

Adaptive learning module for a conversational agent to support MOOC learners

Nuria González-Castro, Pedro J. Muñoz-Merino, Carlos Alario-Hoyos, Carlos Delgado Kloos
Universidad Carlos III de Madrid

Massive open online courses (MOOCs) pose a challenge for instructors when trying to provide personalised support to learners, due to large numbers of registered participants. Conversational agents can be of help to support learners when working with MOOCs. This article presents an adaptive learning module for JavaPAL, a conversational agent that complements a MOOC on Java programming, helping learners review the key concepts of the MOOC. This adaptive learning module adapts the difficulty of the questions provided to learners considering their level of knowledge using item response theory (IRT) and also provides recommendations of video fragments extracted from the MOOC for when learners fail questions. The adaptive learning module for JavaPAL has been evaluated showing good usability and learnability through the system usability scale (SUS), reasonably suitable video fragments recommendations for learners, and useful visualisations generated as part of the IRT-based adaptation of questions for instructors to better understand what is happening in the course, to design exams, and to redesign the course content.

Implications for practice or policy:

- A conversational agent that adapts the questions provided to learners using Item Response Theory (IRT) can be helpful for learners to review the concepts of a MOOC.
- A conversational agent that provides video fragments recommendations can be helpful for learners to improve their performance when answering questions from a MOOC.
- IRT-based visualisations of item characteristic curves and item information curves can be helpful to redesign the contents of a MOOC.

Keywords: conversational agent, adaptive learning, Item Response Theory, MOOC, Java programming, expert evaluation

Introduction

Massive open online courses (MOOCs) require learners to have certain self-regulated learning skills, as they must work more autonomously compared to other types of courses (Littlejohn et al., 2016). Large numbers of registered participants become a problem when instructors try to offer some personalised support to MOOC learners (Atiaja & Proenza, 2016). The combination of limited support from instructors and the need for self-regulated learning skills in MOOCs results in low average retention rates and higher achievement by individuals who already have higher education studies (Jung et al., 2019), which is against the initial idea of the “democratization of higher education” announced with the emergence of MOOCs (Littlejohn & Hood, 2018, p. 22). Therefore, there are research opportunities related to supporting learners, in different ways, when they are working in a MOOC.

Conversational agents are a technology that has increased in popularity in recent years, and has already been explored to accompany learners in MOOCs (Caballé, & Conesa, 2018; Demetriadis et al. 2018). For example, conversational agents may be used by learners to ask questions about the MOOC and get quick answers, since the instructor cannot answer all the questions posed by learners, while getting answers from peers in the course forum may not be as immediate. Alternatively, conversational agents may accompany learners taking a MOOC, helping them to review the main concepts or to search more easily information on certain concepts among the available course materials. In any case, the potential of conversational agents lies in the possibility of communication using natural language, either through a text-based interface (e.g., chatbots) or a voice-based interface (e.g., virtual assistants), instead of the traditional web-based interface used in MOOCs. A voice-based interface may be particularly useful in cases where learners cannot use their hands due to other duties, for example, while driving to work or college.

JavaPAL is an example of a voice-based conversational agent that complements a MOOC on Java programming and aims to help learners review the main concepts of the course (Catalán Aguirre et al., 2018; Delgado Kloos et al., 2019). JavaPAL asks multiple-choice questions to learners related to the topics of that MOOC and provides definitions of related concepts upon learners' requests as a way to help learners understand and practice these concepts. Nevertheless, JavaPAL, as presented in its initial form, has important limitations. A first limitation is that JavaPAL asks random multiple-choice questions to learners, so these may receive questions far above or below their current level of knowledge. As a consequence, learners with limited knowledge may receive questions that they might consider too difficult (which could lead to a sense of frustration), while more advanced learners may receive questions that they might consider too easy (which could lead to a sense of boredom). A second limitation is that JavaPAL offers standard definitions for the concepts addressed in the MOOC, although these definitions may not be sufficient for some learners who would also need to review the specific parts of the related videos to see practical examples on the application of these concepts. Nevertheless, to do that, learners would have to leave the conversational agent, go to the MOOC, and search for the appropriate video fragments manually, which can be very time consuming and may even lead to learners not selecting the right content they need at the right time.

In this context, this article presents an adaptive learning module for the JavaPAL conversational agent. This adaptive learning module focuses on addressing the two main limitations identified in JavaPAL. First, this module adapts the multiple-choice questions to be asked to learners by the conversational agent using item response theory (IRT) (Hambleton, Swaminathan, Rogers, 1991), thus providing adaptive content. IRT is a paradigm used to design tests and questionnaires and can model the relationship between an individual's response to a single test item and his or her performance on an overall measure of the ability that this item was intended to measure. In the case of JavaPAL, IRT is used to calculate the relationship between learners' ability and the probability of correctly answering a question, and the relationship between the information provided by a question and the ability of each learner; these two values may be of interest to instructors when redesigning the MOOC and, therefore, are also presented in the form of visualisations aimed at instructors. Second, the adaptive learning module recommends video fragments extracted from the MOOC and mapped with the ontology of concepts addressed in the MOOC, thus providing content recommendation. Video fragments recommendations are provided when the learner fails a question asked by the conversational agent, which then provides the link to the related video fragment so that the learner can watch this video fragment to review the concept being asked at the very moment (without having to go to the MOOC, search for the video and watch it entirely to locate the necessary information). In this way, the specific content the learners need at a given time is integrated within the interactions with the conversational agent.

In addition to presenting the adaptive learning module for JavaPAL, this work analyses some implications of including this module in the conversational agent from the perspective of end-users. To this end, three research questions (RQs) are defined. The first RQ (RQ1) is a general question related to the overall operation of the conversational agent with the adaptive learning module and whose objective is to see if the integration of the adaptive learning module presents any problems for end-users (learners in this case) that could limit the use of the conversational agent due to, for example, an excessive complexity or a poor experience. The second and third RQs focus on the specific functionality added in the adaptive learning module: the video fragments recommendations provided to the learners (RQ2), and the adaptation of multiple-choice questions using IRT and its importance from the point of view of the instructor who creates these questions (RQ3). The three RQs are stated as follows.

- RQ1: How is the usability and learnability of the conversational agent with the adaptive learning module and how do they compare to those of the conversational agent without the adaptive learning module?
- RQ2: Are the video fragments recommended in the adaptive learning module of the conversational agent suitable for learners to answer the proposed questions?
- RQ3: Are the visualisations generated by the adaptive learning module as part of the IRT-based adaptation of questions useful for instructors when redesigning the MOOC?

Literature review

Conversational agents for education

Conversational agents (whether text-based or voice-based) have increased their presence in our daily lives, with Amazon Alexa and Google Assistant as two of the most popular voice services that are integrated into devices (e.g., smart speakers such as Amazon Echo or Google Home), and upon which specific-purpose applications are built. The human-like interaction, the multi-tasking possibility, and the capability of using the voice to access a large amount of information are some of the features that make the use of conversational agents on the rise at this time (McLean & Osei-Frimpong, 2019). Nevertheless, the communication with machines using natural language is not new at all, with pioneering work in the field such as ELIZA (back in 1964), a chatbot that received inputs and generated outputs based on predefined built-in rules (Weizenbaum, 1966), or more recently, ALICE, a chatbot that was able to generate outputs matching inputs using internal templates (AbuShawar & Atwell, 2015).

In the field of education, the evolution of artificial intelligence and machine learning (ML) techniques has allowed the creation of powerful conversational agents to help both teachers and students (Winkler & Söllner, 2018). For example, in the case of teachers, ProblemPal (Trivedi, 2018) is an Amazon Alexa skill that helps in the automatic generation of practice content. Teachers say the types of exercises they want and ProblemPal uses ML to generate these exercises and share them with students through Google Classroom. In addition, LTKA-Bot (Mulyana & Hakimi, 2018) helps teachers in administrative tasks, such as task assignment, group management, and scoring management, among others. In the case of students, Oscar (Latham et al., 2010) is a conversational agent which mimics the behaviour of a normal tutor and self-adjusts to the learning style of the learner. In addition, Scarlet (Ilhan et al., 2017) is an artificial teacher assistant that uses natural language processing to provide learners with the materials they need to learn at a given moment.

The particular case of MOOCs combines the challenges of online education with the limited support from instructors, so MOOCs represent an interesting case in which conversational agents can be useful, especially for learners. However, this technology has not yet been widely explored in this context (Caballé & Conesa, 2018). QuickHelper (Howley et al., 2015) was one of the first MOOC-related conversational agents and was aimed at increasing the number of questions in the MOOC forum by making learners lose their fear of asking questions. Bazaar (Tomar, Sankaranarayanan, & Rosé, 2016) was another conversational agent aimed at facilitating group work in MOOCs. The conversational agent from the colMOOC project (Demetriadis, et al. 2018) was aimed at triggering learners' constructive dialogue by posing challenging questions. All of these conversational agents were text-based and focused on addressing collaboration and group work within MOOCs.

Finally, JavaPAL (Delgado Kloos et al., 2018; Delgado Kloos et al., 2019) is a voice-based conversational agent designed to support learners in reviewing, studying, and practicing the contents of a MOOC individually. The limitations found about this conversational agent in previous studies lead to the need to develop an adaptive learning module with a twofold functionality: (1) to provide adaptive content (concerning the questions drawn from the MOOC and asked to learners); and (2) to provide content recommendation (concerning video fragments drawn from the MOOC and provided to learners when failing questions).

Adaptive learning

Adaptive learning is a method used to automatically provide learning resources and activities customised to each student's needs (Brusilovsky, Specht, & Weber, 1995). In the particular case of MOOCs, the educational resources primarily used are videos (video lectures), while activities typically refer to (formative and/or summative) assessments, which in most cases include closed-ended questions, and in some cases also include open-ended questions to be peer (and sometimes automatically) assessed. The recommendation of learning resources may improve learners' motivation because with the wealth of information available today, it is sometimes complicated for learners to find the right learning resources useful to them at the right time (Thyagarajan & Nayak, 2007). In the particular case of MOOCs, learners that have to spend too much time looking for the appropriate resources to meet their needs may become unmotivated and drop the course (Onah, Sinclair, & Boyatt, 2014). When recommending learning

resources, an ontology with the key concepts of the course can be defined to map recommended resources and concepts. For example, Shishehchi, Banihashem, and Zin, (2010) proposed a recommender system that uses an ontology to provide students with content suitable for their interests. Regarding video recommendations in MOOCs, SeqSense (Bhatt, Cooper, & Zhao, 2018) is a recommender system that allows learners to access videos from MOOCs that are interesting to them. Another example is MOOCex (Zhao et al., 2018), which recommends videos from different MOOCs according to the learner's preferences. However, and to the best of the authors' knowledge, no research work has been published regarding recommendations of video fragments in MOOCs, nor videos recommended by conversational agents.

The adaptation of activities, especially for closed-ended questions, has been discussed to some extent in the literature (Chrysafiadi, Troussas, & Virvou, 2018). In the case of closed-ended questions, such as those used in most MOOCs, IRT can be an interesting solution to adapt the questions to the level and skills of the students (Cui et al., 2019; Hambleton, Swaminathan, Rogers, 1991; Mahmud, 2017). IRT models can include up to three parameters (Reeve & Fayers, 2005): (1) difficulty (the ability a student needs to have to be able to answer correctly a question with a 50% probability); (2) discrimination (how good a question is to differentiate between students with an ability higher or lower than the needed to answer that question correctly); and (3) guessing (the probability to answer a question correctly by guessing). These three parameters can be used to estimate the user's ability and, therefore, provide students with materials adapted to their abilities. IRT has already been used in educational contexts to adapt the questions asked to the learners, for example with parametric exercises (Muñoz-Merino, Novillo, Delgado Kloos, 2018). In the case of MOOCs, IRT has been used to avoid cheating on graded tests (Alexandron et al., 2016; Meyer & Zhu, 2013), or to improve peer assignment in peer assessment activities (Uto, Nguyen, & Ueno, 2019; Uto, & Ueno, 2018). However, and to the best of the authors' knowledge, no research work has been published regarding the adaptation of questions through IRT in conversational agents.

Visual analytics

The large amount of data collected from learners' performance in learning platforms (e.g., interactions with videos and activities) can be of great interest to its instructors to redesign the course (macro-level) and improve the quality of each educational resource individually (micro-level). Nevertheless, most instructors lack the necessary skills to process the data collected and, therefore, need tools that provide the necessary visualisations to allow them to reach conclusions and make decisions; what is called visual analytics (Vieira, Parsons, & Byrd, 2018). Learning dashboards are typically used to show data to instructors (Schwendimann et al., 2016), although it is important to involve the end-users (instructors in this case) in the process of designing the learning dashboard to provide more meaningful and accurate visualisations of the collected information. For example, Gutiérrez et al. (2020) developed LADA, a learning dashboard to support academic advisors in the decision-making process using comparative and predictive analysis.

In the case of MOOCs, some learning dashboards have already been developed aimed at instructors to redesign and improve the quality of their courses. For example, Coffrin et al. (2014) developed visualisations of learners' engagement in a MOOC using participation, grades, and interaction. Fu, Zhao, Cui, and Qu (2016) developed iForum, a tool to show visualisations related to the use of the forum in MOOCs (e.g., most active posts, the social connection between learners, etc.). Similarly, Moreno-Marcos et al. (2018) developed LATÉS, a tool to show visualisations related to the use of the forum in MOOCs from three perspectives: social, sentiments, and skills. Nevertheless, and to the best of the authors' knowledge, no research work has been published regarding learning dashboards that provide an IRT analysis of the activities in a MOOC, even though the interpretation of such an analysis may be challenging and not always easy to understand by instructors.

The evaluation of learning dashboards is a key element to assess different aspects. There are several criteria to be considered in the evaluation of learning dashboards, such as usability, usefulness, understanding, usage, agreement, or impact at different levels (Jivet et al., 2018). Many evaluations of learning dashboards have assessed their usability using the system usability scale (SUS) (Brooke, 1996), such as Santos et al. (2013), or their interpretability, by providing a real or fictitious learning situation in which stakeholders must interpret the specific visualisations and the results are compared with the correct ones, for example, Sedrakyan et al. (2017). This way it is possible to compare if the proposed cases were interpreted correctly by the end-users through the learning dashboards.

JavaPAL conversational agent

JavaPAL is a conversational agent that complements a MOOC on Java Programming offered in edX (Catalán Aguirre et al., 2021). The non-adaptive version of JavaPAL included two main operation modes (Delgado Kloos et al., 2019): Quiz mode, to provide learners with random multiple-choice questions extracted from the MOOC; and Review mode, to provide learners with standard definitions of key Java concepts on demand.

Adaptive learning module

The adaptive learning module aims to overcome the main limitations found in JavaPAL, providing adaptive content (an adaptation of questions in the Quiz mode through IRT) and content recommendation (recommended video fragments in the Review mode). Regarding the adaptation of questions, a two-parameter model of IRT was used, in which two parameters (difficulty and discrimination) were calculated based on previous data, while the third parameter (guessing) received a fixed value depending on the number of possible answers for each question. The decision to assign a fixed value to the third parameter for each question was because it can be initially estimated and due to the limitation in the amount of data available, since calculating the value for the third parameter in each question requires a larger amount of data for the training. In contrast, the difficulty and discrimination values were calculated with the following procedure.

- Step 1: The difficulty and discrimination values for each of the closed-ended questions in the MOOC were calculated using the data collected from the first edition of this MOOC (2015). The difficulty of the questions was obtained using R language code based on the answers provided by learners to each question. Questions were annotated with the difficulty and discrimination values using resource description framework (RDF) and were connected to JavaPAL through a REST (REpresentational State Transfer) application programming interface (API). It is important to note that in some cases, and due to the design of the MOOC, the questions were presented to the learner as sets of related questions in the MOOC; in such cases, the difficulty and discrimination values were calculated for the sets of questions and not for individual questions.
- Step 2: Learners' ability is calculated as learners use JavaPAL based on their answers to the questions provided and taking into account the previously calculated values of difficulty and discrimination. JavaPAL gives the learner a set of questions, which are then answered correctly or incorrectly. The next set of questions the learner receives will be easier or more difficult depending on his/her previous outcomes, which serve to recalculate the ability of that learner.
- Step 3: Visualisations are generated to show the difficulty and the information that each question provides about the users. These visualisations are intended for the instructors of the MOOC (not for the learners) as an aid to course redesign, for example, for detecting (and fixing) problems in some exercises.

Regarding video fragments recommendations, it is important to note that all the recommended video fragments belong to videos available from the MOOC JavaPAL complements. The MOOC contains 97 videos which can also be accessed through their YouTube links (unlisted videos). These videos are distributed in 5 weeks and explain the main concepts with examples. A video may explain several concepts or provide several examples and hence the importance of recommending the right video fragment at the right moment. Each video has been manually annotated in RDF with the following procedure.

- Step 1: Comprehensive visualisation of the videos. Each video was entirely watched to get a general idea of its structure and to determine the possible fragments into which the video could be divided.
- Step 2: Division of the video into topics. Videos were composed of several parts (e.g., "introduction", "definition of concept X"), so the different parts (fragments) were identified, and the start time and end time of each fragment were collected (409 fragments identified).
- Step 3. Selection of useful fragments for JavaPAL. The decision was made considering the parts in which the videos were divided. For example, for a video divided into "introduction", "definition of concept X", and "conclusion" only the fragment "definition of concept X" was selected (95 video fragments selected).

- Step 4. Generation of links for video fragments. YouTube links (URLs) were generated with the start times and end times collected in the second step for each video fragment.
- Step 5. Mapping of video fragments to Java concepts. The Java concepts were arranged according to an ontology (Delgado Kloos et al., 2019) that was created to provide learners with the standard definitions of the concepts included in this ontology.
- Step 6. RDF annotations. The relationships between the concepts and video fragments were annotated using RDF and were connected to JavaPAL through a REST API. The REST API supports queries to access the information needed to recommend the video fragments within JavaPAL depending on the concepts to be reviewed.

In addition, other related features added to JavaPAL as a result of including the adaptive learning module are learners can also practice a specific concept in the Review mode by receiving only questions related to that concept; and learners can check in a new mode (called Performance mode), the concepts that they need to review because there are questions related to these concepts that the learner failed to answer.

Architecture

The architecture of JavaPAL is presented in Figure 1. Learners access JavaPAL through Google Assistant using any compatible device (smartphone, tablet, Google Home, etc.). JavaPAL is integrated into Dialogflow, a natural language understanding platform by Google, that understands the learners' inputs (either text-based or voice-based). Therefore, when the learner types or talks to JavaPAL (e.g., "I want to play the Quiz mode"), Dialogflow understands this message and translates it to a user intention that has been previously defined in JavaPAL. Then, JavaPAL interprets this intention and, using Node.js code, displays the appropriate response to the learner. In addition, Dialogflow is connected to a database called Firebase (Firebase Realtime Database), a NoSQL cloud-hosted database recommended by Google for the development of conversational agents; this database stores the relevant information, such as the questions and their answers, and learners' performance, among others. Dialogflow accesses the information stored in Firebase using Node.js code. In addition, RDF files including annotations with some relevant relationships (relationships among concepts, and relationships between concepts and questions) are outside JavaPAL and can be accessed through a REST API. This REST API is written in Java and responds to queries aimed at obtaining the required information. Therefore, Dialogflow makes an HTTP request to the REST API asking for the information that the actions of the user inside JavaPAL require.

The adaptive learning module adds some new elements to the architecture. More specifically, it adds RDF files containing annotations with the relationships between the questions and their difficulty and discrimination values, as well as between the Java concepts and the video fragments. These RDF files can also be accessed through a REST API written in Java so that DialogFlow makes an HTTP request to this REST API to collect the needed information. An additional REST API, this time written in Node.js, is also added to access the user estimated ability calculated through an R file outside JavaPAL. Finally, Firebase is also used to store the new needed information, such as the calculated ability, the difficulty of the questions, or the concepts related to the questions that users fail in the Quiz mode.

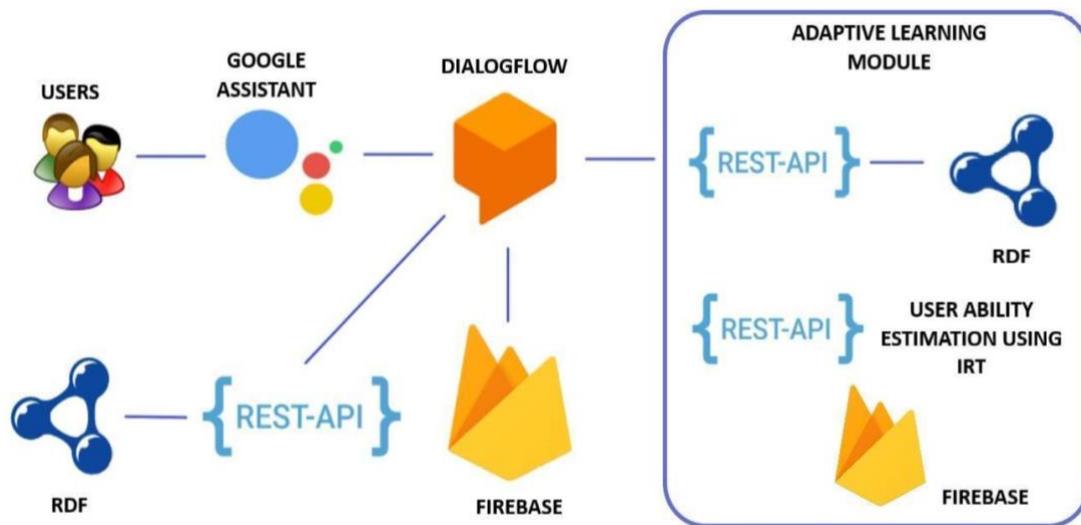


Figure 1. The architecture of JavaPAL

Methodology

The adaptive learning module in JavaPAL has been evaluated through an expert evaluation, which is a method frequently used for the formative evaluation of novel technologies, especially with regard to usability aspects of a system (Landauer, 1997; Triantafyllou, Pomportsis, & Demetriadis, 2003). An expert evaluation is easier to arrange, can be conducted in a more controlled environment, and has shown to be useful to spot problems and recommend changes in novel technologies before handing them over to end-users. Expert evaluations have previously been used in relation to MOOCs, for example, to assess the accessibility in Coursera MOOCs (Al-Mouh, Al-Khalifa, & Al-Khalifa, 2014), or to develop a MOOC design model (Gayoung et al., 2016). Expert evaluations have also been used in relation to conversational agents, for example, to design and develop a conversational agent to guide group discussion (Tavanapour, Theodorakopoulos, & Bittner 2020) or to evaluate chat experience with smart conversational agents (Chen et al. 2019). In the particular case of the MOOCs, the introduction of complementary external technologies requires a certain maturity level in these to avoid that if they are negatively perceived by learners this could also lead to a negative perception of the MOOC. Therefore, the expert evaluation was chosen for the first evaluation of the adaptive learning module of the conversational agent before releasing it to be used together with the MOOC.

In this case, eight experts reviewed and used JavaPAL with the adaptive learning module following a sequence of actions provided. Then, each expert filled in three questionnaires aimed at collecting information to answer the three research questions of this paper (Appendices A, B, C). The selected experts were proficient in the domain of the MOOC JavaPAL complements (i.e., computer science), had experience with the use of technology in education, and had acted as instructors or assistants in higher education courses. The data collection process complies with the general data protection regulation and received institutional approval to carry out this research activity.

The first questionnaire was aimed at evaluating the usability and learnability of the conversational agent with the adaptive learning module. This questionnaire included the 10 elements of SUS (Brooke, 1996) (a well-known tool for measuring usability) plus some general complementary questions (Appendix A). The usability and learnability of JavaPAL without the adaptive learning module had already been measured (Catalán Aguirre et al., 2021) which enables a comparison between the two versions of the conversational agent.

The second questionnaire was aimed at evaluating the suitability of video fragments recommendations. In the first step, six concepts defined in the ontology of concepts addressed in the MOOC were chosen randomly. For each concept, each expert received a subset of related questions and a subset of video fragments. Then, each expert was asked to assess the suitability of each video fragment to answer the related

questions, with free text to justify the decision made (Appendix B). Each expert assessed a total of 28 video fragments.

The third questionnaire was aimed at evaluating the usefulness for instructors of the visualisations generated by the adaptive learning module on the IRT-based adaptation of questions (Appendix C). Visualisations containing item characteristic curves (ICCs) and item information curves (IICs) were presented to experts. ICCs indicate, for each activity, the probability of correctly answering a question according to learners' ability, while IICs show the information provided by each activity regarding learners' abilities (Reeve & Fayers, 2005). Experts were asked several questions about these two visualisations to see if they could understand them correctly, and especially if they were able to get from them the difficulty and discrimination values for each question as an aid for course redesign. Finally, the experts were also asked for their opinions about the use of these visualisations for course redesign.

The three questionnaires served to obtain data that were both of a quantitative nature (e.g., SUS and rating of video fragments recommendations), and of a qualitative nature (e.g., open questions on the functionalities of the adaptive module for the conversational agent and on the application of ICCs and IICs). In the case of data of a quantitative nature, the mean, standard deviation, median and quartiles were calculated, and possible outliers were identified. The results obtained were compared with the reference values in the case of SUS, or were interpreted, in the case of the recommendations of video fragments from the MOOC as a support to solve the proposed questions. In the case of data of a qualitative nature, some categories were predefined and comments to open-ended questions were grouped according to these categories, adding more categories as needed based on the answers obtained. In addition, the answers obtained in relation to the application of the ICCs and IICs from the examples given were classified as correct or incorrect, calculating the percentages of correct answers. Overall, qualitative data was used to explore the findings from quantitative data through a mixed methods approach.

Results

RQ1: Usability and learnability

This section presents the results from the information obtained with the first questionnaire answered by the experts (Appendix A). Figure 2 presents the SUS global score, SUS learnability score, and SUS usability score for JavaPAL with the adaptive learning module (Brooke, 2013; Sauro 2011).

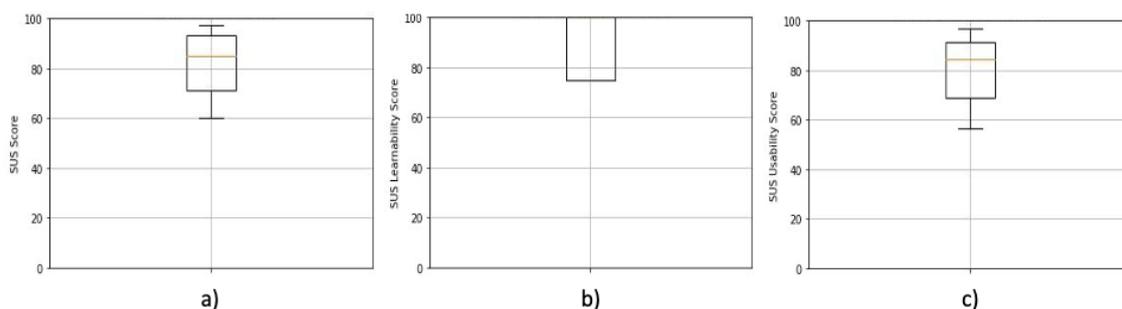


Figure 2. SUS for JavaPAL with the adaptive learning module: a) SUS global score (M 81.88, SD 13.81, median 85); b) SUS learnability score (M 90.63, SD 12.94, median 100.0); c) SUS usability score (M 79.69, SD 15.49, median 84.38).

JavaPAL with the adaptive learning module obtained a mean value for SUS global score of 81.88. This value can be considered “Good” and is matched with the “B” grade (Bangor, Kortum, & Miller, 2009). In addition, the SUS learnability and usability values are also high (mean values 90.63 and 79.69, respectively). Moreover, both learnability and usability are positively skewed, so the results obtained in JavaPAL with the adaptive learning module in terms of learnability and usability can be considered good. It should be noted that, due to the low number of experts filling in the questionnaire, only exploratory results can be drawn from these data.

JavaPAL without the adaptive learning module had been assessed by 39 users and obtained a mean value

for SUS global score of 74.71 (SD 16.48, median 77.5), a mean value for SUS learnability score of 83 (median 87.5), and a mean value for SUS usability score of 72.51 (median 75) (Catalán Aguirre et al., 2021). Therefore, JavaPAL with the adaptive learning module has further improved on the values of usability and learnability obtained by JavaPAL without the adaptive learning module. This may be due precisely to the fact that the adaptation of questions together with the in-place video fragments recommendations leads to an improvement in the overall usability. In any case, it is important to emphasise that the inclusion of more functionality has not worsened the usability and learnability of JavaPAL.

The experts also indicated positive aspects and aspects to be improved in JavaPAL in the first questionnaire. Both the positive aspects and the aspects to be improved have been grouped into related categories from most to least mentioned. Regarding positive aspects, the experts highlighted: adaptation of questions (5 experts); interactivity (4); recommendation of concepts (3); ease of use (3); recommendation of video fragments (2); availability for self-testing (2); and dynamism (1). Among the positive aspects are those related to the adaptive learning module, which was not present in the previous version of the conversational agent; this reinforces the appropriateness of their inclusion. Regarding aspects to be improved, the experts highlighted: to get more visual content (4); to get points after each correct question individually (and not after sets of questions) (3); more variety of questions (3); to be able to interrupt sets of questions (3); to provide further explanations on the score obtained (3); to get more general information on the video fragments recommended (duration and content) (3); to carry out the adaptation using individual questions instead of sets of questions (2). Among the aspects of improvement are those that refer to taking individual questions as a reference instead of sets of questions. While most questions are processed individually in JavaPAL, some others are processed as sets of questions. This is not a limitation of JavaPAL itself as it derives from the design of the MOOC JavaPAL complements, in which some questions were presented as sets of questions, and thus it was not possible to calculate the difficulty and discrimination values individually and JavaPAL had to treat these also as sets of questions.

All in all, it is possible to assert that both the usability and learnability of the conversational agent with the adaptive learning module are good and even better than those obtained without the adaptive learning module. In addition, other aspects of improvement have been detected, especially related to the organisation of the questions, and the information provided to the learner when using JavaPAL with the adaptive learning module.

RQ2: Video fragment recommendations

This section presents the results from the information obtained with the second questionnaire answered by the experts (Appendix B). Each expert rated a total of 28 video fragments recommendations. The video fragments were presented in six rounds to the experts, associated with concepts and related questions. Actually, in each round, each expert received first a group of questions, then the concept or concepts that were asked in the questions, and finally, the video fragments recommended to answer these questions when learners answered them incorrectly. The questionnaire was aimed at assessing whether a learner who receives all recommended video fragments finds them suitable to answer the related questions, regardless of whether one specific video fragment does not fit perfectly with the concept asked in that question. This was done by experts in the field, so they were in a good position to assess the suitability of the video fragments recommendations. Table 1 shows the mean, median, and standard deviation for each round of recommended video fragments based on the experts' opinions. Figure 3 presents a boxplot with the aggregated results for the 28 recommended video fragments.

Table 1
Mean, standard deviation, and median for each round of video fragments recommended

	Number of video fragments recommended	Mean	Standard deviation	Median
Round 1	3	3.42	0.29	3.25
Round 2	4	3.38	0.38	3.38
Round 3	1	4.13	-	4.13
Round 4	5	3.98	0.19	4
Round 5	7	3.85	0.16	3.88
Round 6	8	4.03	0.56	4.13

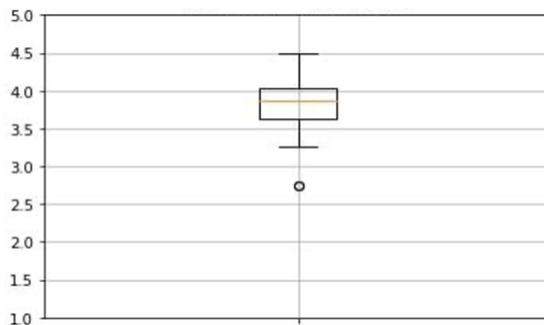


Figure 3. Video fragments recommendation rated by the experts (M 3.8, SD 0.44, median 3.88)

Video fragments recommendations were rated with a mean value of 3.8 on a scale from 1 to 5 (SD 0.44), with an outlier in 2.75. These are positive results, although some values could be improved. Concerning the six rounds, the first and second rounds had the lowest values regarding the recommended video fragments (M 3.42, SD 0.29, and M 3.38, SD 0.38, respectively), while the third and sixth rounds had the highest values (M 4.13 with one video fragment recommend, and M 4.03, SD 0.56, respectively). It is important to consider the effect that the course design had on these results, especially about the relationship between the videos generated for the MOOC, and the questions related to those videos. For example, in the case of the third round, with only one recommended video fragment, the experts considered that this single video fragment was already suitable to solve all the related questions.

The experts also justified their decisions on whether they believed that the video fragments recommended were suitable or not to better answer the questions proposed. Experts' opinions were classified into categories and arranged according to the six rounds. The experts believed that, in general, the video fragments recommended were well selected (in the six rounds), that the video fragments covered more general concepts and not just the questions (in five of the rounds), that the video fragments did not fit perfectly with the questions (in three of the rounds), that the video fragments could be more visual (in three of the rounds), and that all the video fragments were required to fully understand the related concept (in one of the rounds). Overall, experts believed that learners who failed a question could understand their mistake and improve their answers in subsequent opportunities using the recommended video fragments.

All in all, it is possible to assert that the recommended video fragments are reasonably suitable for the learners who have failed some of the proposed questions, although in some cases these video fragments might be more generic than expected by the learner. Aspects to be improved mainly refer to the design of the MOOC and not to JavaPAL itself. This study can also serve to redesign some of the materials. For example, some experts mentioned that some questions were very basic compared to the concept explained in the recommended video fragment. As a result, the concept explained in the video fragment, although related to the question, was much broader. One of the possible options to address this problem is to create more basic or more question-oriented video fragments from scratch. Another possible improvement refers to annotations. For example, annotations could be refined by including more concepts or finer-grain annotations could be proposed to distinguish better among video fragments. Annotations are a very time-consuming task for teachers, and although the automatic processes for making annotations could be improved, it is still necessary to analyse if this can reduce the accuracy of the recommendations.

RQ3: Visualisations of the IRT-based adaptation of questions

This section presents the results from the information obtained with the third questionnaire answered by the experts (Appendix C). First, experts were presented with visualisations of ICCs and IICs for four of the questions extracted from the MOOC and used by JavaPAL. These four questions were selected because they clearly showed different ranges of difficulties and discrimination values so that it could be possible to find out whether the experts could understand these visualisations or not.

The visualisation of ICCs serves to interpret the concept of difficulty, while the visualisation of IICs serves to interpret the concept of discrimination within IRT. The obtained results from experts' answers indicate 97.5% of correct answers in the case of ICCs and 87.5% of correct answers in the case of IICs. Consequently, it can be stated that the experts understood the concepts of difficulty and discrimination associated with ICCs and IICs. This was reinforced by the expert themselves, who were asked whether they considered they had understood the concepts of difficulty (M 4.34, SD 0.52, median 4, scale from 1 to 5) and discrimination (M 4.13, SD 0.84, median 4, scale from 1 to 5). When interpreting these results, it is important to keep in mind that the experts were proficient in the computer science domain, so they were used to interpreting mathematical functions.

Once it was established that the experts understood the concepts of difficulty and discrimination associated with the proposed visualisations with ICCs and IICs, the next step was to collect experts' opinions on possible uses of these visualisations, more specifically to know what is happening in the course, to design exams, and to redesign the course content. Figure 4 presents the boxplots with the experts' opinions on these three possible uses.

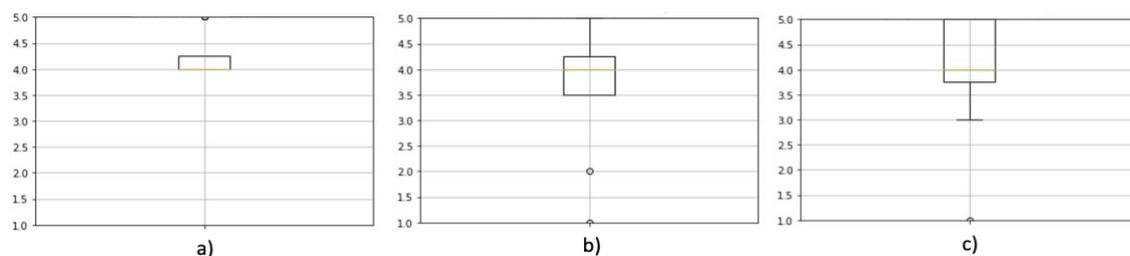


Figure 4. Experts' opinions on the use of ICCs and IICs visualisations: (a) to better understand what is happening in the course (M 4.25, SD 0.46, median 4); (b) to design exams (M 3.63, SD 1.41, median 4); (c) to redesign the course content (M 3.88, SD 1.36, median 4).

The visualisations were considered useful to know what is happening in the course (M 4.24, SD 0.46) (Figure 4a). As part of the justification for their answers, experts felt that the visualisations provided could be useful to better design specific formative questions for learners and to detect when learners are having problems and provide them with additional materials as a way of reinforcement.

The visualisations were also considered useful to design exams (M 3.63, SD 1.41) (Figure 4b), although in this case two outliers were detected, which resulted in polarised opinions. These two experts considered that an instructor should not adapt the difficulty of the exam to the level of the learners and that the difficulty level should be the same for all the learners. In this case, the criticism does not refer to the visualisations themselves, but to the purpose of the visualisations here, which is to adapt the summative assessment activities to the actual level of the learners.

The visualisations were considered also useful to redesign the course content (M 3.88, SD 1.36) (Figure 4c), with one outlier in this case. The expert with the dissenting opinion also considered here that the contents of a course should not be redesigned to be adapted to the level of the learners. Again, the criticism does not refer to the visualisations themselves, but the ultimate purpose of the visualisations.

All in all, it is possible to assert that the experts correctly understood the main concepts related to IRT presented as visualisations (ICCs for difficulty and IICs for discrimination), although it is worth noting the background of the experts in this case. Furthermore, the experts considered that the visualisations were useful to know what is happening in the course. They also considered them useful for designing exams and

redesigning contents, although there are discrepancies with respect to the need to adapt the exams and contents to the level of the learners.

Discussion

The adaptive learning module that was developed for JavaPAL conversational agent has its foundations in the research on intelligent tutoring systems (ITSs) (Kulik & Fletcher, 2016), but adding the possibility of voice-based communication between the learner and the system (agent) (e.g., Hobert & Meyer von Wolff, 2019). Although the use of IRT has been widely analysed in the context of ITSs (e.g., Kavitha, Vijaya, & Saraswathi, 2012; Ma et al., 2014) the novelty of this article lies precisely in the combination of content adaptation (questions) based on IRT and conversational agents, to which is added also the content recommendation (video fragments); all this in a learning environment where such adaptation and recommendation are more relevant as is the case with MOOCs. It is also important to note that most MOOCs are designed to combine educational videos with question activities, so the adaptive learning module presented here covers the two main types of educational content that learners find in MOOCs (Conole, 2015).

Therefore, it can be argued that this work covers a research gap found in the literature and also results in important implications for the practice of MOOC instructors and learners. According to the expert evaluation carried out, the conversational agent with the adaptive learning module showed good usability and learnability (and no negative effect was detected with the addition of the adaptation functionality), which might help in its adoption by MOOC learners, especially considering the importance of enhancing learnability of online education technologies in order to enhance their adoption (Hakami, White, & Chakaveh, 2017). The reasonable suitability of video recommendations according to the expert evaluation carried out is supported by the need for just-in-time teaching to better fit the needs of learners and enhance their learning process (Jonsson, 2015), in the particular case here of failing a question from the MOOC. Finally, the use of IRT to adapt the questions, which is in principle transparent to learners, has demonstrated in the expert evaluation its potential to close the loop and allow instructors to rethink the design of different aspects of the course thanks to IRT-based visualisations (Mattingly, Rice, & Berge, 2012).

Finally, an important lesson learned concerns the synergies that should be established between the MOOC and the conversational agent with the adaptive learning module that complements the MOOC, since it is important to take into account the restrictions that one can impose on the other. If the conversational agent with the adaptive learning module is developed after the MOOC (as has been the case in this study), some difficulties may arise when trying to transform contents from the MOOC to be compatible with the conversational agent. For example, the selection of questions from the MOOC for the content adaptation of the conversational agent was subjected to certain restrictions, such as the selection of closed-ended questions only (and from them, the selection of multiple-choice questions only), the need for shortening some questions that had too many characters in the MOOC, or the need to work with sets of questions when implemented this way in the MOOC instead of with individual questions to calculate the IRT parameters. In the same way, the recommendations of video fragments in the conversational agents were subjected to some restrictions, such as the division of the explanation of a concept into several videos, or the difficulty of mapping questions, concepts, and video fragments. Therefore, if a conversational agent with an adaptive module is going to be developed to complement MOOCs on other topics, it would be advisable to take these restrictions into account in the design phase of the MOOC to facilitate the compatibility of the contents with the conversational agent.

Conclusions

This paper has presented an adaptive learning module for JavaPAL, a conversational agent aimed at supporting learners enrolled in a MOOC on Java programming. This adaptive learning module provides learners with adaptive content and content recommendation. Adaptive content consists of the adaptation of the questions provided to learners using IRT. Content recommendation consists of the recommendation of video fragments from the MOOC depending on the questions failed by the learner. The adaptive learning module of JavaPAL has been evaluated, showing good usability and learnability (better than that of JavaPAL without the adaptive learning module), reasonably suitable video fragments recommendations, and useful visualisations based on the IRT adaptation of questions.

Nevertheless, this study is not without limitations. The evaluation of the adaptive learning module was carried out with eight experts, which were proficient in the domain of a specific MOOC in the area of programming. Further evaluation with end-users, including learners and instructors, is advisable, also with MOOCs from other areas of knowledge. Moreover, it is important to note that JavaPAL and its adaptive learning module were developed after the MOOC was released, which imposes certain constraints to the types of questions that could be adapted and the video fragments that could be recommended. Thus, it would be interesting to create specific content for JavaPAL from scratch, such as video fragments with more examples and definitions of concepts, more related to the specific questions asked by JavaPAL.

Acknowledgements

This work was supported in part by the FEDER/Ministerio de Ciencia, Innovación y Universidades– Agencia Estatal de Investigación, through the Smartlet Project under Grant TIN2017-85179-C3-1-R, and in part by the Madrid Regional Government through the e-Madrid-CM Project under Grant S2018/TCS-4307, a project which is co-funded by the European Structural Funds (FSE and FEDER). Partial support has also been received from the European Commission through Erasmus+ Capacity Building in the Field of Higher Education projects, more specifically through projects LALA, InnovaT and PROF-XXI (586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP), (598758-EPP-1-2018-1-AT-EPPKA2-CBHE-JP), (609767-EPP-1-2019-1-ES-EPPKA2-CBHE-JP). This work has also been supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M21), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation). This publication reflects the views only of the authors and funders cannot be held responsible for any use which may be made of the information contained therein.

References

- AbuShawar, B., & Atwell, E. (2015). ALICE chatbot: Trials and outputs. *Computación y Sistemas*, 19(4), 625-632. <https://doi.org/10.13053/CyS-19-4-2326>
- Alexandron, G., Lee, S., Chen, Z., & Pritchard, D. E. (2016, July). Detecting cheaters in MOOCs using item response theory and learning analytics. *Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (Extended proceedings)* (pp. 1-4). <http://ceur-ws.org/Vol-1618/PALE9.pdf>
- Al-Mouh, N. A., Al-Khalifa, A. S., & Al-Khalifa, H. S. (2014, July). A first look into MOOCs accessibility. *Proceedings of the International Conference on Computers for Handicapped Persons* (pp. 145-152). Springer. https://doi.org/10.1007/978-3-319-08596-8_22
- Atiaja, L. A., & Proenza, R. (2016). The MOOCs: Origin, characterization, principal problems and challenges in higher education. *Journal of e-Learning and Knowledge Society*, 12(1), 65-76. <https://www.learntechlib.org/p/171428/>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123. <https://dl.acm.org/doi/10.5555/2835587.2835589>
- Bhatt, C., Cooper, M., & Zhao, J. (2018, February). SeqSense: Video recommendation using topic sequence mining. *Proceedings of the International Conference on Multimedia Modeling* (pp. 252-263). Springer. https://doi.org/10.1007/978-3-319-73600-6_22
- Brooke, J. (1996). SUS: A 'quick and dirty' usability. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry*, (pp. 189-194). Taylor & Francis.
- Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29-40. http://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Brooke_February_2013.pdf
- Brusilovsky, P., Specht, M., & Weber, G. (1995, September). Towards adaptive learning environments. *Proceedings of the Herausforderungen eines globalen Informationsverbundes für die Informatik (GISI 95)* (pp. 322-329). Springer. https://doi.org/10.1007/978-3-642-79958-7_41
- Caballé, S., & Conesa, J. (2018, September). Conversational agents in support for collaborative learning in MOOCs: An analytical review. *Proceedings of the International Conference on Intelligent Networking and Collaborative Systems* (pp. 384-394). Springer. https://doi.org/10.1007/978-3-319-98557-2_35
- Catalán Aguirre, C., Delgado Kloos, C., Alario-Hoyos, C., & Muñoz-Merino, P. J. (2018, September). Supporting a MOOC through a conversational agent. Design of a first prototype. *Proceedings of the*

- 2018 International Symposium on Computers in Education (pp. 1-6). IEEE.
<https://doi.org/10.1109/SIIE.2018.8586694>
- Catalán Aguirre, C., González-Castro, N., Delgado Kloos, C., Alario-Hoyos, C., & Muñoz-Merino, P. J. (2021). Conversational agent for supporting learners on a MOOC on programming with Java. *Computer Science and Information Systems* (accepted).
- Chen, X., Mi, J., Jia, M., Han, Y., Zhou, M., Wu, T., & Guan, D. (2019). Chat with smart conversational agents: How to evaluate chat experience in smart home. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-6). Association for Computing Machinery. <https://doi.org/10.1145/3338286.3344408>
- Chrysafiadi, K., Troussas, C., & Virvou, M. (2018, October). A framework for creating automated online adaptive tests using multiple-criteria decision analysis. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (pp. 226-231). IEEE.
<https://doi.org/10.1109/SMC.2018.00049>
- Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 83-92). Association for Computing Machinery.
<https://doi.org/10.1145/2567574.2567586>
- Conole, G. (2015). Designing effective MOOCs. *Educational Media International*, 52(4), 239-252.
<https://doi.org/10.1080/09523987.2015.1125989>
- Cui, W., Xue, Z., Shen, J., Sun, G., & Li, J. (2019, September). The Item Response Theory Model for an AI-based Adaptive Learning System. *Proceedings of the 18th International Conference on Information Technology Based Higher Education and Training* (pp. 1-6).
<https://doi.org/10.1109/ITHET46829.2019.8937383>
- Delgado Kloos, C., Alario-Hoyos, C., Muñoz-Merino, P. J., Catalán Aguirre, C., & González-Castro, N. (2019, April). Principles for the design of an educational voice assistant for learning Java. *Proceedings of the International Conference on Sustainable ICT, Education, and Learning* (pp. 99-106). https://doi.org/10.1007/978-3-030-28764-1_12
- Delgado Kloos, C., Catalán, C., Muñoz-Merino, P. J., & Alario-Hoyos, C. (2018, September). Design of a conversational agent as an educational tool. *Proceedings of the Learning with MOOCs* (pp. 27-30). IEEE. <https://doi.org/10.1109/LWMOOCs.2018.8534591>
- Demetriadis, S., Karakostas, A., Tsiatsos, T., Caballé, S., Dimitriadis, Y., Weinberger, A., Papadopoulos, P. M., Palaigeorgiou, G., Tsimpanis, C., & Hodges, M. (2018, March). Towards integrating conversational agents and learning analytics in MOOCs. *Proceedings of the International Conference on Emerging Networking, Data & Web Technologies* (pp. 1061-1072). Springer.
https://doi.org/10.1007/978-3-319-75928-9_98
- Fu, S., Zhao, J., Cui, W., & Qu, H. (2016). Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 201-210.
<https://doi.org/10.1109/TVCG.2016.2598444>
- Gayoung, L. E. E., Sunyoung, K. E. U. M., Myungsun, K. I. M., Yoomi, C. H. O. I., & Ilju, R. H. A. (2016). A study on the development of a MOOC design model. *Educational technology international*, 17(1), 1-37. http://kset.or.kr/eti_ojs/index.php/instruction/article/view/69
- Gutiérrez, F., Seipp, K., Ochoa, X., Chiluiza, K., De Laet, T., & Verbert, K. (2020). LADA: A learning analytics dashboard for academic advising. *Computers in Human Behavior*, 107, 105826.
<https://doi.org/10.1016/j.chb.2018.12.004>
- Hakami, N., White, S., & Chakaveh, S. (2017, April). Motivational factors that influence the use of MOOCs: Learners' perspective. *Proceedings of the 9th International Conference on Computer Supported Education* (pp. 323-331). <https://www.scitepress.org/papers/2017/62595/62595.pdf>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Hobert, S., & Meyer von Wolff, R. (2019, February). Say hello to your new automated tutor—a structured literature review on pedagogical conversational agents. *Proceedings of the Wirtschaftsinformatik* (pp. 301-314). <https://aisel.aisnet.org/wi2019/track04/papers/2/>
- Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, 98, 157-168. <https://doi.org/10.1016/j.compedu.2016.03.016>
- Howley, I., Tomar, G., Yang, D., Ferschke, O., & Rosé, C. P. (2015, June). Alleviating the negative effect of up and downvoting on help seeking in MOOC discussion forums. *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 629-632). Springer.
https://doi.org/10.1007/978-3-319-19773-9_78

- Ilhan, K., Mušić, D., Junuz, E., & Mirza, S. (2017, May). Scarlet-Artificial teaching assistant. *Proceedings of the International Conference on Control, Artificial Intelligence, Robotics & Optimization* (pp. 11-14). IEEE. <https://doi.org/10.1109/ICCAIRO.2017.11>
- Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 31-40). <https://doi.org/10.1145/3170358.3170421>
- Jonsson, H. (2015, October). Using flipped classroom, peer discussion, and just-in-time teaching to increase learning in a programming course. *Proceedings of the IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. <https://doi.org/10.1109/FIE.2015.7344221>
- Jung, E., Kim, D., Yoon, M., Park, S., & Oakley, B. (2019). The influence of instructional design on learner control, sense of achievement, and perceived effectiveness in a supersize MOOC course. *Computers & Education*, 128, 377-388. <https://doi.org/10.1016/j.compedu.2018.10.001>
- Kavitha, R., Vijaya, A., & Saraswathi, D. (2012, January). Intelligent item assigning for classified learners in ITS using item response theory and point biserial correlation. *Proceedings of the International Conference on Computer Communication and Informatics* (pp. 1-5). IEEE. <https://doi.org/10.1109/ICCCI.2012.6158813>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42-78. <https://doi.org/10.3102/0034654315581420>
- Landauer, T. K. (1997). Behavioral research methods in human-computer interaction. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 203-227). North-Holland. <https://doi.org/10.1016/B978-044481862-1.50075-3>
- Latham, A. M., Crockett, K. A., McLean, D. A., Edmonds, B., & O'Shea, K. (2010, July). Oscar: An intelligent conversational agent tutor to estimate learning styles. *Proceedings of the International Conference on Fuzzy Systems* (pp. 1-8). IEEE. <https://doi.org/10.1109/FUZZY.2010.5584064>
- Littlejohn, A., & Hood, N. (2018). *Reconceptualising learning in the digital age: The [un] democratising potential of MOOCs*. Springer. <https://www.springer.com/gp/book/9789811088926>
- Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, 29, 40-48. <https://doi.org/10.1016/j.iheduc.2015.12.003>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901-918. <https://psycnet.apa.org/fulltext/2014-25074-001.html>
- Mahmud, J. (2017). Item response theory: a basic concept. *Educational Research and Reviews*, 12(5), 258-266. <https://doi.org/10.5897/ERR2017.3147>
- Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal*, 4(3), 236-247. <https://doi.org/10.34105/j.kmel.2012.04.020>
- Meyer, J. P., & Zhu, S. (2013). Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating. *Research & Practice in Assessment*, 8, 26-39. <https://eric.ed.gov/?id=EJ1062822>
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, 99, 28-37. <https://doi.org/10.1016/j.chb.2019.05.009>
- Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., Estevez-Ayres, I., & Delgado Kloos, C. (2018). A learning analytics methodology for understanding social interactions in MOOCs. *IEEE Transactions on Learning Technologies*, 12(4), 442-455. <https://doi.org/10.1109/TLT.2018.2883419>
- Mulyana, E., & Hakimi, R. (2018, July). Bringing automation to the classroom: A chatOps-based approach. *Proceedings of the 4th International Conference on Wireless and Telematics* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICWT.2018.8527810>
- Muñoz-Merino, P. J., Novillo, R. G., & Delgado Kloos, C. (2018). Assessment of skills and adaptive learning for parametric exercises combining knowledge spaces and item response theory. *Applied Soft Computing*, 68, 110-124. <https://doi.org/10.1016/j.asoc.2018.03.045>
- Onah, D. F., Sinclair, J., & Boyatt, R. (2014, July). Dropout rates of massive open online courses: Behavioural patterns. *Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14)*, (pp. 5825-5834). <http://wrap.warwick.ac.uk/65543/>
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. M. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (pp. 55-73). Oxford University Press.

- Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013). Addressing learner issues with StepUp! an evaluation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 14-22). Association for Computing Machinery. <https://doi.org/10.1145/2460296.2460301>
- Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC.
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41. <https://doi.org/10.1109/TLT.2016.2599522>
- Sedrakyan, G., Leony, D., Muñoz-Merino, P. J., Delgado Kloos, C., & Verbert, K. (2017, September). Evaluating student-facing learning dashboards of affective states. *Proceedings of the European Conference on Technology Enhanced Learning* (pp. 224-237). Springer. https://doi.org/10.1007/978-3-319-66610-5_17
- Shishehchi, S., Banihashem, S. Y., & Zin, N. A. M. (2010, June). A proposed semantic recommendation system for e-learning: A rule and ontology based e-learning recommendation system. *Proceedings of the 2010 International Symposium on Information Technology* (pp. 1-5). IEEE. <https://doi.org/10.1109/ITSIM.2010.5561329>
- Tavanapour, N., Theodorakopoulos, D., & Bittner, E. A. (2020, July). A conversational agent as facilitator: Guiding groups through collaboration processes. *International Conference on Human-Computer Interaction* (pp. 108-129). Springer. https://doi.org/10.1007/978-3-030-50506-6_9
- Thyagarajan, K. K., & Nayak, R. (2007). Adaptive content creation for personalized e-learning using web services. *Journal of Applied Sciences Research*, 3(9), 828-836. <https://doi.org/10.1109/IAMA.2009.5228081>
- Tomar, G. S., Sankaranarayanan, S., & Rosé, C. P. (2016, March). Intelligent conversational agents as facilitators and coordinators for group work in distributed learning environments (MOOCs). *Proceedings of the 2016 Association for the Advancement of Artificial Intelligence Spring Symposium Series*.
- Triantafillou, E., Pomportsis, A., & Demetriadis, S. (2003). The design and the formative evaluation of an adaptive educational system based on cognitive styles. *Computers & Education*, 41(1), 87-103. [https://doi.org/10.1016/S0360-1315\(03\)00031-9](https://doi.org/10.1016/S0360-1315(03)00031-9)
- Trivedi, N. (2018, October). ProblemPal: Generating autonomous practice content in real-time with voice commands and Amazon Alexa. *Proceedings of the E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 80-82). Association for the Advancement of Computing in Education. <https://www.learntechlib.org/p/184950/>
- Uto, M., Nguyen, D. T., & Ueno, M. (2019). Group optimization to maximize peer assessment accuracy using item response theory and integer programming. *IEEE Transactions on Learning Technologies*, 13(1), 91-106. <https://doi.org/10.1109/TLT.2019.2896966>
- Uto, M., & Ueno, M. (2018, June). Item response theory without restriction of equal interval scale for rater's score. *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 363-368). Springer. https://doi.org/10.1007/978-3-319-93846-2_68
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122, 119-135. <https://doi.org/10.1016/j.compedu.2018.03.018>
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. <https://doi.org/10.1145/365153.365168>
- Winkler, R. & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Proceedings of the Academy of Management Annual Meeting*, Chicago, IL. (pp. 1-40), <https://www.alexandria.unisg.ch/254848/>
- Zhao, J., Bhatt, C., Cooper, M., & Shamma, D. A. (2018). Flexible learning with semantic visual exploration and sequence-based recommendation of MOOC videos. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, (pp. 1-13). <https://doi.org/10.1145/3173574.3173903>

Corresponding author: Carlos Alario-Hoyos, calario@it.uc3m.es

Copyright: Articles published in the Australasian Journal of Educational Technology (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: González-Castro, N., Muñoz-Merino, P. J., Alario-Hoyos, C., & Delgado Kloos, C. (2021). Adaptive learning module for a conversational agent to support MOOC learners. *Australasian Journal of Educational Technology*, 37(2), 24-44. <https://doi.org/10.14742/ajet.6646>

Appendix A

This appendix presents the first questionnaire used to collect information. This questionnaire contains the ten statements of the SUS standard version, plus some complementary general questions on the adaptive learning module of JavaPAL.

Table A1
First questionnaire used to collect information

ID	Sentence	Strongly disagree	1	2	3	4	Strongly agree	5
1	I think that I would like to use this system frequently	1	2	3	4	5		
2	I found the system unnecessarily complex	1	2	3	4	5		
3	I thought the system was easy to use	1	2	3	4	5		
4	I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5		
5	I found the various functions in the system were well integrated	1	2	3	4	5		
6	I thought there was too much inconsistency in this system	1	2	3	4	5		
7	I would imagine that most people would learn to use this system very quickly	1	2	3	4	5		
8	I found the system very awkward to use	1	2	3	4	5		
9	I felt very confident using the system	1	2	3	4	5		
10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5		
	The questions you received were adapted correctly depending on your previous questions solved correctly and incorrectly Justify your previous answer (free text)	1	2	3	4	5		
	The recommendations of video fragments were appropriate as an aid to solving the questions correctly Justify your previous answer (free text)	1	2	3	4	5		
	Overall opinion on the app (1 min. value, 5 max. value)	1	2	3	4	5		
	Indicate three positive aspects of the app (free text)							
	Indicate three aspects to be improved in the app (free text)							

Appendix B

This appendix presents the second questionnaire used to collect information. In the first step, six random concepts from the ontology of concepts of this MOOC were chosen. This ontology relates concepts, questions, and video fragments. The chosen concepts were: (1) statement/variable/expression; (2) array; (3) constructor; (4) method; (5) program, and (6) comment. For each concept, the expert was provided with a subset of multiple-choice questions related to that concept with their correct answers. These are questions that the conversational agent may provide to a learner. Then, the expert received a subset of video fragments and was asked to evaluate the suitability of each video fragment for the concept.

Table B1
Second questionnaire used to collect information

Round	Concept	Number of questions	Number of video fragments		Min. value					Max. value
1	Statement/variable/expression	3	3	The video fragment	1	2	3	4	5	
2	Array	2	4	recommended	1	2	3	4	5	
3	Constructor	5	1	is suitable for	1	2	3	4	5	
4	Method	8	5	the concept	1	2	3	4	5	
5	Program	8	7	Justify your	1	2	3	4	5	
6	Comment	12	8	answer (free text)	1	2	3	4	5	
Total		38	28							

Appendix C

This appendix presents the third questionnaire used to collect information. Experts receive a brief description of IRT as well as of difficulty and discrimination parameters. Then, experts receive two figures, one with ICC, and the other one with IIC. For each figure, the expert must answer a set of questions, with some more general questions at the end.

Part 1: Item characteristic curves ICCs)

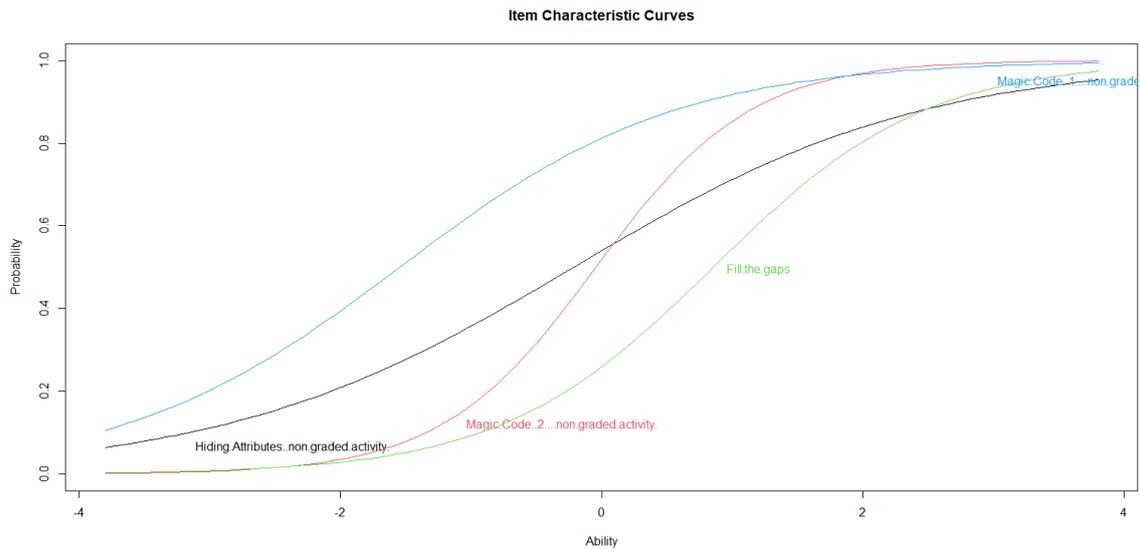


Figure C1. ICCs for four questions taken from the MOOC (blue, black, red, and green lines)

Table C1

Questionnaire related to ICCs

Sentence	Options				
Which question do you think is more complicated?	Blue line	Black line	Red line	Green line	I do not know
Which question do you think is easier?	Blue line	Black line	Red line	Green line	I do not know
Which questions do you think a learner with ability equals to 0 could answer better? (multiple options can be marked)	Blue line	Black line	Red line	Green line	I do not know
Which question would you give to a learner with a low ability (less than -3)?	Blue line	Black line	Red line	Green line	I do not know
Which question would you give to a learner with high ability (more than 3)?	Blue line	Black line	Red line	Green line	I do not know

Part 2: Item information curves (IICs)

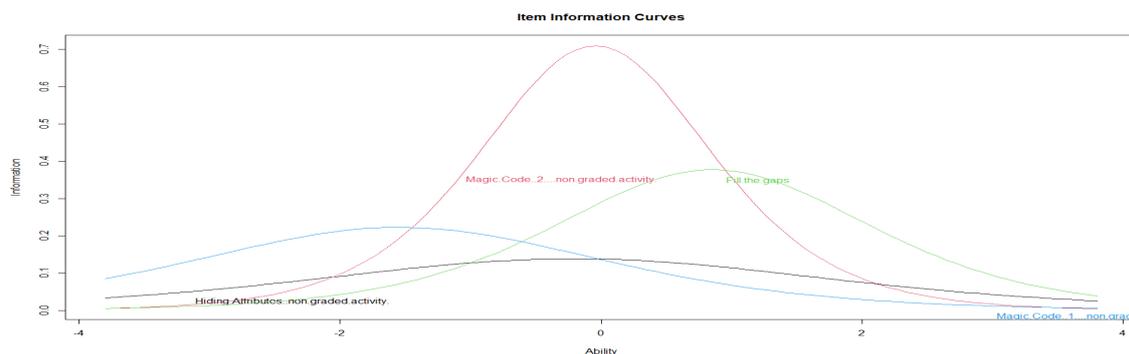


Figure C2. IICs for four questions taken from the MOOC (blue, black, red, and green lines)

Table C2

Questionnaire related to IIC

Sentence	Options				
What do you think this image represents (free text)					
What do you think the term “information” means in the Item Response Theory context? (free text)					
Imagine you have a group of learners with the ability equals to 0. Which of these questions would you give them to differentiate among them? (multiple options can be marked)	Blue line	Black line	Red line	Green line	I do not know
Take a look at the red line. As you see, it provides a higher information rate compared to the other questions. Would you use this question to discriminate between learners with ability higher than 1? If not, which one would you choose?		Yes		No	
Would you use this question to discriminate between learners with an ability lower than -1? If not, which one would you choose?		Yes		No	
If you had to design an exam and the mean ability of your learners would be 0.5, which questions would you pick? (multiple options can be marked)	Blue line	Black line	Red line	Green line	I do not know

Part 3: General questions

Table C3

Questionnaire with additional general questions

Sentence	Strongly disagree					Strongly agree				
These visualisations give me a better understanding of what is happening in my course	1	2	3	4	5					
Justify your answer (free text)										
I find these visualisations useful when it comes to designing my exams	1	2	3	4	5					
Justify your answer (free text)										
I would use these visualisations to redesign the content of my course	1	2	3	4	5					
Justify your answer (free text)										
I understand the concept of difficulty	1	2	3	4	5					
I understand the concept of discrimination	1	2	3	4	5					
Indicate three aspects to be improved in the visualisations	1	2	3	4	5					