et al. (2012)		Mobile Learning Application		x	x				Expert	
N26	Fournier-Viger et al. (2012)		ITS	x		x	x			Mixed	Sequence mining
N27	del Mar Sánchez-Vera et al. (2012)	x	E-learning platform		x					Data	Natural language processing, text mining
N28	Johnson and Zaiane (2012)		ITS	x		x				Data	Clustering, reinforcement learning, optimization
N29	Rivers and Koedinger (2013)		Approach		x	x				Data	Optimization
N30	Roscoe et al. (2013)	x	ITS		x					Data	Natural language processing, regression, statistical modeling
N31	Chung et al. (2013)		ITS	x				x		Expert	
N32	Jordan et al. (2013)	x	ITS		x					Expert	

Table C.15: Classification of the selected papers based on TAF-ClaF - Part1-3

				Student data						Feedback generation	
ID	Study	Public availability	Terminology	Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other	model	Technique
N33	Kennedy et al. (2013)		VLE				x			Data	Sequence mining
N34	Mich et al. (2013)	x	Web tool	x						Expert-driven	
N35	Mitrovic et al. (2013)	x	ITS		x	x				Expert	
N36	Shekhar et al. (2014)	x	Educational software tool	x		x				Expert	
N37	Zinn (2014)		ITS		x					Expert	
N38	Dzikovska et al. (2014)		ITS	x		x				Data	Natural language processing
N39	Narciss et al. (2014)	x	Intelligent learning environment	x						Expert	
N40	Kão and Niitsoo (2014)	x	Web application				x			Data	Audio processing, time series analysis
N41	Akçapınar (2015)		Course Management System		x					Data	Text mining
N42	Hull and du Boulay (2015)		ITS			x		x		Expert	
N43	Sullins et al. (2015)	x	ITS		x					Data	Natural language processing, text mining
N44	Barmaki and Hughes (2015)	x	Teaching platform			x	x			Data	Audio and video-processing
N45	de Vries et al. (2015)		Computer-assisted language learning system			x	x			Data	Audio and video processing
N46	Fung et al. (2015)	x	System				x			Mixed	Classification, text mining
N47	Grigoriou et al. (2015)		Web-based learning system	x						Expert	
N48	Shaalan et al. (2015)	x	ITS	x		x				Mixed	Natural language processing
N49	Höhn and Ras (2016)		E-learning platform	x		x				Expert	

Table C.16: Classification of the selected papers based on TAF-ClaF - Part1-4

ID	Study	Public availability	Terminology	Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other	Feedback generation model	Technique
				Student data					
N50	Schneider et al. (2016)		Tool				x			Mixed	Audio and video-processing
N51	Wilson and Czik (2016)	x	Automated Essay Evaluation (AEE) system		x					Data-driven	Natural language processing, text mining
N52	Choi (2016)		Intelligent computer-assisted language learning tutoring system	x						Expert	
N53	Frenoy et al. (2016)		Adaptive training system		x	x				Data	Classification
N54	Belcadhi (2016)		Framework	x				x		Expert	
N55	Lukáč and Sekerák (2015)	x	Digital learning environment	x						Expert	
N56	Mirzaeian et al. (2016)		Intelligent Feedback Generator		x					Data	Natural language processing
N57	Sung et al. (2016)		Summary assessment and feedback system		x					Data	Text mining
N58	Sedrakyan et al. (2017)	x	Feedback-inclusive rapid prototyping environment		x					Expert	
N59	Grawemeyer et al. (2017)	x	Intelligent learning environment			x	x	x		Data	Statistical modeling
N60	Correia et al. (2017)	x	Web learning environment		x					Expert	
N61	Hodgkinson et al. (2016)		Automated feedback tool		x					Mixed	Classification, linear transformations
N62	Liu et al. (2017)		Automated feedback system		x					Mixed	Classification
N63	Wang et al. (2017)	x	Feedback generator		x					Data	Optimization
N64	Cheng (2017)	x	Automated feedback tool		x					Data	Text mining

Table C.17: Classification of the selected papers based on TAF-ClaF - Part1-5

ID	Study	Public availability	Terminology	Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other	Feedback generation model	Technique
				Student data							
N65	Dias Pereira dos Santos et al. (2017)	x	Tool				x			Expert	
N66	Cutumisu et al. (2017)		Automated feedback system		x					Expert	
N67	McGuire et al. (2017)	x	Online learning platform	x						Expert	
N68	Perikos et al. (2017)		ITS	x		x				Expert	
N69	Gerdes et al. (2017)	x	ITS		x					Expert	
N70	Karkalas et al. (2016)		Interactive Learning Environment			x				Expert	
N71	Schaffer et al. (2017)		Tool	x		x				Expert	
N72	Ai (2017)		Intelligent computer-assisted language learning environment	x		x		x		Expert	
N73	Zhu et al. (2017)	x	Online module		x					Mixed	Classification, regression
N74	Jiménez et al. (2018)		ITS		x			x		Expert	
N75	Benotti et al. (2018)	x	Web-based coding tool		x					Expert	
N76	Chu et al. (2018)		Techniques		x					Data	Optimization, classification
N77	Demaidi et al. (2018)		VLE	x		x		x		Expert	
N78	Fujita et al. (2018)	x	Learning support system	x						Expert	
N79	Hautala et al. (2018)	x	Virtual tutoring system	x						Expert	
N80	Ho et al. (2018)		Online assessment tool		x					Expert	
N81	Monahan et al. (2018)		Virtual agent		x					Expert	

Table C.18: Classification of the selected papers based on TAF-ClaF - Part1-6

ID	Study	Public availability	Terminology	Student data						Feedback generation model	Technique
				Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other		
N82	Naismith and Lajoie (2017)		Computer learning environment	x						Expert	
N83	Ochoa et al. (2018)		System					x		Mixed	Classification
N84	Rajendran et al. (2018)	x	ITS	x		x				Mixed	Optimization
N85	Cabestrero et al. (2018)		ITS	x		x		x		Mixed	Classification
N86	Hülsmann et al. (2018)		Training environment				x			Data	Time series analysis, classification
N87	Sun et al. (2018)	x	Intelligent learning environment			x				Data	Sequence mining
N88	Holland et al. (2009)		ITS		x					Expert	
N89	Rhienmora et al. (2011)		Intelligent training simulator				x			Mixed	Hidden Markov Models
N90	Chan et al. (2010)		Virtual reality training system				x			Expert	
N91	Ward et al. (2013)		ITS		x					Mixed	Natural language processing, audio and video-processing
N92	Arevalillo-Herráez et al. (2013)		ITS		x	x				Expert	
N93	Wong et al. (2014)		ITS				x			Expert	
N94	Taele et al. (2015)		Educational tool		x					Mixed	Audio and video-processing
N95	Kyan et al. (2015)		Virtual environment				x			Mixed	Self-organizing maps

Table C.19: Classification of the selected papers based on TAF-ClaF - Part1-7

ID	Study	Public availability	Terminology	Student data						Feedback generation model	Technique
				Pre-defined answer	Open answer	In-system behavior	Real life behavior	Student characteristics	Other		
N96	Price et al. (2016)		ITS		x					Data	Tree decomposition
N97	Liu et al. (2016)		E-learning platform			x				Data	Audio and video-processing, classification
N98	Baneres and Saíz (2016)	x	ITS		x					Expert	
N99	Basu et al. (2017)		Open-ended learning environment	x	x	x				Expert	
N100	Arevalillo-Herráez et al. (2017)		ITS		x					Mixed	Classification
N101	VanLehn et al. (2016)		ITS	x						Expert	
N102	Tiam-Lee and Sumi (2018)		System		x	x	x	x		Mixed	Hidden Markov models
N103	Bjorn et al. (2018)		Digital learning environment		x	x				Expert	
N104	Lee et al. (2019)		Feedback system	x	x					Mixed	Natural language processing, regression
N105	Li and Hsieh (2019)		ITS		x	x				Data	Text mining
N106	Qian and Lehman (2019)		Automated assessment system		x					Expert	
N107	Romano et al. (2019)	x	System				x			Mixed	Classification
N108	Silva et al. (2019)		ITS		x					Mixed	Text mining
N109	Wünsche et al. (2019)	x	Interactive web-based tool		x					Expert	

Table C.20: Classification of the selected papers based on TAF-ClaF - Part2-1

ID	Web-based	Connection	Language/tool	Domain	Level	Setting	Adaptiveness	Learner control
N1	x	Plug-in	Python	STEM	Adult education	Online	Student	High
N2		Not specified	Web Ontology Language	STEM	Non-level-specific	Not specified	No	No
N3		Standalone	Other	Other skills	Non-level-specific	In-class	No	No
N4		Standalone	C++, C#	Special education	Non-level-specific	In-class	Task	No
N5		Plug-in	Not specified	STEM	Secondary	In-class	Student & Task	No
N6	x	Standalone	Not specified	Medicine	Tertiary	In-class	Student	No
N7		Standalone	Java	STEM	Tertiary	In-class	No	Mild
N8	x	Standalone	Not specified	Languages	Non-level-specific	Online	Student	No
N9	x	Plug-in	Not specified	STEM	Tertiary	In-class	Student & Task	Mild
N10	x	Plug-in	Not specified	STEM	Secondary	In-class	Student & Task	No
N11	x	Not specified	Not specified	Not domain specific	Non-level-specific	In-class	Student	No
N12		Standalone	Java	STEM	Tertiary	In-class	Student & Task	No
N13		Plug-in	Web Ontology Language	Not domain specific	Tertiary	In-class	No	No
N14	x	Standalone	Not specified	Languages	Tertiary	In-class	No	No
N15		Standalone	Rapid Miner	STEM	Tertiary	In-class	Student & Task	Mild
N16		Standalone	Java	Medicine	Tertiary	In-class	Task	No
N17		Standalone	Not specified	Medicine	Tertiary	In-class	Task	No
N18		Standalone	Java	STEM	Tertiary	In-class	Task	No
N19		Plug-in	Not specified	STEM	Secondary	In-class	Student	Mild
N20		Not specified	Not specified	Languages	Tertiary	In-class	Student & Task	No
N21	x	Not specified	Not specified	STEM	Tertiary	Blended	Student	No
N22		Not specified	Not specified	Art	Non-level-specific	Not specified	Task	High
N23		Standalone	Not specified	Art	Non-level-specific	Not specified	Task	No
N24	x	Standalone	Microsoft Kinect	Other skills	Non-level-specific	Not specified	Task	No
N25		Plug-in	JavaScript	STEM	Non-level-specific	Online	Task	No

Table C.21: Classification of the selected papers based on TAF-ClaF - Part2-2

ID	Web-based	Connection	Language/tool	Domain	Level	Setting	Adaptiveness	Learner control
N26		Standalone	Java	STEM	Adult education	In-class	Student & Task	No
N27	x	Standalone	Web Ontology Language	Not domain specific	Tertiary	In-class	Task	No
N28	x	Standalone	Not specified	Medicine	Tertiary	Online	Student	Mild
N29		Standalone	Not specified	STEM	Non-level-specific	In-class	Task	No
N30		Standalone	Other	Other skills	Secondary	In-class	Task	Mild
N31	x	Standalone	Other	Languages	Tertiary	In-class	Student	No
N32		Standalone	Other	STEM	Secondary	In-class	Student	No
N33		Standalone	Not specified	Medicine	Tertiary	In-class	Student & Task	No
N34	x	Standalone	HTML, JavaScript	Special education	Primary	Online	No	Mild
N35		Standalone	Not specified	STEM	Tertiary	In-class	No	Mild
N36		Plug-in	Not specified	STEM	Tertiary	In-class	No	High
N37		Standalone	Not specified	STEM	Non-level-specific	Not specified	Task	Mild
N38		Standalone	Other	STEM	Tertiary	In-class	Task	No
N39	x	Standalone	Not specified	STEM	Primary	Online	Task	No
N40	x	Standalone	JavaScript, Python	Art	Non-level-specific	Online	Task	No
N41	x	Plug-in	C#, SQL, RapidMiner	Other skills	Tertiary	Blended	Task	No
N42		Standalone	Other	STEM	Tertiary	In-class	Student	Mild
N43		Standalone	Not specified	STEM	Tertiary	Blended	Student & Task	No
N44		Standalone	Microsoft Kinect	Other skills	Adult education	Blended	Task	Mild
N45		Standalone	Not specified	Languages	Adult education	In-class	Task	No
N46	x	Standalone	JavaScript	Other skills	Non-level-specific	Online	Task	Mild
N47	x	Standalone	Not specified	Not domain specific	Non-level-specific	Blended	No	No
N48	x	Standalone	Not specified	Languages	Non-level-specific	Online	Student & Task	No
N49		Standalone	Not specified	Other skills	Adult education	Not specified	Student	No
N50		Standalone	C#, Microsoft Kinect	Other skills	Adult education	Not specified	Task	No

Table C.22: Classification of the selected papers based on TAF-ClaF - Part2-3

ID	Web-based	Connection	Language/tool	Domain	Level	Setting	Adaptiveness	Learner control
N51	x	Standalone	Not specified	Languages	Secondary	In-class	Task	No
N52	x	Standalone	Not specified	Languages	Secondary	In-class	No	No
N53		Plug-in	Other	Other skills	Non-level-specific	Not specified	Student & Task	No
N54	x	Standalone	Not specified	Other skills	Adult education	Online	Student	No
N55	x	Plug-in	JavaScript, HTML	STEM	Secondary	Blended	No	No
N56		Standalone	Java	Languages	Non-level-specific	Not specified	Task	No
N57		Standalone	Other	Languages	Primary	Blended	Task	Mild
N58		Plug-in	Java	STEM	Tertiary	In-class	Task	No
N59		Plug-in	Not specified	STEM	Secondary	In-class	Student & Task	No
N60	x	Plug-in	JavaScript	STEM	Non-level-specific	Not specified	Task	Mild
N61		Plug-in	C++	STEM	Tertiary	Not specified	Task	No
N62	x	Plug-in	Other	Languages	Tertiary	Not specified	Task	No
N63	x	Plug-in	Not specified	STEM	Non-level-specific	Online	Task	No
N64	x	Plug-in	Not specified	Languages	Tertiary	Not specified	Task	No
N65	x	Standalone	Not specified	Art	Non-level-specific	Online	Student & Task	No
N66		Standalone	Not specified	STEM	Secondary	In-class	Task	Mild
N67	x	Standalone	Not specified	STEM	Secondary	Not specified	No	No
N68	x	Standalone	Not specified	STEM	Tertiary	Not specified	No	Mild
N69	x	Standalone	JavaScript	STEM	Tertiary	Not specified	Task	Mild
N70	x	Plug-in	HTML, JavaScript	Not domain specific	Non-level-specific	Not specified	No	Mild
N71	x	Plug-in	Not specified	Not domain specific	Tertiary	Blended	Student	No
N72	x	Standalone	Python, Java, MySQL	Languages	Tertiary	Not specified	Task	No
N73	x	Plug-in	Not specified	STEM	Secondary	Online	Task	No
N74	x	Plug-in	HTML, JavaScript, Python	STEM	Tertiary	Blended	Task	Mild
N75	x	Standalone	Not specified	STEM	Tertiary	Blended	Task	No
N76		Standalone	Not specified	Languages	Tertiary	In-class	Task	No
N77	x	Plug-in	Not specified	Not domain specific	Non-level-specific	Not specified	Student	No

Table C.23: Classification of the selected papers based on TAF-ClaF - Part2-4

ID	Web-based	Connection	Language/tool	Domain	Level	Setting	Adaptiveness	Learner control
N78	x	Standalone	Other	STEM	Secondary	In-class	No	No
N79		Standalone	Not specified	STEM	Primary	Blended	No	No
N80	x	Standalone	Not specified	Medicine	Tertiary	Blended	Task	No
N81		Not specified	Not specified	Other skills	Non-level-specific	In-class	No	No
N82		Standalone	Not specified	Medicine	Tertiary	In-class	No	No
N83		Standalone	C++	Other skills	Tertiary	In-class	Student & Task	No
N84	x	Plug-in	Not specified	STEM	Secondary	Blended	Student	No
N85		Standalone	Not specified	STEM	Secondary	In-class	Student	Mild
N86		Plug-in	Python	Other skills	Non-level-specific	In-class	Task	No
N87	x	Standalone	Not specified	Other skills	Tertiary	Online	Student	No
N88		Standalone	Not specified	STEM	Tertiary	In-class	Task	Mild
N89		Plug-in	C++	Medicine	Tertiary	In-class	Task	No
N90		Standalone	Other	Art	Tertiary	In-class	Task	No
N91		Standalone	Not specified	STEM	Primary	In-class	Task	No
N92		Standalone	Not specified	STEM	Tertiary	In-class	Student & Task	Mild
N93		Plug-in	Not specified	Other skills	Tertiary	In-class	Task	No
N94		Standalone	Not specified	Art	Tertiary	Not specified	Task	No
N95		Plug-in	Microsoft Kinect	Art	Tertiary	In-class	Task	High
N96		Plug-in	Not specified	STEM	Tertiary	Not specified	Task	No
N97	x	Standalone	Not specified	Medicine	In-class	Online	Task	No
N98	x	Plug-in	Not specified	STEM	Tertiary	Online	Task	No
N99		Standalone	Java	STEM	Secondary	In-class	Student & Task	Mild
N100		Standalone	Not specified	STEM	Secondary	In-class	Student & Task	No
N101	x	Standalone	Not specified	STEM	Secondary	In-class	Task	No
N102		Standalone	Other	STEM	Tertiary	In-class	Student & Task	Mild
N103		Standalone	Not specified	Medicine	Tertiary	In-class	Task	No
N104		Standalone	Not specified	Other skills	Secondary	In-class	Task	No
N105	x	Standalone	Not specified	STEM	Tertiary	In-class	Task	Mild
N106		Standalone	Java	STEM	Secondary	In-class	Task	No
N107		Standalone	Kinect SDK	Art	Non-level-specific	Not specified	Student & Task	No
N108		Standalone	Not specified	STEM	Tertiary	In-class	Task	High
N109	x	Plug-in	OpenGL, C++, HTML, JavaScript	STEM	Tertiary	In-class	Task	No

Table C.24: Classification of the selected papers based on TAF-ClaF - Part3-1

ID	Timing			Purpose				Evaluation approach									Positive outcome	# of students	Group size
	Action-based	On request	End-of-task	Informative	Corrective	Suggestive	Motivational	Control group	Student surveys	Pre-& post-test	Statistical testing	Repeated measures	Algorithm performance	Experts evaluation	Behavioral analysis	Other			
N1	x	x	x	x	x	x		x	x									8	Small
N2			x	x	x						No evaluation							No evaluation	
N3	x			x	x	x		x		x	x			x			x	35	Medium
N4	x				x								x				x	2	Small
N5	x			x		x		x	x	x	x						x	42	Medium
N6	x			x	x	x		x	x	x	x						x	85	Medium
N7		x		x	x			x	x		x			x			x	75	Medium
N8			x	x	x						No evaluation							No evaluation	
N9			x	x	x	x			x		x						x	22	Small
N10	x			x		x		x	x								x	584	Massive
N11			x	x	x	x			x								x	500	Massive
N12	x	x	x	x	x	x	x	x	x	x	x			x			x	200	Large
N13			x	x	x	x			x	x							x	21	Small
N14			x	x	x				x			x					x	29	Small
N15	x			x	x	x	x		x				x				x	10	Small
N16			x		x	x		x			x			x			x	10	Small
N17	x			x	x						No evaluation							No evaluation	
N18			x		x				x					x			x	54	Medium
N19	x			x		x		x		x	x	x				x	x	125	Large
N20	x			x	x			x			x		x	x			x	39	Medium
N21	x		x	x	x	x					No evaluation							No evaluation	
N22	x			x	x			x	x					x			x	16	Small
N23	x			x	x				x			x					x	15	Small
N24	x			x	x						No evaluation							No evaluation	
N25		x			x	x					No evaluation							No evaluation	
N26	x	x		x	x	x			x			x					x	10	Small
N27			x	x	x	x			x	x	x			x			x	22	Small
N28	x		x	x	x	x	x				No evaluation							No evaluation	
N29			x		x	x					No evaluation							No evaluation	
N30			x	x		x		x	x		x	x					x	65	Medium
N31	x			x	x	x		x	x	x							x	45	Medium
N32	x			x	x	x		x							x		x	93	Medium
N33	x				x	x					No evaluation							No evaluation	
N34	x				x	x	x	x			x						x	30	Small
N35	x	x			x		x	x		x	x				x		x	41	Medium

72

Table C.25: Classification of the selected papers based on TAF-ClaF - Part3-2

ID	Action-based	On request	End-of-task	Informative	Corrective	Suggestive	Motivational	Control group	Student surveys	Pre-& post-test	Statistical testing	Repeated measures	Algorithm performance	Experts evaluation	Behavioral analysis	Other	Positive outcome	# of students	Group size
		Timing			Purpose							Evaluation approach							
N36	x	x	x	x	x	x						No evaluation							
N37	x	x			x							No evaluation							
N38			x	x	x	x	x	x	x	x	x		x				x	120	Large
N39			x	x	x	x		x	x	x	x				x			186	Large
N40			x	x	x					x		x		x			x	108	Large
N41	x				x	x				x	x						x	59	Medium
N42		x	x				x	x	x	x	x				x		x	36	Medium
N43	x			x	x	x	x	x		x	x						x	17	Small
N44	x	x			x			x	x								x	34	Medium
N45	x				x	x	x	x	x	x	x		x		x		x	29	Small
N46			x	x		x		x	x								x	23	Small
N47			x	x	x				x		x						x	65	Medium
N48	x				x	x							x				x	No	No
N49			x		x	x						No evaluation							
N50	x			x	x			x	x		x				x		x	40	Medium
N51			x		x	x		x	x		x						x	145	Large
N52	x				x	x		x	x	x	x						x	156	Large
N53	x				x	x		x		x	x	x				x	x	44	Medium
N54			x	x	x	x			x								x	No	No
N55	x				x	x						No evaluation							
N56	x			x	x										x			No	No
N57		x	x	x	x			x		x	x	x		x			x	154	Large
N58	x				x	x		x	x	x	x						x	65	Medium
N59	x				x	x	x	x		x	x				x		x	77	Medium
N60		x		x	x							No evaluation							
N61			x	x	x			x					x	x			x	19	Small
N62	x				x	x		x	x		x			x	x		x	110	Large
N63			x		x	x							x				x	3361	Massive
N64			x	x	x	x		x	x		x	x			x		x	138	Large
N65			x	x	x				x								x	7	Small
N66			x	x	x		x		x		x							89	Medium
N67	x			x	x	x		x			x							196	Large
N68	x	x	x	x	x	x		x	x	x	x						x	226	Large
N69		x	x	x	x	x			x				x	x	x		x	116	Large
N70		x	x	x	x		x		x						x		x	8	Small

Table C.26: Classification of the selected papers based on TAF-ClaF - Part3-3

ID	Timing			Purpose				Evaluation approach									Positive outcome	# of students	Group size
	Action-based	On request	End-of-task	Informative	Corrective	Suggestive	Motivational	Control group	Student surveys	Pre-& post-test	Statistical testing	Repeated measures	Algorithm performance	Experts evaluation	Behavioral analysis	Other			
N71			x	x	x			No evaluation											
N72	x				x	x			x						x		x	6	Small
N73			x	x	x	x			x	x	x		x	x	x		x	183	Large
N74		x		x			x		x		x						x	40	Medium
N75			x		x	x	x	x	x		x	x					x	114	Large
N76			x		x			x	x				x				x	19	Small
N77	x				x	x		x	x	x	x		x	x			x	88	Large
N78			x	x	x										x		x	15	Small
N79	x			x	x		x		x	x	x						x	61	Medium
N80	x				x				x		x			x			x	137	Large
N81	x			x				x			x						x	60	Medium
N82			x	x	x				x		x	x			x		x	28	Small
N83			x	x	x	x			x					x			x	83	Medium
N84	x						x				x	x					x	188	Large
N85	x	x	x		x	x	x	x			x	x			x			48	Medium
N86	x				x	x							x				x	74	Medium
N87	x				x	x	x	x	x		x	x						191	Large
N88	x			x	x		x	x	x	x							x	26	Small
N89	x				x	x	x						x	x			x	5	Small
N90	x		x	x	x			x	x	x	x		x				x	6	Small
N91	x			x		x		x	x	x	x		x				x	1478	Massive
N92	x	x	x	x	x	x		x		x	x	x					x	38	Medium
N93	x				x	x		x	x		x			x			x	20	Small
N94			x	x	x	x						x	x				x	10	Small
N95	x		x		x	x	x					x	x				x	3	Small
N96	x				x	x	x				x		x	x			x	51	Medium
N97			x	x					x	x			x	x			x	18	Small
N98			x		x				x							x	x	508	Massive
N99	x	x	x	x	x	x		x		x	x	x					x	98	Medium
N100	x		x		x	x		x		x	x	x	x				x	48	Medium
N101	x				x	x		x			x						x	34	Medium
N102			x	x	x	x		x	x				x				x	35	Medium
N103	x				x	x		x	x								x	19	Small
N104			x	x	x	x				x	x				x		x	343	Large
N105		x	x	x	x	x		x	x	x	x		x				x	93	Medium
N106			x	x	x	x		x			x						x	23	Small
N107	x		x	x	x	x			x		x		x				x	25	Small
N108	x	x		x	x					x	x						x	34	Medium
N109			x	x	x				x								x	300	Large

References

Ai, H., 2017. Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. ReCALL 29, 313–334.

Akçapınar, G., 2015. How automated feedback through text mining changes plagiaristic behavior in online assignments. Computers & Education 87, 123–130.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C., 2001. A taxonomy for learning, teaching, and assessing: A revision of blooms taxonomy of educational objectives, abridged edition. White Plains, NY: Longman .

Arevalillo-Herráez, M., Arnau, D., Ferri, F.J., Santos, O.C., 2017. Gui-driven intelligent tutoring system with affective support to help learning the algebraic method, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE. pp. 2867–2872.

Arevalillo-Herráez, M., Arnau, D., Marco-Giménez, L., 2013. Domain-specific knowledge representation and inference engine for an intelligent tutoring system. Knowledge-Based Systems 49, 97–105.

Bandara, W., Miskon, S., Fielt, E., 2011. A systematic, tool-supported method for conducting literature reviews in information systems, in: ECIS 2011 Proceedings [19th European Conference on Information Systems]:, AIS Electronic Library (AISeL)/Association for Information Systems. pp. 1–13.

Baneres, D., Saíz, J., 2016. Intelligent tutoring system for learning digital systems on mooc environments, in: 2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), IEEE. pp. 95–102.

Barker, T., 2010. An automated feedback system based on adaptive testing: extending the model. International Journal of Emerging Technologies in Learning (iJET) 5.

Barmaki, R., Hughes, C.E., 2015. Providing real-time feedback for student teachers in a virtual rehearsal environment, in: Proceedings of the 2015

ACM on International Conference on Multimodal Interaction, ACM. pp. 531–537.

Basu, S., Biswas, G., Kinnebrew, J.S., 2017. Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. User Modeling and User-Adapted Interaction 27, 5–53.

Belcadhi, L.C., 2016. Personalized feedback for self assessment in lifelong learning environments based on semantic web. Computers in Human Behavior 55, 562–570.

Benotti, L., Aloi, F., Bulgarelli, F., Gomez, M.J., 2018. The effect of a web-based coding tool with automatic feedback on students' performance and perceptions, in: Proceedings of the 49th ACM Technical Symposium on Computer Science Education, ACM. pp. 2–7.

Bimba, A.T., Idris, N., Al-Hunaiyyan, A., Mahmud, R.B., Shuib, N.L.B.M., 2017. Adaptive feedback in computer-based learning environments: a review. Adaptive Behavior 25, 217–234.

Bjorn, M., Ravyse, W., Villafruella, D.S., Luimula, M., Leivo, S., 2018. Higher education learner experience with fuzzy feedback in a digital learning environment, in: 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), IEEE. pp. 000253–000260.

Bloom, B.S., 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational researcher 13, 4–16.

Boud, D., Molloy, E., 2013. Rethinking models of feedback for learning: the challenge of design. Assessment & Evaluation in Higher Education 38, 698–712.

Cabestrero, R., Quirós, P., Santos, O.C., Salmeron-Majadas, S., Uria-Rivas, R., Boticario, J.G., Arnau, D., Arevalillo-Herráez, M., Ferri, F.J., 2018. Some insights into the impact of affective information when delivering feedback to students. Behaviour & Information Technology 37, 1252–1263.

Caiza, J.C., Del Alamo, J.M., 2013. Programming assignments automatic grading: review of tools and implementations, in: 7th International Technology, Education and Development Conference (INTED2013), pp. 5691–5700.

Calvo, R.A., Ellis, R.A., 2010. Students' conceptions of tutor and automated feedback in professional writing. Journal of Engineering Education 99, 427–438.

Chan, J.C., Leung, H., Tang, J.K., Komura, T., 2010. A virtual reality dance training system using motion capture technology. IEEE Transactions on Learning Technologies 4, 187–195.

Cheng, G., 2017. The impact of online automated feedback on students' reflective journal writing in an efl course. The Internet and Higher Education 34, 18–27.

Choi, I.C., 2016. Efficacy of an icall tutoring system and process-oriented corrective feedback. Computer Assisted Language Learning 29, 334–364.

Chu, T., Taele, P., Hammond, T., 2018. Supporting chinese character educational interfaces with richer assessment feedback through sketch recognition .

Chung, Z., Nagai, H., Rodrigo, M.M.T., 2013. An intelligent tutoring system for japanese language particles with user assessment and feedback., in: AIED Workshops, Citeseer.

Cook, D.A., Erwin, P.J., Triola, M.M., 2010. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Academic Medicine 85, 1589–1602.

Correia, H.P., Leal, J.P., Paiva, J.C., 2017. Enhancing feedback to students in automated diagram assessment .

Cutumisu, M., Blair, K.P., Chin, D.B., Schwartz, D.L., 2017. Assessing whether students seek constructive criticism: The design of an automated feedback system for a graphic design task. International Journal of Artificial Intelligence in Education 27, 419–447.

Demaidi, M.N., Gaber, M.M., Filer, N., 2018. Ontopefege: Ontology-based personalized feedback generator. IEEE Access 6, 31644–31664.

Di Eugenio, B., Fossati, D., Haller, S., Yu, D., Glass, M., 2008. Be brief, and they shall learn: Generating concise language feedback for a computer tutor. International Journal of Artificial Intelligence in Education 18, 317–345.

Dominguez, A.K., Yacef, K., Curran, J., 2010. Data mining to generate individualised feedback, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 303–305.

Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., Campbell, G., 2014. Beetle ii: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. International Journal of Artificial Intelligence in Education 24, 284–332.

Economides, A.A., 2005. Personalized feedback in cat. WSEAS Transactions on Advances in Engineering Education 2, 174.

Ferguson, P., 2011. Student perceptions of quality feedback in teacher education. Assessment & Evaluation in Higher Education 36, 51–62.

Fleming, M.L., Levie, W.H., 1978. Instructional message design: Principles from the behavioral sciences. Educational Technology.

Fossati, D., Di Eugenio, B., Ohlsson, S., Brown, C., Chen, L., 2010. Generating proactive feedback to help students stay on track, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 315–317.

Fossati, D., Di Eugenio, B., Ohlsson, S., Brown, C., Chen, L., 2015. Data driven automatic feedback generation in the ilist intelligent tutoring system. Technology, Instruction, Cognition and Learning 10, 5–26.

Fournier-Viger, P., Nkambou, R., Mayers, A., Mephu-Nguifo, E., Faghihi, U., 2012. Multi-paradigm generation of tutoring feedback in robotic arm manipulation training, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 233–242.

Frenoy, R., Soullard, Y., Thouvenin, I., Gapenne, O., 2016. Adaptive training environment without prior knowledge: Modeling feedback selection as a multi-armed bandit problem, in: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, ACM. pp. 131–139.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The elements of statistical learning. volume 1. Springer series in statistics New York.

Frutos-Morales, F., Sánchez-Vera, M., Castellanos-Nieves, D., Esteban-Gil, A., Cruz-Corona, C., Prendes-Espinosa, M., Fernández-Breis, J., 2010. An

extension of the oele platform for generating semantic feedback for students and teachers. Procedia-Social and behavioral sciences 2, 527–531.

Fujita, T., Jones, K., Miyazaki, M., 2018. Learners use of domain-specific computer-based feedback to overcome logical circularity in deductive proving in geometry. ZDM , 1–15.

Fung, M., Jin, Y., Zhao, R., Hoque, M.E., 2015. Roc speak: semi-automated personalized feedback on nonverbal behavior from recorded videos, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM. pp. 1167–1178.

Gallien, T., Oomen-Early, J., 2008. Personalized versus collective instructor feedback in the online courseroom: Does type of feedback affect student satisfaction, academic performance and perceived connectedness with the instructor? International Journal on E-learning 7, 463–476.

Gerdes, A., Heeren, B., Jeuring, J., van Binsbergen, L.T., 2017. Ask-elle: an adaptable programming tutor for haskell giving automated feedback. International Journal of Artificial Intelligence in Education 27, 65–100.

Gimeno Sanz, A., De-Sequeira, J.M., 2009. Designing feedback to support language acquisition using the "ingenio" authoring tool. Online Submission 1, 1239–1243.

Golonka, E.M., Bowles, A.R., Frank, V.M., Richardson, D.L., Freynik, S., 2014. Technologies for foreign language learning: a review of technology types and their effectiveness. Computer Assisted Language Learning 27, 70–105.

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., Rummel, N., 2017. Affective learning: improving engagement and enhancing learning with affect-aware feedback. User Modeling and User-Adapted Interaction 27, 119–158.

Grigoriou, T., Cheong, C., Cheong, F., 2015. Improving quality of feedback using a technology-supported learning system., in: PACIS, p. 86.

Gutierrez, F., Atkinson, J., 2011. Adaptive feedback selection for intelligent tutoring systems. Expert Systems with Applications 38, 6146–6152.

Hattie, J., Timperley, H., 2007. The power of feedback. Review of educational research 77, 81–112.

Hautala, J., Baker, D.L., Keurulainen, A., Ronimus, M., Richardson, U., Cole, R., 2018. Early science learning with a virtual tutor through multimedia explanations and feedback on spoken questions. Educational Technology Research and Development 66, 403–428.

Hepplestone, S., Holden, G., Irwin, B., Parkin, H.J., Thorpe, L., 2011. Using technology to encourage student engagement with feedback: a literature review. Research in Learning Technology 19.

Hevner, A., Chatterjee, S., 2010. Design science research in information systems, in: Design research in information systems. Springer, pp. 9–22.

Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design science in information systems research. MIS Quarterly 28, 75–105. URL: http://www.jstor.org/stable/25148625.

Ho, V.W., Harris, P.G., Kumar, R.K., Velan, G.M., 2018. Knowledge maps: a tool for online assessment with automated feedback. Medical education online 23, 1457394.

Hodgkinson, B., Lutteroth, C., Wünsche, B., 2016. glgetfeedbacktowards automatic feedback and assessment for opengl 3d modelling assignments, in: Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on, IEEE. pp. 1–6.

Höhn, S., Ras, E., 2016. Designing formative and adaptive feedback using incremental user models, in: International Conference on Web-Based Learning, Springer. pp. 172–177.

Holland, J., Mitrovic, A., Martin, B., 2009. J-latte: a constraint-based tutor for java .

Hull, A., du Boulay, B., 2015. Motivational and metacognitive feedback in sql-tutor. Computer Science Education 25, 238–256.

Hülsmann, F., Göpfert, J.P., Hammer, B., Kopp, S., Botsch, M., 2018. Classification of motor errors to provide real-time feedback for sports coaching in virtual realitya case study in squats and tai chi pushes. Computers & Graphics 76, 47–59.

Hwang, G.J., 2014. Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. Smart Learning Environments 1, 4.

Ihantola, P., Ahoniemi, T., Karavirta, V., Seppälä, O., 2010. Review of recent systems for automatic assessment of programming assignments, in: Proceedings of the 10th Koli calling international conference on computing education research, pp. 86–93.

Inventado, P.S., Legaspi, R., Bui, T., Suarez, M., 2010. Predicting students appraisal of feedback in an its using previous affective states and continuous affect labels from eeg data, in: Proceedings of the 18th International Conference on Computers in Education, Putrajaya, Malaysia.

Jiménez, S., Juárez-Ramírez, R., Castillo, V.H., Licea, G., Ramírez-Noriega, A., Inzunza, S., 2018. A feedback system to provide affective support to students. Computer Applications in Engineering Education 26, 473–483.

Johnson, S., Zaiane, O., 2012. Intelligent feedback polarity and timing selection in the shufti intelligent tutoring system, in: International Conference on Computers in Education, Citeseer.

Jordan, P.W., Albacete, P.L., Katz, S., Ford, M., Lipschultz, M., 2013. Eliciting student explanations during tutorial dialogue for the purpose of providing formative feedback., in: AIED Workshops.

Jylhä, A., Erkut, C., 2011. Auditory feedback in an interactive rhythmic tutoring system, in: Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound, ACM. pp. 109–115.

Kão, K., Niitsoo, M., 2014. Matchmysound: Introducing feedback to online music education, in: International Conference on Web-Based Learning, Springer. pp. 217–225.

Karavirta, V., Helminen, J., Ihantola, P., 2012. A mobile learning application for parsons problems with automatic feedback, in: Proceedings of the 12th Koli Calling International Conference on Computing Education Research, ACM. pp. 11–18.

Karkalas, S., Mavrikis, M., Xenos, M., Kynigos, C., 2016. Feedback authoring for exploratory activities: The case of a logo-based 3d microworld, in:

International Conference on Computer Supported Education, Springer. pp. 259–278.

Kazi, H., Haddawy, P., Suebnukarn, S., 2010. Leveraging a domain ontology to increase the quality of feedback in an intelligent tutoring system, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 75–84.

Kazi, H., Haddawy, P., Suebnukarn, S., 2012. Employing umls for generating hints in a tutoring system for medical problem-based learning. Journal of biomedical informatics 45, 557–565.

Kelly, D., McDonald, J., Markham, C., 2008. A system for teaching sign language using live gesture feedback, in: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, IEEE. pp. 1–2.

Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., O'Leary, S., 2013. Mining interactions in immersive learning environments for real-time student feedback. Australasian Journal of Educational Technology 29.

Keuning, H., Jeuring, J., Heeren, B., 2018. A systematic literature review of automated feedback generation for programming exercises. ACM Transactions on Computing Education (TOCE) 19, 3.

Kitchenham, B., 2004. Procedures for performing systematic reviews. Keele, UK, Keele University 33, 1–26.

Kyan, M., Sun, G., Li, H., Zhong, L., Muneesawang, P., Dong, N., Elder, B., Guan, L., 2015. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. ACM Transactions on Intelligent Systems and Technology (TIST) 6, 1–37.

Le, N.T., Strickroth, S., Gross, S., Pinkwart, N., 2013. A review of ai-supported tutoring approaches for learning programming, in: Advanced Computational Methods for Knowledge Engineering. Springer, pp. 267–279.

Lee, H.S., Pallant, A., Pryputniewicz, S., Lord, T., Mulholland, M., Liu, O.L., 2019. Automated text scoring and real-time adjustable feedback: Supporting revision of scientific arguments involving uncertainty. Science Education 103, 590–622.

Li, Q., Hsieh, S.J., 2019. An intelligent tutoring system for computer numerical control programming. The International journal of engineering education 35, 252–261.

Liu, B., Chen, H., He, W., 2008. A framework of deriving adaptive feedback from educational ontologies, in: Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for, IEEE. pp. 2476–2480.

Liu, C., Calvo, R.A., Lim, R., 2016. Improving medical students awareness of their non-verbal communication through automated non-verbal behavior feedback. Frontiers in ICT 3, 11.

Liu, M., Li, Y., Xu, W., Liu, L., 2017. Automated essay feedback generation and its impact on revision. IEEE Transactions on Learning Technologies 10, 502–513.

Lu, X., Di Eugenio, B., Ohlsson, S., Fossati, D., 2008. Simple but effective feedback generation to tutor abstract problem solving, in: Proceedings of the Fifth International Natural Language Generation Conference, Association for Computational Linguistics. pp. 104–112.

Lukáč, S., Sekerák, J., 2015. Implementation of feedback in digital learning environment with a focus on system geogebra, in: Emerging eLearning Technologies and Applications (ICETA), 2015 13th International Conference on, IEEE. pp. 1–7.

Ma, X., Wijewickrema, S., Zhou, Y., Zhou, S., OLeary, S., Bailey, J., 2017. Providing effective real-time feedback in simulation-based surgical training, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 566–574.

del Mar Sánchez-Vera, M., Fernández-Breis, J.T., Castellanos-Nieves, D., Frutos-Morales, F., Prendes-Espinosa, M.P., 2012. Semantic web technologies for generating feedback in online assessment environments. Knowledge-Based Systems 33, 152–165.

Marcus, N., Ben-Naim, D., Bain, M., 2011. Instructional support for teachers and guided feedback for students in an adaptive elearning environment, in: Information Technology: New Generations (ITNG), 2011 Eighth International Conference on, IEEE. pp. 626–631.

Martin, F., Chen, Y., Moore, R.L., Westine, C.D., 2020. Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. Educational Technology Research and Development 68, 1903–1929.

McGuire, P., Tu, S., Logue, M.E., Mason, C.A., Ostrow, K., 2017. Counterintuitive effects of online feedback in middle school math: results from a randomized controlled trial in assistments. Educational Media International 54, 231–244.

Mich, O., Pianta, E., Mana, N., 2013. Interactive stories and exercises with dynamic feedback for improving reading comprehension skills in deaf children. Computers & Education 65, 34–44.

Mirzaeian, V.R., Kohzadi, H., Azizmohammadi, F., 2016. Learning persian grammar with the aid of an intelligent feedback generator. Engineering Applications of Artificial Intelligence 49, 167–175.

Mitrovic, A., Ohlsson, S., Barrow, D.K., 2013. The effect of positive feedback in a constraint-based intelligent tutoring system. Computers & Education 60, 264–272.

Monahan, S., Johnson, E., Lucas, G., Finch, J., Gratch, J., 2018. Autonomous agent that provides automated feedback improves negotiation skills, in: International Conference on Artificial Intelligence in Education, Springer. pp. 225–229.

Naismith, L.M., Lajoie, S.P., 2017. Motivation and emotion predict medical students attention to computer-based feedback. Advances in Health Sciences Education , 1–21.

Narciss, S., 2008. Feedback strategies for interactive learning tasks. Handbook of research on educational communications and technology 3, 125–144.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., Melis, E., 2014. Exploring feedback and student characteristics relevant for personalizing feedback strategies. Computers & Education 71, 56–76.

Ng, K., 2011. Interactive multimedia for technology-enhanced learning with multimodal feedback, in: Musical Robots and Interactive Multimodal Systems. Springer, pp. 105–126.

Nguyen, A.T., Chen, W., Rauterberg, M., 2012. Online feedback system for public speakers, in: E-Learning, E-Management and E-Services (IS3e), 2012 IEEE Symposium on, IEEE. pp. 1–5.

Normadhi, N.B.A., Shuib, L., Nasir, H.N.M., Bimba, A., Idris, N., Balakrishnan, V., 2019. Identification of personal traits in adaptive learning environment: Systematic literature review. Computers & Education 130, 168–190.

O'Brien, H.L., Toms, E.G., 2008. What is user engagement? a conceptual framework for defining user engagement with technology. Journal of the American society for Information Science and Technology 59, 938–955.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., Castells, J., 2018. The rap system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors, in: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, ACM. pp. 360–364.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., Mirriahi, N., 2019. Using learning analytics to scale the provision of personalised feedback. British Journal of Educational Technology 50, 128–138.

Parvez, S.M., Blank, G.D., 2008. Individualizing tutoring with learning style based feedback, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 291–301.

Perikos, I., Grivokostopoulou, F., Hatzilygeroudis, I., 2017. Assistance and feedback mechanism in an intelligent tutoring system for teaching conversion of natural language into logic. International Journal of Artificial Intelligence in Education 27, 475–514.

Perlmuter, L.C., Monty, R.A., 1977. The importance of perceived control: Fact or fantasy? experiments with both humans and animals indicate that the mere illusion of control significantly improves performance in a variety of situations. American Scientist 65, 759–765.

Price, T.W., Dong, Y., Barnes, T., 2016. Generating data-driven hints for open-ended programming. International Educational Data Mining Society .

Pridemore, D.R., Klein, J.D., 1991. Control of feedback in computer-assisted instruction. Educational Technology Research and Development 39, 27–32.

Qian, Y., Lehman, J.D., 2019. Using targeted feedback to address common student misconceptions in introductory programming: A data-driven approach. SAGE Open 9, 2158244019885136.

Rahati, A., Kabanza, F., 2010. Delivering tutoring feedback using persuasive dialogues, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 251–253.

Rajendran, R., Iyer, S., Murthy, S., 2018. Personalized affective feedback to address students frustration in its. IEEE Transactions on Learning Technologies .

Rhienmora, P., Haddawy, P., Suebnukarn, S., Dailey, M.N., 2011. Intelligent dental training simulator with objective skill assessment and feedback. Artificial intelligence in medicine 52, 115–121.

Rivers, K., Koedinger, K.R., 2013. Automatic generation of programming feedback: A data-driven approach, in: The First Workshop on AI-supported Education for Computer Science (AIEDCS 2013).

Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R., 2011. Improving students help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learning and Instruction 21, 267–280.

Romano, G., Schneider, J., Drachsler, H., 2019. Dancing salsa with machinesfilling the gap of dancing learning solutions. Sensors 19, 3661.

Roscoe, R.D., Snow, E.L., McNamara, D.S., 2013. Feedback and revising in an intelligent tutoring system for writing strategies, in: International Conference on Artificial Intelligence in Education, Springer. pp. 259–268.

Roselli, R., Gilbert, S., Howard, L., Blessing, S., Raut, A., Pandian, P., 2008. Integration of an intelligent tutoring system with a web-based authoring system to develop online homework assignments with formative feedback, in: American Society for Engineering Education 2008 Conference.

Dias Pereira dos Santos, A., Yacef, K., Martinez-Maldonado, R., 2017. Forró trainer: Automated feedback for partner dance learning, in: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, ACM. pp. 103–104.

Schaffer, H.E., Young, K.R., Ligon, E.W., Chapman, D.D., 2017. Automating individualized formative feedback in large classes based on a directed concept graph. Frontiers in psychology 8, 260.

Schneider, J., Börner, D., Van Rosmalen, P., Specht, M., 2016. Can you help me with my pitch? studying a tool for real-time automated feedback. IEEE Transactions on Learning Technologies 9, 318–327.

Sedrakyan, G., Poelmans, S., Snoeck, M., 2017. Assessing the influence of feedback-inclusive rapid prototyping on understanding the semantics of parallel uml statecharts by novice modellers. Information and Software Technology 82, 159–172.

Seechaliao, T., 2017. Instructional strategies to support creativity and innovation in education. Journal of education and learning 6, 201–208.

Serral, E., Snoeck, M., 2016. Conceptual framework for feedback automation in sles, in: Smart Education and e-Learning 2016. Springer, pp. 97–107.

Serral Asensio, E., Ruiz, J., Elen, J., Snoeck, M., 2019. Conceptualizing the domain of automated feedback for learners, in: Proceedings of the XXII Iberoamerican Conference on Software Engineering, CIbSE 2019,, Curran Associates 2019.

Shaalan, K., Magdy, M., Fahmy, A., 2015. Analysis and feedback of erroneous arabic verbs. Natural Language Engineering 21, 271–323.

Shekhar, V.S., Agarwalla, A., Agarwal, A., Nitish, B., Kumar, V., 2014. Enhancing jflap with automata construction problems and automated feedback, in: Contemporary Computing (IC3), 2014 Seventh International Conference on, IEEE. pp. 19–23.

Shute, V.J., 2008. Focus on formative feedback. Review of educational research 78, 153–189.

Silva, P., Costa, E., de Araújo, J.R., 2019. An adaptive approach to provide feedback for students in programming problem solving, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 14–23.

Singh, R., Gulwani, S., Solar-Lezama, A., 2013. Automated feedback generation for introductory programming assignments. ACM SIGPLAN Notices 48, 15–26.

Stranieri, A., Yearwood, J., 2008. Enhancing learning outcomes with an interactive knowledge-based learning environment providing narrative feedback. Interactive Learning Environments 16, 265–281.

Sullins, J., Craig, S.D., Hu, X., 2015. Exploring the effectiveness of a novel feedback mechanism within an intelligent tutoring system. International Journal of Learning Technology 10, 220–236.

Sun, J.C.Y., Yu, S.J., Chao, C.H., 2018. Effects of intelligent feedback on online learners engagement and cognitive load: the case of research ethics education. Educational Psychology , 1–18.

Sung, Y.T., Liao, C.N., Chang, T.H., Chen, C.L., Chang, K.E., 2016. The effect of online summary assessment and feedback system on the summary writing on 6th graders: The lsa-based technique. Computers & Education 95, 1–18.

Suselo, T., Wünsche, B.C., Luxton-Reilly, A., 2019. Technologies and tools to support teaching and learning computer graphics: A literature review, in: Proceedings of the Twenty-First Australasian Computing Education Conference, Association for Computing Machinery, New York, NY, USA. p. 96105. URL: `https://doi.org/10.1145/3286960.3286972`, doi:10.1145/3286960.3286972.

Taele, P., Barreto, L., Hammond, T., 2015. Maestoso: an intelligent educational sketching tool for learning music theory, in: Twenty-Seventh IAAI Conference.

Tiam-Lee, T.J., Sumi, K., 2018. Adaptive feedback based on student emotion in a system for programming practice, in: International Conference on Intelligent Tutoring Systems, Springer. pp. 243–255.

VanLehn, K., 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist 46, 197–221.

VanLehn, K., Wetzel, J., Grover, S., Van De Sande, B., 2016. Learning how to construct models of dynamic systems: an initial evaluation of the dragoon intelligent tutoring system. IEEE Transactions on Learning Technologies 10, 154–167.

Voerman, L., Meijer, P.C., Korthagen, F.A., Simons, R.J., 2012. Types and frequencies of feedback interventions in classroom interaction in secondary education. Teaching and Teacher Education 28, 1107–1115.

de Vries, B.P., Cucchiarini, C., Bodnar, S., Strik, H., van Hout, R., 2015. Spoken grammar practice and feedback in an asr-based call system. Computer Assisted Language Learning 28, 550–576.

Wang, K., Lin, B., Rettig, B., Pardi, P., Singh, R., 2017. Data-driven feedback generator for online programing courses, in: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, ACM. pp. 257–260.

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., Weston, T., 2013. My science tutor: A conversational multimedia virtual tutor. Journal of Educational Psychology 105, 1115.

Watson, C., Li, F.W., Lau, R.W., 2011. Learning programming languages through corrective feedback and concept visualisation, in: International Conference on Web-Based Learning, Springer. pp. 11–20.

Webster, J., Watson, R.T., 2002. Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly , xiii–xxiii.

Wilson, J., Czik, A., 2016. Automated essay evaluation software in english language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. Computers & Education 100, 94–109.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering, in: Proceedings of the 18th international conference on evaluation and assessment in software engineering, pp. 1–10.

Wong, J.H., Ogren, L., Peters, S., Bratt, E.O., Newport, R., 2014. Developing and evaluating an intelligent tutoring system for advanced shiphandling, in: Proceedings of the Interservice/Industry, Training, Simulation, and Education Conference (I/ITSEC) Annual Meeting.

Wong, W.K., Yang, M.L., Chaoc, T.K., 2010. A language learning system with automatic feedback: An application based on a english-chinese parallel corpus .

Wünsche, B.C., Huang, E., Shaw, L., Suselo, T., Leung, K.C., Dimalen, D., van der Mark, W., Luxton-Reilly, A., Lobb, R., 2019. Coderunnergl-an interactive web-based tool for computer graphics teaching and assessment, in: 2019 International Conference on Electronics, Information, and Communication (ICEIC), IEEE. pp. 1–7.

Zacharia, Z.C., Manoli, C., Xenofontos, N., De Jong, T., Pedaste, M., van Riesen, S.A., Kamp, E.T., Mäeots, M., Siiman, L., Tsourlidaki, E., 2015. Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. Educational technology research and development 63, 257–302.

Zhu, M., Lee, H.S., Wang, T., Liu, O.L., Belur, V., Pallant, A., 2017. Investigating the impact of automated feedback on students scientific argumentation. International Journal of Science Education 39, 1648–1668.

Zinn, C., 2014. Algorithmic debugging and literate programming to generate feedback in intelligent tutoring systems, in: Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz), Springer. pp. 37–48.