Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-17 (2022) 7–12

Towards an open database of assessment material for STEM subjects: requirements and recommendations from early field trials

Ali Mohammadi [***], **Adriano Fagiolini** [*], **Maurizio Cirrincione** [***], **Emanuele Garone** [**], **Anja Garone** [****], and **Damiano Varagnolo** [†]

[*] *Mobile & Intelligent Robots @ Panormous Laboratory (MIRPALab) University of Palermo*

[**] *Université Libre de Bruxelles*

[***] *University of the South Pacific*

[****] *Vrije Universiteit Brussel*

[†] *Norwegian University of Science and Technology*

Abstract: If appropriately implemented, open databases of instruction material may help teaching and learning by providing content for teaching activities, scaffolding, and self-assessment. The paper presents the current results of the development and implementation of a database that is expressly built for promoting exchange of questions and exercises, together with the associated solutions among teachers for STEM subjects. Besides presenting and motivating the initiative (together with reporting its current status), the manuscript lists a series of lessons that have been learned while executing the project – including the need for proper management of authorship and version control of the uploaded material. Moreover, the manuscript describes which features any open database of instruction material should implement to aid improved usability, together with a series of nontrivial theoretical and practical problems for future scientific investigations (e.g., developing taxonomies for indexing the difficulty levels of the instruction material uploaded in the database that do not suffer from the subjective interpretability associated with the existing taxonomies).

Keywords: teaching material repositories; open instruction; collaborative teaching; learning analytics

1. INTRODUCTION

Our anecdotal experience suggests that more or less every teacher in STEM subjects, especially in higher education, suffers from the following problem: finding, inventing and documenting high quality instruction material (such as exercises, questions, and examples to be used either for creating exams or to be given to the students so that they can self-assess).

Every year most of the colleagues around the world spend a non-negligible part of their time on developing new questions and exercises for the purposes above for their courses. This means that every year, all over the world, this amounts to a huge creation of content.

However, most of this content is virtually lost for the rest of the colleagues. In the best possible case, it is uploaded on a variety of more or less reachable local institutions/personal websites. This is an enormous loss of content and time for the various communities. However, if properly collected, indexed, made accessible, and available for effective search, this content could be an invaluable resource for teachers and students alike.

In a sense, in the XX century this problem was generally solved by the presence of some books that collected exercises, statements and solutions, which were published with a certain regularity (typically as companion to theoretical books). For the automatic control community, for example, noticeable publications in this sense are Schoukens et al. [2012], Nahvi and Edminister [2003] (But, every national community has its own famous books, e.g., Marcellini and Sbordone [1995], Haugen [1994]).

Over time, though, and arguably due to the high effort and low economic/career-boosting value of this kind of activity, the publication of this type of book has virtually disappeared, and the old ones are often not easily accessible. Students who are willing to improve and assess their exercise solving capabilities, often cannot access the contents of the syllabus, or have to resort to a time-consuming web-hunt of uncategorized and often unreliable material. Our experience in this activity is futile as students without the highest motivations simply "relax" their attitudes and do just what is suggested by the teachers (potentially, later

⋆ The research leading to these results has received funding from pedagogical funds at Norwegian University of Science and Technology, from the International Federation of Automatic Control, and the European Community through the Erasmus+ project 2019-1-NO01-KA203-060257 "Face It".

10.1016/j.ifacol.2022.09.217

on, complaining that what is provided by the teachers is not enough).

Another observation is the problem of the lack of properly collected, indexed, accessible, and available resources for effective search and self-assessment material for students.It is particularly relevant for students from developing countries that often, due to the lack of resources and local scientific communities, have little (if any) access to expert teachers to get feedback on their progress.

The free availability of a sufficient corpus of high quality student-oriented self-training/self-assessment material may instead mitigate the difficulties of students in such an unfavorable situation. Clearly, this would only be a mitigating factor as the real problem is the lack of a sufficient well-trained teachers. However, from that students' standpoint the more they can access such material the better it would be.

Thus, finding, inventing, and documenting high quality exercises and examples to be used for teaching, assessment, and self-assessment is an issue. Potentially, this may be addressed with a proper and effective way to gather the new questions and exercises that the various teacher come up with their courses every year.

Literature review The paper that seem most related to developing question and answer database is by Pinar-Pérez et al. [2021], which deals with establishing a methodology to construct a university exams database. The authors in Pinar-Pérez et al. [2021] discuss the problems related to creating question and answer database, and their main suggestions are: *1)* the database should be first pioneered by a small group of interdisciplinary scholars, offering experience and testing advantages and challenges. *2)* each teacher from this group should design 20 questions covering their syllabus. *3)* questions should preferably be multiple choice, and to avoid duplications, be sufficiently specific. *4)* the first list of questions should be first peer-reviewed by the creating group. *5)* the reviewed list should then be opened for review to the entire community, then *6)* made available for everybody, students included, *7)* serve the purpose of self-evaluating to international standards. Finally, *8)* it is necessary to maximize meetings to suggest improvements and developments. Pinar-Pérez et al. [2021] provide a series of valuable insights that serve as a basis for the suggestions and experiences we provide in this manuscript.

We then note that the envisioned database offers self-assessment possibilities that can be used for take-home exams. With this regard, Bengtsson [2019] provide a systematic review of the research on take-home exams in tertiary education, showing that different teachers have different opinions on the benefit of take-home exams. Indeed, some believe the preferred choice of assessment method on the higher taxonomy levels, because they promote higher-order thinking skills and allow time for reflection, but some others fear the obvious risks of unethical student behavior (especially when questions are on the lowest taxonomy levels).

We also note Pinar-Pérez et al. [2021] emphasis on the need for maintaining the learning objects in the database. This is a well known problem; see also Vossen and Westerkamp [2004] (that, discuss problems that may arise from storing data on different servers).

Hameed [2016] describes instead how a database as we envision it may be built from the technical, software and engineering-oriented perspective. The authors propose a relational database so that each academic department at the various universities can have their individual database, along with a master database in a Three-Tier architecture combining a web application.

Husemann et al. [2002] looks at web-based platforms, and proposes a setup to enable students to exercise and teachers to implement blended learning. What is more important, the authors describe how web platforms may help implement personalized training concepts while allowing transparent access to third-party systems – a concept that, we believe, it is essential for the early phases of implementation of the envisioned platforms to survive and grow.

Another interesting concept in the literature may complement the database is proposed here. Especially, on the part related to procedural exercises is the one of generating and correcting exercises automatically, as in Prados et al. [2005]. The key concept and requirement for this to work, though, is that of creating base exercises that can be varied according to the topic - something that may be done to a certain extent when considering specific procedural exercises in mathematics, physics and computer science. Such automatic generation may then be cast as an automatic control problem (a formalism that is known to be suitable to design education activities that autonomously adapt to individual students' abilities and learning goals Knorn and Varagnolo [2020]).

Finally, the database proposed here is related to the broader push towards the development of data-driven digital tools approaches and tools for the management of future curricula, e.g., the digital tools that analyze the structure and coherence Teixeira et al. [2020]. The database may indeed help with the conceptualization processes highlighting what students are required to know.

Finally, we cite Telang [2020], who reports some experiences from teachers that, forced by Covid-19 related lockdowns, delivered theoretical contents in dentistry remotely. The authors believe that, it would be feasible to do the same for a large part of theoretical dentistry contents, even if they recognize that face-to-face and online learning experiences have pros and cons. At the same time, they show that the pandemic has brought about innovations to teaching, and conclude by stating that the longer the pandemic continues, the more innovative teaching may become. Our experience with the automatic control community is aligned with the one of these authors.

Statement of contributions Our contributions can be summarized as follows:

(1) we present the initiative and its current status to the public, with the explicit purpose of motivating and attracting potential contributors;

(2) we list a series of lessons that we learned while executing the project;
(3) we define which features the database should implement to improve its usability;
(4) we report a series of nontrivial theoretical and practical problems, which in our opinion require solutions by thorough scientific investigation.

Structure of the manuscript each section in this document answers a specific question. More precisely:

Section 2 answers "*why are we pushing this initiative forward?*";
Section 3 answers "*what has been done until now?*";
Section 4 answers "*what has been learned until now?*";
Section 5 answers "*which features should an ideal database have, and why?*".

Section 6 finally concludes the paper with some remarks and plans for the next future.

2. OUR MOTIVATIONS FOR PUSHING FORWARD AN OPEN DATABASE OF ASSESSMENT MATERIAL FOR STEM SUBJECTS

Our first objective is to facilitate drafting, publishing and keeping updated an extensive collection of exercises with associated (verified) detailed solutions to be available in an 'open source' modality for the whole community. The second, objective, is to learn through our experience what are the effective modalities to create open source exercises and other communities.

The reasons behind this are several:

students can benefit from having additional self-assessment material, that can also be used for scaffolding purposes, since this material provides temporary support, in adjustment with their learning objectives and the tasks assigned, as well as support to accomplish. The absence of costs for obtaining such material is respectful of the economic conditions of the learners – especially for students in underdeveloped countries, as mentioned in the introduction;
teachers can benefit from finding high-quality material with lower efforts than now. Moreover, teachers can also find novel ways of collaborating and novel collaborators by helping improve the exercises as they use them. As described below, indeed, we envision that the database should enable wiki-like editing of the material;
universities , as a whole, can also have benefits. Besides potentially saving costs through the availability of instruction material from other teachers, the material itself may be, at least hypothetically and in a distant future, serve as a standard to define the knowledge levels the learners should reach at the end of the studies.

Finally, we emphasize that there is a strong trend towards open source solutions currently, and both modern & traditional organizations are in the process of developing software with increasingly open-source resources. We forecast this trend will be adopted also in education.

3. WHAT HAS BEEN DONE UNTIL NOW

The main accomplishments until now have been:

- implementing a first version of the publicly accessible repository, currently hosted at
  `https://faceit.pythonanywhere.com/`
  Where teachers can already search, edit, and download questions & exercises, upload their ones, and create tests (i.e., collections of questions & exercises that may be used for learning analytics purposes). These tests can be directly downloaded in a LaTeX format, so as to be immediately importable by teachers in their own slides or handouts (in the assumption that they use LaTeX for their instruction material) [1] ;
- analyzed which type of licensing the material that is being uploaded in the database should follow. We propose CC BY-SA [2] , i.e., the same that is used in the Wikimedia projects (including thus Wikipedia); the reason behind this is that it seems the best compromise in terms of openness & promotion of collaboration vs. possibility of laddering on the material for the creation of commercial contents;
- started filling up the database through a series of exercises and associated solutions that have been provided by colleagues willing to share their material under the licensing above. The database currently consists of around 2000 questions (mainly in English) focusing on basic control courses. In addition, the material has been manually verified, to guarantee its quality.

We have also started making tests about the usability and usefulness of the database with a limited set of colleagues and students. Even if more definitive results will be presented in the future, all the participants have found helpful database, even if the usability may be improved (something that is under development right now). The first online version of the open book is currently hosted at

`www.mirpalab.it/openbook/`

4. LESSONS LEARNED UNTIL NOW

The project started officially at the beginning of 2021, even if prototypes of the main idea and first uncoordinated attempts towards a database of questions have been envisaged by the team since 2019. From these initial phases until now, we have been thus moving: *1)* from local `.tex` files formatted opportunely to a full fledged web repository

[1] From a technical perspective, we note that the database currently uses a standard to define the questions where is purposefully close to the IMS Question & Test Interoperability Specification IMS Global Learning Consortium [2021], Smythe and Roberts [2000], a series of different versions of data structures that provide interoperability between question and test systems in different Learning Management Systems (LMSs) (that in practice, they can be used to download exercises from one LMS as an opportune markdown file, and then reuploaded in another LMS). Currently, we are developing a code to perform this task automatically.

[2] The "BY" part implies that "Licensees may copy, distribute, display, perform and make derivative works and remixes based on it only if they give the author or licensor the credits (attribution) in the manner specified by these". The "SA" part implies that "Licensees may distribute derivative works only under a license not more restrictive than the license that governs the original work".

(with a MongoDB as a kernel and dedicated back and frontends); *2)* from being individuals developing their files to a set of more than 30 persons working on the project; *3)* from having some naïve thoughts about what it means to openly share instruction material to be much more aware of the authors' needs, e.g., granting authorship and guaranteeing quality. In the process, the most important lessons we have likely learned are:

- there is the need to implement Wikipedia-like mechanisms, since questions and exercises can continuously be improved or branched into different versions that comply with different needs of different learners / teachers. Thus, there should also be the possibility for the users to ask the authors of the various items to revise their material;
- there is the need for flags or fields that dynamically represent the estimated verification level of the questions (and solutions, even more importantly). Indeed, since the material is editable online, the following situation occurs (and more frequently than we initially assumed): an author uploads a question, somebody modifies it thinking that this improves the material, the original author also believes this is an improvement, but actually, for the students the first version was cleared. This highlights the need for a rating the mechanism that also accounts for material that may be modified dynamically;
- the material should be indexed accurately, to enable an effective search (generally, the less effective the search, the less helpful the database is). Our experience suggests that the fields that play the most important roles are:
  type of instruction material, i.e., multiple-choice questions vs. open answers. We note that there is a growing need for exercises that require coding and performing simulations, so we expect that some fields like "required level of programming experience" will become important too in the near future;
  content units involved in the material, i.e., which prerequisite the question or exercise requires. We note that it is better to index the material in terms of what the various *solutions* involve, rather than the questions themselves. Moreover, we note the need be to able to query the database inflexible ways, e.g., be able to ask for "material whose solutions use both concepts X and Y, but do not use Z". This flexibility is required to get material that suits to the course;
  difficulty levels of the solutions. As for this point, we note that to the best of our knowledge, standardizing how to describe the difficulty level of what is being asked or taught is an open problem (see also Section 5.1). However, there is the need to associate a "difficulty level" to an exercise (by proxy, though: formulating such difficulty level based on the body of the exercise seems to be an ineffective strategy; it seems more appropriate to associate a difficulty level to each solution of that exercise, since different solutions may be more or less involved, and one way of assessing students is also, through their capability of find different ways of solving a problem).

5. WHAT AN IDEAL DATABASE SHOULD FEATURE, AND WHY

Section 4 describes what has been learned while executing the project. This section instead describes what we believe should be implemented and tested in the next steps.

The first consideration follows from the fact that the frontend of the database enables users to log in, and once they are logged in they may answer the questions directly through their browser. This allows learning analytics to be implemented, i.e., collecting data about which questions the learners answer, how, and when. This information can then be used to understand and optimize the learning processes (e.g., a teacher may push students to answer questions in class about what is being taught at the moment, and use this information on the fly to adapt to the pace of teaching). Teachers may also use historical data about their class to identify the dynamics of learning, and ideally, design education activities that autonomously adapt to individual students' abilities and learning goals, as suggested in Knorn and Varagnolo [2020].

Another essential step that needs to be executed is creating, all together and as a community, a standardized nomenclature to label the knowledge and concepts that are mentioned or laddered in the material uploaded in the database. A simple example showing the need for this is "simple stability" / "marginal stability", the same concept with different names. To enable effective searching in the database, there is need to define how the concepts can be hierarchically nested in terms of prerequisite, e.g., complex numbers as prerequisites for Laplace transforms, plus identify which synonyms are present. Such work would translate into a concepts map of the field which, would, be also a useful instruction material *per se*.

The third step is to enable interoperability of the database with the most common LMSs, i.e., adding interoperability features so that the database can export / import items to / from the LMSs. Our experience says that this feature will enable several teachers to offer their items with minimal effort, as well as facilitate ease for several other teachers using the existing material for their exams and tests. Expanding the set of users is then expected to increment the chances of securing funds and personnel resources to implement the proposed improvements to their database.

5.1 Open questions

We note several features that we believe should be somehow implemented or considered for, but we still do not have clear opinions on how to address them. These are, in our opinion, a set of important open research & development questions:

(1) if a user suggests a significant edit to an existing item in the database (e.g., a way to make a conceptual question more interesting, to the point, and useful for the learner), then how shall the authorship of the item be "blended" with the set of already existing authors? For now, we envision a growing list of authors, where the last user is appointed to the list. However, the importance of the authors in this list dynamically changes in time, and it is not clear how to algorithmically solve this problem in a fair way;

(2) as written above, the indexing of the items should be about the content units they involve. The definition of the conceptual map of such content, units should be unique, and there should be a curated list of synonyms. Who curates that list? And, even more importantly, who should have the authority / physical possibility to execute this curation process? We believe this should be organized by the IFAC Technical Committee on Control Education, but there is the need to encourage the members to be responsible for this duty;
(3) dissemination and promotion activities are essential, since we believe that the database follows Metcalfe's law effects. Thus, for the successfulness of the initiative there may be the need to reach a critical mass as soon as possible. What is the best strategy to disseminate and promote the database?
(4) verifying the quality of the uploaded material (correctness and well-posedness of the exercises and the solutions) is a burden. Who is going to be responsible for it? Which level of "authorship" does this verification entail? And which mechanism design will promote a feeling of responsibility in the users, so that they will improve the quality? We think that the mechanisms that led to the success of Wikipedia may be repeated here. At the same time, there may also be benefits in implementing a rating system for which both the teachers and the students can rate the items they use in the portal. Having good ratings may then stimulate the improvement of material. At the same time, rating systems may introduce distorting effects as in social media, even if these may not become as evident and pervasive in the envisioned database, since they would get just technical information and not opinions;
(5) as written above, to the best of our knowledge, there is no taxonomy for classifying the difficulty levels of the solutions that serve our purposes. Bloom's taxonomy Bloom et al. [1956], the SOLO taxonomy Biggs and Collis [2014], and the other ones we found in the literature (e.g., Fink [2013], Man Choi [2019], Sedelmaier and Landes [2012], Crawley et al. [2011]) do not serve the purpose of describing and quantifying the difficulty levels of exam questions and exercises encountered in engineering-related contexts. Therefore, initiatives need to formally describe and assess the difficulty of learning and teaching material, and to facilitate communications among teachers and students.

6. CONCLUDING REMARKS AND FUTURE PLANS

Every year colleagues across the world spend a non-negligible part of their time developing new instruction material that ends up being virtually unavailable to the whole community, since at best shared among the local networks of the producers. Was this material instead effectively stored and indexed in a computer and accessible through a database, then everybody would benefits.

Aware of this and motivated by the desire to help the community, we have been pushing for such a database to be implemented and deployed through a dedicated project and associated fundraising activities. This has been leading to a web portal that enables collecting and sorting exercises with solutions as above, and that currently contains around 2000 questions related to basic automatic control concepts. This database is thus actually open for being used and filled up with material, even if for now, it has not been publicly advertised - having for now indeed focused on early development and testing so to learn lessons and experience before opening to the public.

The lessons we learned up to now are in fact several: first of all, we believe (now even more than before starting the project) that implementing a genuinely open database of instruction material requires wiki-type features for collaborative editing and updating the material. We also learned that users appreciate keeping track of authorship, and that thus guaranteeing proper attributions to the authors requires implementing such an authorship tracking system. With this comes, however, the currently unsolved problem of updating the relative importance of the authorship of the persons were contributing to each material in time. Still related to this, the possibility of editing induces the need to keep track of the validity and usability of the various items, since edits may affect the precision and correctness of the questions and solutions. Finding a sustainable mechanism to guarantee this seems, however, not an easy task, and to the best of our knowledge, there are no off-the-shelf solutions for this.

We also learned about the importance of ensuring satisfactory user experiences and smooth usability of the database. Our initial tests indicate that all early users appreciate the initiative and the contents inserted until now, but that impaired usability directly translates into dropping off from using the service (both in terms of uploading material and in downloading it).

We then confirmed accurate indexing of the instruction material is essential to improve the overall usability of the service. To develop this indexing, there is the need for a communal effort to define a standard map of the concepts that are taught or used in the discipline, and how they relate to each other (especially in terms of what is a prerequisite, what can be considered as a sub-topic, what is a synonym, etc.). It is unclear who should have to responsible for this and how to implement such mapping in practice, though.

While executing the project, we also realized that a series of open questions need scientific investigations. The most important is likely the one about how to index the difficulty of the material, not only in terms of which knowledge content units it requires, but also in terms of how these should be combined (e.g., in a procedural way or in a conceptual one) to successfully answer the questions. Indeed, measuring in a valid and reproducible way the difficulty levels associated with the various questions would enable implementing much more effective learning analytics, i.e., information about which topics the learners seem to have understood which is, useful to optimize the learning processes in the classroom.

In summary, creating an open database of assessment material for STEM subjects is both socially helpful and technically feasible. Still, it is a process that requires time, effort, and solving a series of other related problems that affect the overall usability of the service.

REFERENCES

Lars Bengtsson. Take-home exams in higher education: a systematic review. *Education Sciences*, 9(4):267, 2019.

John B Biggs and Kevin F Collis. *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press, 2014.

Benjamin S Bloom, Max D Engelhart, Edward J Furst, Walquer H Hill, David R Krathwohl, et al. Taxonomy of educational objetives: the classification of educational goals: handbook i: cognitive domain. Technical report, New York, US: D. Mckay, 1956.

Edward F Crawley, Johan Malmqvist, William A Lucas, and Doris R Brodeur. The cdio syllabus v2. 0: An updated statement of goals for engineering education", in proceedings of the 7 th intl. cdio conference. In *University of Demark, Copenhagen*. Citeseer, 2011.

L Dee Fink. *Creating significant learning experiences: An integrated approach to designing college courses*. John Wiley & Sons, 2013.

Hind Khalid Hameed. Building distributed database system for development of students' exams. *Al-Mustansiriyah Journal of Science*, 27(3), 2016.

Finn Haugen. *Regulering av dynamiske systemer, oppgaver med løsninger*. Tapir, 1994.

B Husemann, J Lechtenborger, Gottfried Vossen, and Peter Westerkamp. Xlx-a platform for graduate-level exercises. In *International Conference on Computers in Education, 2002. Proceedings.*, pages 1262–1266. IEEE, 2002.

IMS Global Learning Consortium. Ims question & test interoperability (qti) specifications, 2021. URL `https://www.imsglobal.org/question/index.html`.

Steffi Knorn and Damiano Varagnolo. Automatic control: the natural approach for a quantitative-based personalized education. *IFAC-PapersOnLine*, 53(2):17326–17331, 2020.

Eun Man Choi. Software engineering education for significant learning experience. *International Journal of Information and Education Technology*, 9(12):862–867, 2019. ISSN 20103689. doi: 10.18178/ijiet.2019.9.12.1318.

Paolo Marcellini and Carlo Sbordone. *Esercitazioni di Matematica*. Liguori Editore, 1995.

Mahmood Nahvi and Joseph A Edminister. Theory and problems of electric circuits, 2003.

Jesús María Pinar-Pérez, María Fernandez-Moya, Pedro Cuadros-Solas, Carlos Salvador, and Roberto Morales-Arsenal. A new methodology to construct a database of world university exams. In *Proceedings INN-ODOCT/20. International Conference on Innovation, Documentation and Education*, pages 89–94. Editorial Universitat Politècnica de València, 2021.

Ferran Prados, Imma Boada, Josep Soler, and Jordi Poch. Automatic generation and correction of technical exercises. In *International conference on engineering and computer education: Icece*, volume 5, 2005.

Johan Schoukens, Rik Pintelon, and Yves Rolain. *Mastering system identification in 100 exercises*. John Wiley & Sons, 2012.

Yvonne Sedelmaier and Dieter Landes. A research agenda for identifying and developing required competencies in software engineering. In *2012 15th International Conference on Interactive Collaborative Learning (ICL)*, pages 1–5. IEEE, 2012.

Colin Smythe and P Roberts. An overview of the ims question & test interoperability specification. 2000.

André MH Teixeira, Aida OPDC Guerra, Steffi Knorn, Kjell Staffas, and Damiano Varagnolo. Computer-aided curriculum analysis and design: existing challenges and open research directions. In *2020 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE, 2020.

Ajay Telang. Covid-19 drives innovation using free online resources. *Journal of Dental Education*, 2020.

Gottfried Vossen and Peter Westerkamp. Maintenance and exchange of learning objects in a web services based e-learning system. *Electronic Journal of e-Learning*, 2(2): pp285–296, 2004.