Check for updates

ESSAY

The possibilities of personalized music education through exercise generation

[version 1; peer review: 1 approved with reservations, 1 not approved]

Filippo Carnovalini[1], Sean Scofield[1], Louis Verstraeten[1], Geraint A. Wiggins[1,2]

[1]Vrije Universiteit Brussel, Brussels, Brussels, 1050, Belgium
[2]Queen Mary University of London, London, England, E1 4NS, UK

v1 **First published:** 13 Apr 2026, **6**:108
https://doi.org/10.12688/openreseurope.23292.1
Latest published: 13 Apr 2026, **6**:108
https://doi.org/10.12688/openreseurope.23292.1

Abstract

One of the more prominently advocated advantages of Artificial Intelligence in the context of Music Education and education in general is its possibility of personalizing the learning experience. We agree with that assessment, but must note that this ideal goal has not been satisfactorily attained in many research efforts. Most personalized systems in literature only rely on learner models that allow teachers to track students' progress and suggest adequate resources. Both the models and the resources are however fixed and finite, limiting the amount of personalization that is possible. We take music education in particular as a field that can greatly benefit from a higher grade of personalization, and consider the possibilities offered by AI generation for music. We posit that with the currently available music generation technologies it should be possible to provide students with exercises created ad hoc for them to address their specific needs, guided by their teacher's indications and concerns. Such a system could obtain a far finer grain of personalization than what is currently available. In this article we review the literature to assess the current standing of musical technique education personalization, and what is possible given the current available technology. We then describe prototypes to show how exercise generation can effectively enhance the personalization of music education, and suggest directions for further research on the topic.

Keywords

AI-Education; Generative AI; Personalization; Music Generation; Music Education; IDyOM

Open Peer Review

Approval Status ? ✗

	1	2
version 1 13 Apr 2026	? view	✗ view

1. **Maria Argyriou**, University of the Aegean, Mytilene, Greece

2. **José Luis Aróstegui**, Universidad de Granada, Granada, Spain

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Horizon Europe gateway.



This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

Corresponding author: Filippo Carnovalini (filippo.carnovalini@vub.be)

Author roles: Carnovalini F: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Scofield S**: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Verstraeten L**: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Wiggins GA**: Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme, under grant agreement No 101108690 (Co-creativity And Learning: Interactive Opus generation for Piano Education [CALIOPE]).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Carnovalini F *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Carnovalini F, Scofield S, Verstraeten L and Wiggins GA. **The possibilities of personalized music education through exercise generation [version 1; peer review: 1 approved with reservations, 1 not approved]** Open Research Europe 2026, **6**:108 https://doi.org/10.12688/openreseurope.23292.1

First published: 13 Apr 2026, **6**:108 https://doi.org/10.12688/openreseurope.23292.1

1. Introduction

The history of music education has seen the use of a variety of technological tools to improve the effectiveness of teaching and, by consequence, the technical ability of musicians.[1] One early "analog" example is the metronome: by having a device that keeps a fixed beat it is much easier to exercise precise playing at fast tempos. Artificial intelligence is also a tool that could provide such pedagogical benefits, and indeed there has been interest on the use of AI for the betterment of music education at least since 1993, the date of the first workshop on the subject.[2] Still, music education seems slower than other disciplines in adapting these novel technologies, and there seems to be less research on AI in music education compared to the use of AI in the teaching of other subjects (e.g., computer science, mathematics, language, and medical education.[3,4] Music education is indeed still largely carried out as it was before the personal computer became a common item in the average household, the most common setting being one-to-one sessions with a teacher, often one hour per week, complemented by solo practice the student does (or should do) to prepare for the following lesson. The study material for both the lessons and the solo practice at the beginner level is typically taken from "method" books, i.e. a collection of exercises and short musical pieces of progressive difficulty that a composer/pedagogist curated as a way to gradually introduce new techniques and challenges to the student. One downside of these books is that they are a one-size-fits-all solution, lacking the ability to adapt to different learner's characteristics, strengths, and shortcomings.

In this article we will substantiate why we believe that, given the current available technology, it should be possible to change this status by embedding AI systems into the educational practice. The advancements in deep learning, and most prominently Transformers, has already been applied to music generation.[5,6] While the results of these systems are currently less astonishing than what is possible for natural language with Large Language Models, or for image generation with Stable Diffusion models, it is safe to predict that the technology will continue to advance to a point where music generated scores will be of similar quality to those composed by an human, at least to an inexperienced listener. We are however not particularly interested in systems that generate music for the sake of generating it, or worse to provide marketing companies with cheaper compositions at the expense of human composers. As scientists, it is our duty to try to direct research towards ethical uses of AI systems, that could have the benefit of the human user at its core. While the direction that the tech world will take is out of our control, this article tries to be a contribution towards that goal. Using music generation systems as a tool to improve music education is, in our opinion, a perfect example of the deployment of AI systems that have a human user and human benefit at its core.

In particular, the benefit we envision as achievable by using AI system within education is the betterment of the personalization of the educational experience. As stated above, most of music education in Western-classical environments is done by using method books and fixed curricula. A (good) teacher could search for additional exercises to give to their students and could even write some new ones in some cases, but this can be extremely expensive in terms of time for the teacher, and doing so for each student could quickly prove overwhelming. Notably, the fact that some teachers do try to go these lengths is a testament to how valuable the generation of new exercises could be seen as a valuable teaching tool. Why is that so? Personalization within an educational context is not merely a flourish added on top of a curriculum or a quirk of some more extroverted teachers. Education research has shown that personalizing the learning approach and material can have a tremendous impact on the quality of education. The difference is so stark that this effect is known as the "two-sigma effect",[7] because it has been shown that a personalized education can lead to results that are two standard deviations away from a control group not using personalized approaches. While we don't expect to be able to obtain such results by the use of AI exercise generation alone, we strongly believe this tool would make it much more realistic for teachers to structure their educational sessions in a personalized manner, which in turn should allow them to achieve this two-sigma improvement.

1.1 Personalization in the zone of proximal develpment

More specifically, we believe these tools could help teacher keep each student in their own Zone of Proximal Development.[8] The idea behind the ZPD is that a student is able to learn most by approaching exercises that are above his current skill level, to allow for improvement, but not so far detached as to frustrate the student because of the hardship of approaching the new exercise (see Figure 1(a)). In other words, it is important to keep a smooth difficulty progression curve, not skipping suddenly to exercises that are too hard, but also not remaining too long on a certain level of difficulty wasting time that could be spent improving on more challenging tasks. It may not appear immediately apparent why the ZPD requires personalization: having a (fixed) set of exercises of increasing difficulty could in theory achieve that smoothness of difficulty required to remain in the right level of challenge. The issue is that difficulty is not a unidimensional scalar value, despite we often think about it in such terms. Because of this, different learners may require different time to dedicate to certain skills, techniques, or facets of interpretations of a musical piece. The ZPD is indeed a zone that can be understood as an area of a multidimensional space, rather than simply a range in a line (as in Figure 1(b)), and that area may be different for different individuals.

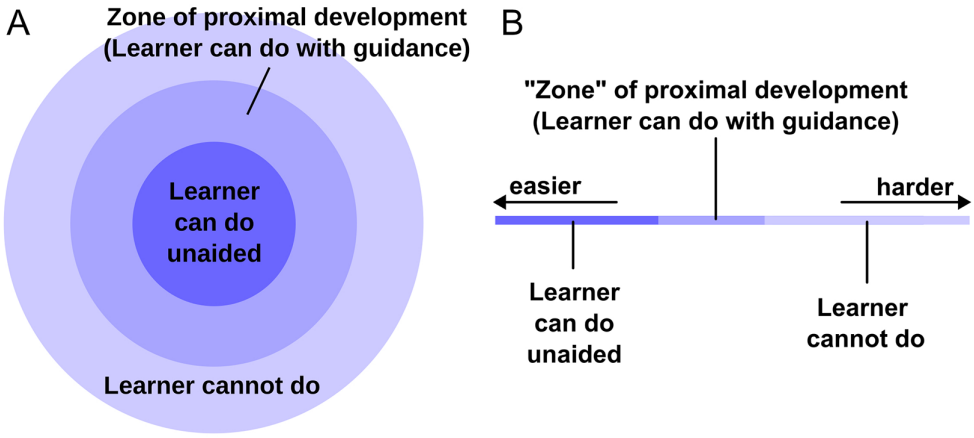


Figure 1. (a) a common representation of the Zone of Proximal Development in a two-dimensional space; (b) what the Zone of Proximal Develpment is reduced to by using only a scalar value of difficulty to evaluate it. Figure (a) has been reproduced under CC0 license from [https://commons.wikimedia.org/wiki/File:Zone_of_proximal_development.svg].

As an example, let us consider the two excerpts of music shown in Figure 2. One is the beginning of an Allegretto by Béla Bartók, the other the beginning of an Ecossaise by Beethoven. ABRSM [1], a well-established music education institution from the UK, included both in exam syllabi for the Grade 3 piano exam, although in different years (2021–2022 for the Ecossaise, 2025–2026 for the Allegretto). We can therefore say they were considered of similar difficulty by expert piano teachers. We can however imagine a piano student who quickly learnt to read notes on a score, accurately estimating the amount of hand movement required for playing large intervals that call for hand position changes, but is maybe less skilled in performing precise rhythms especially if the two hands differ. We can imagine the opposite as well: a student who is good at feeling the rhythm and is good at giving proper accents to notes even if the two hands require different dynamics, but prefers to keep the hand position relatively static. Even if the suggested difficulty is the same for the two pieces mentioned above, the first student will find the Ecossaise easier to perform thanks to their skill at moving hands (as required especially by the left hand part) while the other may prefer the Allegretto, which requires less obvious accenting of off-beat notes but leaves the hands at the same position throughout these bars. The zone of proximal development will be different for the two students, and a good teacher could take this into consideration: while for performing the exam the teacher should suggest playing the piece that comes easier to each student, in order to progress the first student may benefit from focusing on studying the Allegretto and the second may get more benefit from studying the Ecossaise. Or, new exercises could be proposed that show the same rhythmic structures from the Allegretto to the first student, and the hand displacements required for the Ecossaise to the second learner.

2. The state of the art

Having described the way we envision how AI, and more specifically AI based music generation, can be helpful in personalizing music education, we turn to the scientific literature to see how much these ideas have already been explored by researchers.

2.1 Computer-assisted music education

Music Education in a traditional Western context includes a variety of aspects, all of which can benefit from the use of computational systems to some degree. It is possible to divide these aspects into four main categories: fundamentals of music (music theory), performance skills, analysis of music, and composition.[9] In this article, we focus on performance skills, i.e., those skills related to playing one instrument (the piano). Research shows that employing a Computer Assisted Instruction (CAI) framework can effectively increase the effectiveness of piano lessons when compared to a control group.[10] In this framework, the computational system does not substitute the role of the teacher, but rather complements the teaching by offering a learning environment.[11] For example, a CAI system can embed score following to help a student visualize (whether through a display or through augmented reality) what is being played and the mistakes they make.[12,13] Another approach is that of using an Integrated Teaching System (ITS), that is a system that has greater control over the teaching activity by proactively helping the teacher by leveraging knowledge about the subject being taught, the curriculum, and the student. Within this context, there are systems that can select exercises from a given corpus based on

[1]https://www.abrsm.org

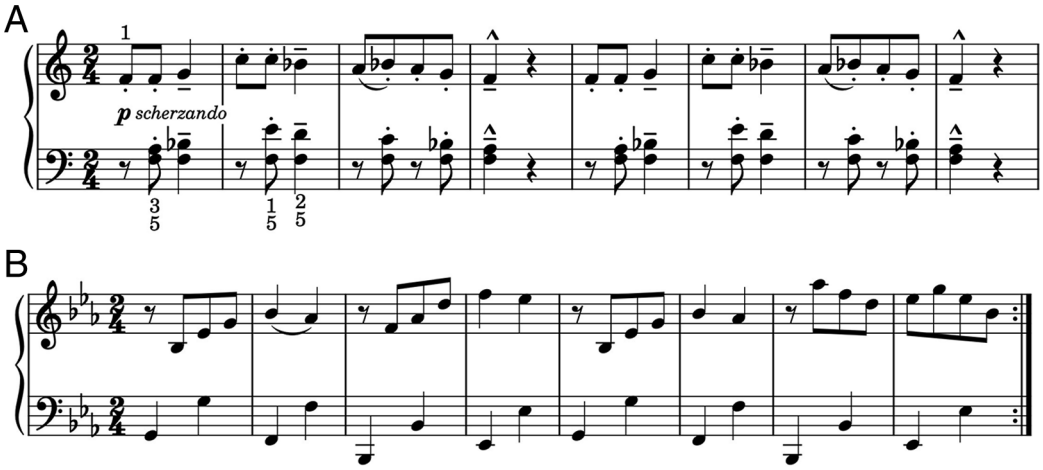


Figure 2. (a) first 8 measures from *Allegretto* by Béla Bartók (no. 3 from the "For Children" album). (b) first 8 measures from *Ecossaise in E flat* by Ludwig van Beethoven (no. 4 from "Six Ecossaises", WoO 83). Both pieces are valued of being of approximately the same difficulty by ABRSM, different students may find one or the other more difficult depending on their specific skills.

the needs of the student and the teacher's concerns.[14,15] To the best of our knowledge, there is no system that generates novel piano exercises to propose to students in a similar fashion. One similar project is the one by Takidaira et al.,[16] that generates rhythmic scores for the popular rhythm game Taiko no Tatsujin to improve the player's ability in the game by analyzing their game logs. While this is not directly meant for musical education, performance in a rhythm game is strongly related to musical skills and thus this can be considered a music education tool. However, that system only focuses on rhythmic aspects, making the system quite limited when compared to the one we propose. Other researchers dealing with music generation have considered applications in the context of learning an instrument, and there is some research towards controlling the difficulty of a generated music piece. Ariga et al.[17] proposed a system to generate guitar solos using fingering as a means of controlling difficulty. Nakamura and Yoshii[18] describe a system that generated piano reduction of ensemble scores, allowing different difficulty settings based on tempo and fingering, and similar techniques have been applied to score difficulty analysis Ramoneda et al.[19,20] However, none of these systems try to actually build a system that uses generation of music for the personalization of piano education.

2.2 Commercial systems

While commercial smartphone applications are often of little scientific interest, it is relevant here to discuss what the industry proposes to the end users in terms of systems meant to aid the learning of piano. A variety of applications exist for this goal, some of which have amassed notable user bases through the years. This section does not aim to be a comprehensive review of all those systems, but some will be taken as example of what more popular apps offer to piano learners.

Simply Piano [2] is one such popular app, having more than 50 million downloads on the google play store at the time of this writing. The main feature of the application is the availability of a large amount of songs to perform, with live score following: the learner is expected to play the piano part (which is often simplified and even monophonic for beginners) while the application performs the part of other instruments to enrich the musicality of the experience. Either by using a MIDI keyboard connected to the device running the application or through using the device's microphone, the app analyzes the student's piano playing in real time to assess whether the notes are being played correctly and giving immediate visual feedback (by coloring the misplayed notes differently), going as far as stopping the execution if the student is struggling too much. Duolingo, the famous language-learning application, has recently started a "language" course in Music, which is similarly based on reading short songs played on screen, but these are played on a digital keyboard shown on the screen of the device rather than on an external piano.

Flowkey [3] offers score following similarly to Simply Piano, but while the latter often uses simplified parts, for example having the player execute the vocal line of songs instead of a piano accompaniment, Flowkey is more focused on actual

[2]https://hellosimply.com/simply-piano
[3]https://www.flowkey.com

piano partitions. The user has the option to only focus on the left hand or the right hand before moving to playing both at the same time, but there is no further simplifications of the part. This app also provides the learner with a video of the hands of an expert playing the same track to help the learner.

Pianote [4] goes a step further in this direction, being more based around video lessons and community support that the learner has access to in-app. Skoove [5] has a similar approach, and even offers online lessons with actual live teachers.

All of the above leverage gamification to some extent, but Musa [6] goes even further by offering an actual game that is played through a piano keyboard (again, either MIDI or by using microphones). In this case the entertainment values comes first by proposing the app as a game with educational benefits, instead of a gamified educational app.[21]

Some common features of these commercial approaches is the ability to follow the learner while playing to provide feedback (which is however often limited to stating if the correct note has been played at the correct time), and the availability of a large amount of scores to play, sometimes offering simplified versions. There is usually a suggested learning path and a planned difficulty curve, but the personalization is mostly left to the student who gets to choose what songs to play. We are not aware of any form of music generation in any of these systems.

2.3 A systematic investigation

Aiming to find out more conclusively how much generative AI is used for the personalization of music education; we performed a systematic review on the topic. While the detailed results of the review are reported elsewhere,[22] here we will sum up the methodology and the main findings.

Our hope was to find other research efforts in generating personalized exercises for Piano education, but we structured the systematic search using the PICO methodology[23] to include more generally any generative AI approach to the personalization of music education, to give a broader description on the subject. We thus retrieved 360 documents from four online repositories: Scopus, IEEE Xplore, ACM Digital Library, and Google Scholar. Following established methodologies,[24] we screened the papers to only include relevant documents that would pass some minimal quality requirements, which led us to 40 relevant papers on the subject.

The in-depth analysis of the papers led us to the conclusion that generative AI is still underutilized in the personalization of music education. Of the 40 papers, 13 were using generation to some extent, but only 8 of those were using generation to enable higher personalization of the educational material. Only 2 of these papers dealt with the generation of exercises, but both were related to the teaching of music theory, rather than piano technique. We therefore concluded that there is currently no active research effort in using music generation for the creation of a personalized set of exercises, as we are envisioning in this article. There were still valuable insights in those papers, beyond the goal that we anticipated, which we sum up here.

2.3.1 Use of generative AI for music education

While we found that generation of technical exercises is still under researched, generation was used by the retrieved papers in a variety of ways:

- Generation of Music: music can be embedded in musical lessons to complement or expand the music that is being performed by the student (and teacher). For example, Li[25] used AI to generate accompaniment for singing lessons, and Addessi et al.[26] created a system to generate responses to children's musical improvisations.

- Feedback: in instances there is some automatic evaluation of performances, instead of giving plain feedback to the learner it is possible to create a natural language response using LLMs to better encourage the learner.[27]

[4]https://www.pianote.com
[5]https://www.skoove.com/
[6]https://play.google.com/store/apps/details?id=com.playmusa.musa

- Music Theory Exercises: as mentioned, the generation of exercises to test the knowledge on theoretical aspects of music has been explored.[28,29] Possibly, this is more common than technical exercises because of the knowledge-based approach, as opposed to the skill-based exercises that are needed for technical practice.

- Visualizations: the musical practice can be enhanced by generating visualizations that complement the normal practice, for example warning the learner of difficult to perform parts[30,31] or to compare movements with those of more expert pianists.[32]

2.3.2 Personalization of music education

The generation of exercises is, of course, not the only possible way to personalize music education. What follows is a list of some of the more common approaches found during the systematic review.

- Recommendation and curriculum adaptation: by far the most common approach to personalization we found is the use of recommender systems. These systems are inherently personal, as they try to suggest items that are relevant to each different user of the system. There can be however different ways to suggest relevant items. One natural way is to select items of appropriate difficulty to suggest to the learner based on their skill level (similarly to what we propose),[19,33,34] but it is also possible to simply select music and musical exercises based on the taste of the student, to improve motivation,[35,36] or more exotic approaches like suggesting music based on the current mood of the student.[37]

- User Interfaces: another well researched approach is that of allowing the user to modify the graphical interface of the system to make it personalized in ways that improve the effectiveness of the educational endeavor. Fitton et al.[38] report the interesting finding that allowing even minimal personalization to an avatar (such as gender, skin color, and hair color) can improve the feeling of embodiment with the avatar, improving the effectiveness of a dance teaching system that was based on imitating the avatar's dance moves. Another example is that of using rich digital score formats that allow the user to display different information at the same time depending on their need, such as color coding for notes (useful to dyslexic students), or pneumatic notation for early music.[39] The above-mentioned generated visualizations also fit in the category of User Interface personalizations.

3. Steps forward

The analysis of the literature shows us the importance of personalization, and yet the use of generation to that goal is currently limited. Our confident guess is that in the near future we are going to witness increasing amounts of generative AI being embedded in music education, but probably exercise generation is still going to take some years before being used in practice, as research on this is still in its preliminary stage, and because music generation systems ate still somewhat lacking in quality. This section tries to foresee some possible paths that these eventual systems could take in creating exercise or learning material that is personalized to their users. To be realistic with our predictions, we shall describe things that are currently possible with the technology available to us. This will result in mostly self-fulfilling prophecies, as the reported examples come from open research venues being explored within project CALIOPE, but we will also describe some more ambitious predictions of things that are yet to be conceived, describing potential research venues.

3.1 Analysis tools

As a first step towards personalization, we wish to provide analytical tools that estimate difficulty of a music piece in a manner that enables learner-specific difficulty ratings. As stated above, only by doing this it would be possible to meaningfully assess whether a new exercise or piece lies within the Zone of Proximal Development for that specific student.

3.1.1 Aspect-specific difficulty

We have discussed how a single scalar value is not sufficient to properly describe the ZPD, which is better understood in a multidimensional feature space. One way to increase the potential personalization of education is therefore to describe difficulty among different axes. This would allow a teacher to specify the desired exercise for a student in terms of different features, which can capture the different abilities of different students, as in the example we made in section 1.1. Going towards that direction, we experimented with the definition of three difficulty aspects that we can compute from a digital version of a music score. Fingering displacement was defined as the distance one's fingers must move from their estimated resting positions to play each note. Playing speed was defined as the inverse of the inter-onset interval (IOI) between each note. Lastly, Polyphony was defined as the ratio of chord counts (or beats) to non-chord counts (or beats),

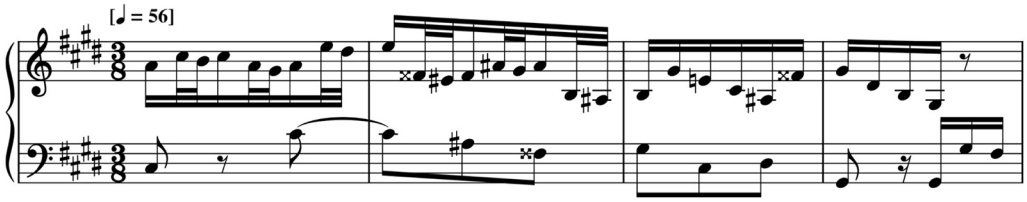


Figure 3. Measures 39–42 of J.S. Bach's Inventio (No. 6 from "Fifteen Two-Part Inventions", BWV 772–786), a tricky passage in terms of finger movements but that does not require chords at all.

computed separately for various chord sizes. Wanting to test the viability of these features, we employed two datasets: Mikrokosmos-difficulty (MKD)[19] and Can I Play It? (CIPI)[20] These datasets don't include the fingering annotations needed for the Fingering displacement feature, but those were added automatically using PianoPlayer [7]. Both datasets include labels of difficulty based on the 9-class difficulty scale designed by music publisher Henle [8] which is used to label each of the pieces in their catalog, allowing us to have an expert-made reference of difficulty for the scores in the datasets. These labels are once again a single scalar value that we wish to move away from, but as a validation for our features we use a Gradient Boost model (XGBoost) to predict the Henle difficulty rating based on the three features we have. The predictions with this model achieved a balanced classification accuracy of 29.48% and an MAE of 0.9040. The low MAE in particular suggests that even if there are wrong estimates the estimated label on the 9-class scale is not too far off from the ground truth (which in itself is open to disagreement, as Henle's expert Ralf Koenen who designed this scale admits). This shows that the designed features do correlate with the expert assessment on difficulty, and that their joint usage does a good job at predicting a piece's overall difficulty.

However, we wish to go a step further from that: we wish to see if the different features capture different aspects of a piece. We visually inspected those scores where the aspect-specific model predictions differed from one another by 2 or more classes of difficulty to assess whether those differences aligned with human intuition. For example, Bach's "Inventio 6" (a snippet of which is shown in Figure 3) received 9-class fingering displacement and polyphony predictions of 5/9 and 2/9, respectively. A visual inspection corroborated those predictions, in that the score contains few chords but numerous wide-ranging finger movements. Ultimately, such inspections provided evidence that the models were capturing unique dimensions of difficulty from one another. Moreover, this multi-feature classification system allows us to act separately on different features of a score, as we investigate in section 3.2.1.

3.1.2 A cognitive approach

A different approach to the personalized modeling of difficulty is to cognitively model the expected difficulty of a piece, using a model that would allow to adapt this expectation to the different users of the system. We argue that Information Content[40] is a valid measure to be used to model the cognitive load of the various elements on a score, which make up the difficulty of a piece: seen before elements will be easier than those that have never been seen, and more common elements will be easier than rare elements. Here "elements" is meant to be as generic as possible, including notes, chords, but also sequences of notes, intervals, and most generally any notation that can be found on a score and that can be recognized later. Indeed, there has been some evidence in prior literature that Information Content correlates to difficulty.[18] For further putting this idea to the test, we leverage IDyOMs a framework for multiple-viewpoint, variable-order Markov modeling as described by Pearce[41] which allows us to model those "elements" as "constituents" of an ad-hoc knowledge representation system. In this investigation, we used pitch, duration and fingering annotations as the viewpoints to be considered by the system. The system was trained on the above mentioned Mikrokosmos-difficulty dataset, but this time using the scale of three difficulty classes that Bartók divided his opus into (Beginner, Intermediate, Professional). The trained model yields probability estimates for all the elements in the analyzed pieces. The distributions of these are shown in Figure 4, divided by the three difficulty levels, numbered here 0 through 2.

A statistical test (Kruskal-Wallis H test, followed by a Dunn post-hoc test) confirmed an initial empirical assessment of the plots, showing that the full viewpoint, containing information about pitch, duration and fingering results in information content distributions that are more significantly different than other viewpoints. This result was stronger for comparisons between difficulties 0 vs 1, as well as 0 vs 2, while 1 vs 2 was not significant for the duration and fingering viewpoints. This confirms the hypothesis that information content is correlated to the concept of difficulty. As we

[7] https://github.com/marcomusy/pianoplayer
[8] https://www.henle.de/en/Levels-of-Difficulty/

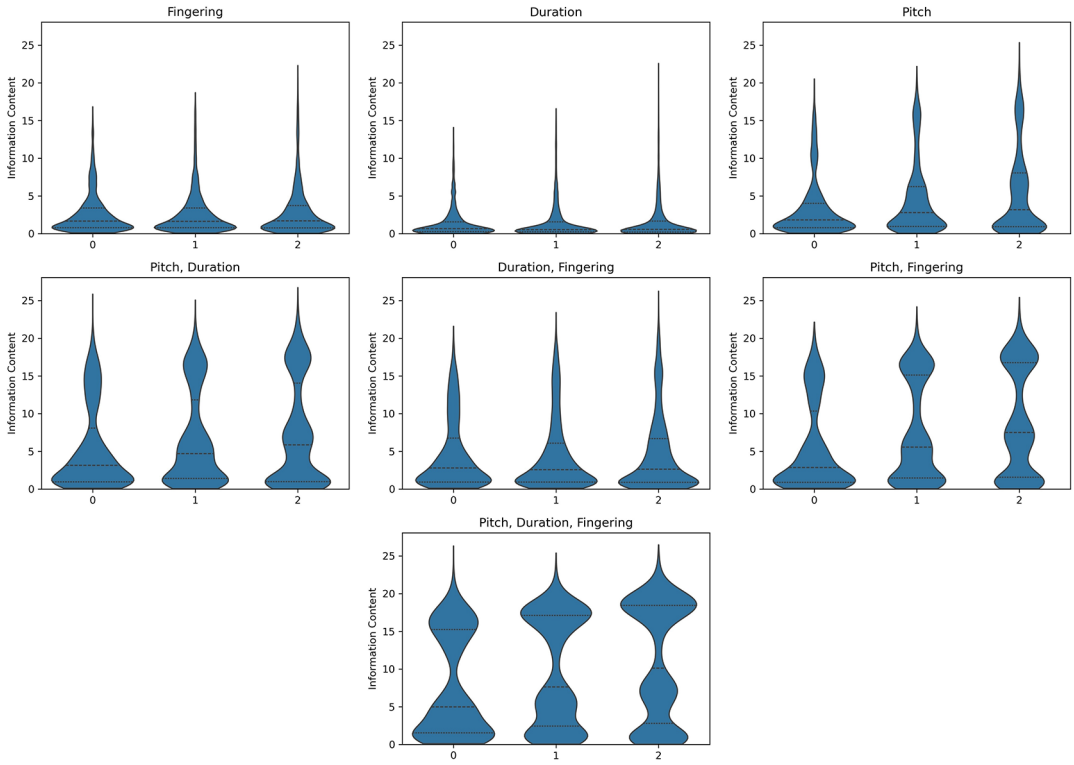


Figure 4. Information Content per level of difficulty (0, 1, and 2 corresponding respectively to Beginner, Intermediate, and Professional) computer over various features on the right hand parts.

mentioned this was expected from prior literature, but this is not merely a confirmatory study. While this is still under investigation, we argue that the approach we describe here can easily adapt to different datasets and more importantly to personalized datasets: providing the system with a list of pieces the student is familiar with should give a good indication of how easily the student can approach a new piece by calculating the information content of the new piece based on the probability distributions learnt on the known pieces. This approach would create a bottom-up description of the capabilities of the student, as opposed to the top-down approach described in the previous section.

3.2 Prototypes

From the results of the analyses above we were able to desing some proof-of-concept prototypes.

3.2.1 Aspect-specific reduction of scores

The model for the estimation of multi-aspect difficulty described above lends itself to some extra uses besides analysis. In particular, we designed a prototype system for the reduction of musical scores that would simplify a musical piece in different ways depending on which difficulty aspect the system is asked to make easier. Given a target difficulty for each of the three aspects and a score to simplify, the system employs a moving window approach that evaluates each difficulty aspect in different measures of the score, finding sections where each aspect is harder than the target. Then, a heuristic-based strategy removes one note at a time from those sections, repeating iteratively until the aspect-specific difficulty targets are met (or other stopping conditions occur to avoid excessive removal of notes). This system was evaluated qualitatively via an interview with a domain expert. Only one of four reductions presented to the domain expert was viewed as practically useful. In most cases, reductions were seen as having maintained or even increased score difficulty because of the introduction of occasional imbalance between playing hands, intermittent notational issues, and a tendency to remove single notes from challenging sequences without actually making those sequences easier to play. However, the different strategies managed to address different musical features as expected. We believe the main shortcoming of this proof-of-concept system is the limited approach to simplification, i.e., the iterative removal of single notes. Despite the unsatisfactory results of this first approach, we have found the windowed approach interesting in its ability to find difficult sections of a score, and we expect that embedding the aspect-specific difficulty estimators in a genetic algorithm capable of more varied and meaningful transformations of the musical content, possible joined with a system to factor the musical importance of the notes in the piece,[42] will lead to better results.

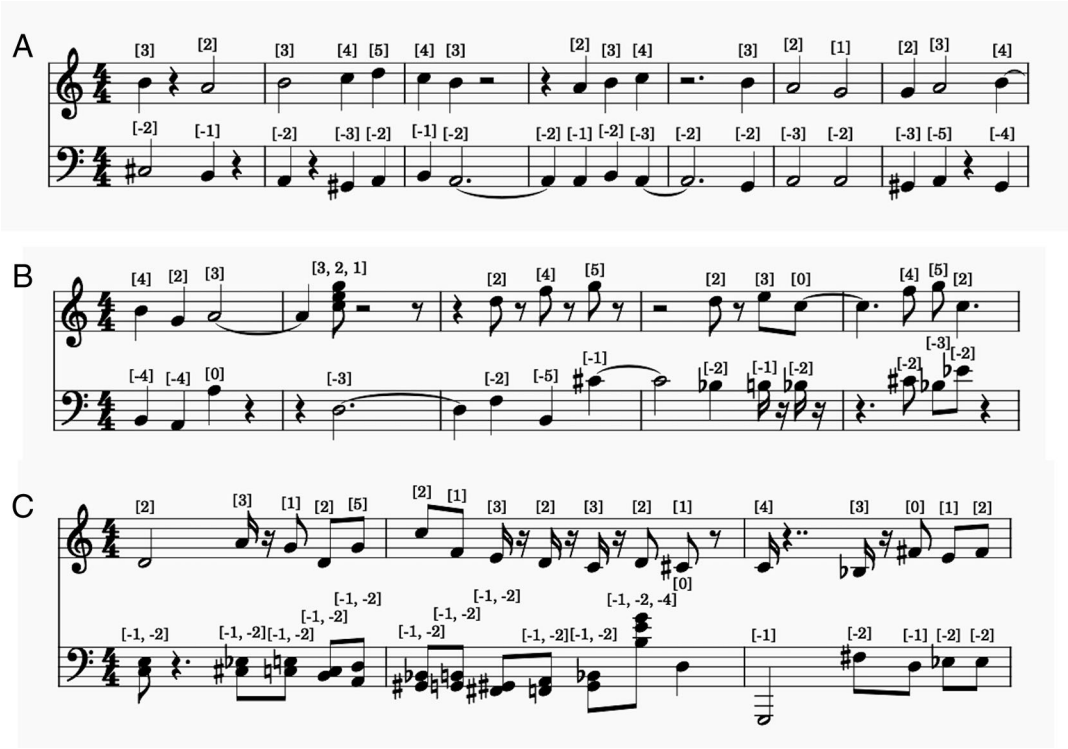


Figure 5. Beginning measures of an (a) easy (b) intermediate and (c) difficult generated pieces.

3.2.2 Learner-specific generation of scores

Going back to the use of IDyOMs to describe a learner's path and expectations when presented with a new piece, we must note that the same statistical models that are learnt to analyze difficulty can also be leveraged for generation of new pieces. Since IDyOMs uses the same concepts for extracting information content and for generating sequences based on information content, the patterns that contribute to the overall difficulty of a piece should transfer to the generated pieces. In Figure 5 we can see some examples of this generative approach. These are based on the generic three difficulty levels described above, but we remind here that it would be possible to change these with a selected dataset to reflect the pieces that are known to the student. On the one hand, it is clear that the model was able to generate simpler sequences when trained on the easy pieces as opposed to when trained on more difficult one. On the other hand, the markovian models used by IDyOMs are known to lack long-term structure, and thus the results lack overall musicality. It is therefore clear that a more sophisticated approach is needed, as we speculate in the next section.

3.3 Futurisms

In the first part of this article we described a path that nobody seems to have walked upon yet, and we described the first steps we took along those paths. Let us now describe what, having walked the beginning meters of this road, we can see up ahead.

The most obvious continuation of this research is to improve on the systems we described here. Our Information Content based approach can benefit from more rigorous testing on the field, and we can improve the its generative abilities by embedding other generation frameworks like Recurrent Neural Networks or the now more popular Transformer. As we mentioned, the reduction system can also leverage a more sophisticated way to alter musical content, and again neural systems could help achieve that goal. Besides these small incremental changes, we are more interested in the big picture developments. A lot of resources are available to learners on the internet, especially musical scores. Yet, it is not easy for an unexperienced player to choose a piece to learn based on its technical difficulty. Using the suggested technologies it would be possible to create a recommender system that chooses pieces based on the characteristics of the user, be it based on an assessment of multi-aspect skills, or on the history of pieces performed by the student. Yet, a student would probably prefer to play their favourite songs, rather than playing a random 18th century piece that was selected because the system established the user could learn it. The

reduction system could then come in handy in simplifying that song that is achievable yet challenging (i.e. it falls in the ZPD) or a generation system could create a new arrangement of the same song, keeping the musical aspect as intact as possible while keeping the accompaniment within the skill level of the learner. Of course this will not always be possible, but a generation system should also be able to create pieces that are within the style favoured by the learner and that are within his capabilities at the same time. Looking at a more traditional perspective, a conservatory teacher finding that a student is uncapable of performing the piece that the normal curriculum would require him to play could turn to a generative system to either create exercises that help the student train for the harder passages of the piece (which could maybe be identified in a manner similar to the sliding window analysis approach from our reduction system), or create a simplified version of the piece to lean into the more difficult exercise, or create an entirely new piece that adapts some of the hard techniques the student struggles with but acts as a bridge from the student's prior capabilities towards the new difficult piece. Commercial systems seem to like employing a gamified approach to learning. This has well-known educational benefits (besides improving customer retention), so it would be worth to investigate what is the best approach to gamification of music lessons. Generation can play a huge role in a gamified context: games are inherently interactive media, and creating music ad-hoc on the fly can open a great deal of novel human-computer interactions as well as serve as a complement for the interactivity of these games. Taking a step back, even more theoretical works on defining what makes a piece difficult would be of great benefit to future educational endeavours: while it is intuitively clear when a piece is hard or easy, it is not an easy task to describe specifically what makes it so. Explainable AI systems could maybe be an additional tool to aid experts in exploring more answers to this question. There are multiple ways to go forward, and while we intend to continue this investigation, we hope we will not be alone on this path, as we alone cannot possible hope to explore all the possibilities that are open.

4. Conclusions

In this article we have outlined the utility of using Generative AI approaches to the teaching of music, more specifically Piano technique, as these approaches can greatly improve the personalization of the provided education and, in turn, improve the effectiveness of education. Yet, we note from performing a rigorous scrutiny of the literature that these techniques are yet to be properly investigated by the academic world despite interest on AI research moving fast in the last couple of years. We then provided ways in which we have started investigating the subject, and ways we argue this exploration could and should continue. Hoping to see more efforts from researchers in this area, we wish to make a call for human-centric approaches to this application and to AI in general. Without considering and including the human actors of the areas we with to change with Artificial Intelligence, we may obtain disruptive change, but if the disruption is only achieved through the removal of the human from the loop, we risk creating an intelligence that is ultimately useless, if not straight up harmful, to us.

Ethics and consent

Ethical approval and consent were not required.

Data and software availability

Source data

The presented explorations used two datasets authored by Pedro Ramoneda and colleagues, both available on Zenodo:

1. Mikrocosmos-difficulty dataset (https://doi.org/10.5281/zenodo.6092709)[19]

2. CIPI Dataset (https://doi.org/10.5281/zenodo.8037327)[20]

Underlying data

Repository: Support data for "The Possibilities of Personalized Music Education Through Exercise Generation".

https://doi.org/10.5281/zenodo.18836323.[43]

The repository contains data and a script used to generate Figure 4, as well as the predictions used to compute the numbers reported in section 3.1.1.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

The authors wish to thank Nick Harley for help provided in the usage of Idyoms.

References

1. Mandanici M, Spagnol S, Ludovico LA, *et al.*: **A Timeline of Music Education Technologies.** *Digital Music Learning Resources, in SpringerBriefs in Education.* Singapore: Springer Nature Singapore; 2023; pp. 1–17.
Publisher Full Text

2. Smith M, Smaill A, Wiggins GA, editors. **Music Education: An Artificial Intelligence Approach: Proceedings of a Workshop held as part of AI-ED 93, World Conference on Artificial Intelligence in Education, Edinburgh, Scotland, 25 August 1993.** *Workshops in Computing.* London: Springer; 1994.
Publisher Full Text

3. Alkhatlan A, Kalita J: **Intelligent Tutoring Systems: A Comprehensive Historical Survey with Recent Developments.** Dec. 23, 2018. arXiv.
Publisher Full Text

4. Chen X, Zou D, Xie H, *et al.*: **Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions.** *J. Educ. Technol. Soc.* 2022; **25**(1): 28–47.

5. Wang Y, *et al.*: **NotaGen: Advancing Musicality in Symbolic Music Generation with Large Language Model Training Paradigms.** Mar. 21, 2025. arXiv: arXiv:2502.18008.
Publisher Full Text

6. Huang C-ZA, *et al.*: **Music Transformer.** Dec. 12, 2018. arXiv: arXiv: 1809.04281.
Publisher Full Text

7. Bloom BS: **The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring.** *Educ. Res.* Jun. 1984; **13**(6): 4–16.
Publisher Full Text

8. Chaiklin S: **The zone of proximal development in Vygotsky's analysis of learning and instruction.** *Vygotsky's educational theory in cultural context.* 2003; **1**(2): 39–64.

9. Brandao M, Wiggins G, Pain H: **Computers in music education.** *Proceedings of the AISB'99 Symposium on Musical Creativity, AISB Society for the Study of Artificial Intelligence and the Simulation of ....* 1999; pp. 82–88.

10. Kaleli YS: **The Effect of Computer-Assisted Instruction on Piano Education: An Experimental Study with Pre-Service Music Teachers.** *International Journal of Technology in Education and Science.* 2020; **4**(3): 235–246.
Publisher Full Text

11. Abdullah ZD, Mustafa KI: **The Underlying Factors of Computer Self-Efficacy and the Relationship with Students' Academic Achievement.** *Int J Res Educ Sci.* 2019; **5**(1): 346–354.

12. Smoliar SW, Waterworth JA, Kellock PR: **pianoFORTE: a system for piano education beyond notation literacy.** *Proceedings of the third ACM international conference on Multimedia - MULTIMEDIA'95.* San Francisco, California, United States: ACM Press; 1995; pp. 457–465.
Publisher Full Text

13. Rigby L, Wünsche BC, Shaw A: **piARno - An Augmented Reality Piano Tutor.** *32nd Australian Conference on Human-Computer Interaction, Sydney NSW Australia.* ACM; Dec. 2020; pp. 481–491.
Publisher Full Text

14. Wiggins GA, Trewin S; **A System for Concerned Teaching of Musical Aural Skills.** *Intelligent Tutoring Systems. Lecture Notes in Computer Science.* Gauthier G, Frasson C, VanLehn K, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000; **1839**: pp. 494–503.
Publisher Full Text

15. Sébastien V, Ralambondrainy H, Sébastien O, *et al.*; **Score analyzer: Automatically determining scores difficulty level for instrumental e-learning.** *13th International Society for Music Information Retrieval Conference (ISMIR 2012).* 2012; pp. 571–576.

16. Takidaira S, Shoji Y, Dürst MJ: **Slow, Repeat, Voice Guidance: Automatic Generation of Practice Charts to Improve Rhythm Action Games** *HCI in Games. Lecture Notes in Computer Science.* Fang X, editors. Cham: Springer International Publishing; 2022; **13334**: pp. 119–138.
Publisher Full Text

17. Ariga S, Fukayama S, Goto M: *Song2Guitar: A Difficulty-Aware Arrangement System for Generating Guitar Solo Covers from Polyphonic Audio of Popular Music.* ISMIR; 2017. 568–574.

18. Nakamura E, Yoshii K: **Statistical piano reduction controlling performance difficulty.** *APSIPA Transactions on Signal and Information Processing.* 2018; **7**.
Publisher Full Text

19. Ramoneda P, Tamer NC, Eremenko V, *et al.*: **Score Difficulty Analysis for Piano Performance Education based Fingering.** *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore.* Singapore: IEEE; May 2022; pp. 201–205.
Publisher Full Text

20. Ramoneda P, Jeong D, Eremenko V, *et al.*: **Combining piano performance dimensions for score difficulty classification.** *Expert Syst. Appl.* Mar. 2024; **238**: 121776.
Publisher Full Text

21. Micheloni E, Tramarin M, Rodà A, *et al.*: **Playing to play: a piano-based user interface for music education video-games.** *Multimed. Tools Appl.* 2019; **78**(10): 13713–13730.
Publisher Full Text

22. Carnovalini F, Espírito Santo L, Wiggins GA: **Personalized Music Education: A Systematic Review of AI Generation Methods.** *IEEE Access.* 2025; **13**: 207433–207447.
Publisher Full Text

23. Schardt C, Adams MB, Owens T, *et al.*: **Utilization of the PICO framework to improve searching PubMed for clinical questions.** *BMC Med. Inform. Decis. Mak.* 2007; **7**: 16.
Publisher Full Text

24. Carrera-Rivera A, Ochoa W, Larrinaga F, *et al.*: **How-to conduct a systematic literature review: A quick guide for computer science research.** *MethodsX.* 2022; **9**: 101895.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Li Y: **Application of Computer-Based Auto Accompaniment in Music Education.** *Int. J. Emerg. Technol. Learn.* 2020; **15**(06): 140.
Publisher Full Text

26. Addessi AR, Anelli F, Benghi D, *et al.*: **Child–Computer Interaction at the Beginner Stage of Music Learning: Effects of Reflexive Interaction on Children's Musical Improvisation.** *Front. Psychol.* Jan. 2017; **8**.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Matzek E, Yankee T, Kohler O, *et al.*: **VRmonic: A VR Piano Playing Form Trainer.** *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR).* Los Angeles, CA, USA: IEEE; Jan. 2024; pp. 330–334.
Publisher Full Text

28. Neubarth K, Weyde T, Bellini P, *et al.*: **Automatic Generation of Test Objects for Music Self-Study.** 2007.

29. Beuls K, Loeckx J: **Chapter 9. Steps Towards Intelligent MOOCs.**

30. Deja JA: **Adaptive Visualisations Using Spatiotemporal and Heuristic Models to Support Piano Learning.** *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization.* Utrecht Netherlands: ACM; Jun. 2021; pp. 286–290.
Publisher Full Text

31. Deja JA: **Piano Learning and Improvisation through Adaptive Visualisation and Digital Augmentation.** *Companion Proceedings of the 2022 Conference on Interactive Surfaces and Spaces.* Wellington New Zealand: ACM; Nov. 2022; pp. 41–45.
Publisher Full Text

32. Xiao X, Ishii H; **Inspect, Embody, Invent: A Design Framework for Music Learning and Beyond.** *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* San Jose California USA: ACM; May 2016; pp. 5397–5408.
Publisher Full Text

33. Yee-King M: **Pedagogical agents for social music learning in Crowd-based Socio-Cognitive Systems.** *Crowd Intelligence: Foundations, Methods, and Practices, CEUR Workshop Proceedings.* 2014; pp. 76–93.

34. Tamir-Ostrover H, *et al.*: **Automatic Evaluation of Aspects of Performance and Scheduling in Playing the Piano.** *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization.* Barcelona Spain: ACM; Jul. 2022; pp. 276–285.
Publisher Full Text

35. Xing W, Slowik A, Peter JD: **Edge-cloud computing oriented large-scale online music education mechanism driven by neural networks.** *J. Cloud Comput.* 2024; **13**(1): 55.
Publisher Full Text

36. Qi W-Q, Cao H-B: **A Study of Information-Based Teaching Strategies for the Saxophone Based on Deep Learning.** *Mob. Inf. Syst.* 2022; **2022**: 1–9.
Publisher Full Text

37. Ran Z: **Computer-Aided Approach for Personalized Vocal Music Instruction and Psychological Awareness in College Settings using Fuzzy Logic and Big Data.** *CADandA.* 2023; 244–264.
Publisher Full Text

38. Fitton I, Clarke C, Dalton J, *et al.*; **Dancing with the Avatars: Minimal Avatar Customisation Enhances Learning in a Psychomotor Task.** *Proceedings of the 2023 CHI Conference on*

Human Factors in Computing Systems. Hamburg Germany: ACM; Apr. 2023; pp. 1–16.
Publisher Full Text

39. Ludovico LA, Rita G: **Teaching Adaptively for Music.** *J e-Learn Knowl Soc.* 2014; **10**(3).

40. Pearce MT: **Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation.** *Ann. N. Y. Acad. Sci.* 2018; **1423**(1): 378–395.
Publisher Full Text

41. Pearce M: *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition.* London: City University;Accessed: Feb. 25, 2026.
Reference Source

42. Simonetta F, Carnovalini F, Orio N, *et al.*: **Symbolic Music Similarity through a Graph-Based Representation.** *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion - AM'18.* Wrexham, United Kingdom: ACM Press; 2018; pp. 1–7.
Publisher Full Text

43. Carnovalini F, Verstraeten L, Scofield S: **Support data for 'The Possibilities of Personalized Music Education Through Exercise Generation'.** *Zenodo.* 2026.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ✗

Version 1

Reviewer Report 30 April 2026

https://doi.org/10.21956/openreseurope.25198.r72496

© **2026 Aróstegui J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✗ **José Luis Aróstegui**
Universidad de Granada, Granada, Andalusia, Spain

The manuscript presents an essay-style, forward-looking discussion on the potential of AI-based music generation for personalized piano education, complemented by three technical "proof-of-concept" explorations: aspect-specific difficulty estimation, an information-theoretic approach using IDyOM, and prototype systems for score reduction and learner-specific generation. The topic is timely and relevant, and the paper is clearly motivated within current debates on personalization, learning analytics, and AI in education.
The combination of a broad conceptual overview with concrete computational experiments is a strength, as it helps bridge music education, music informatics, and AI-based generation. At the same time, because the paper reports specific quantitative results and prototype evaluations, the methodological description would benefit from greater clarity and completeness to support transparency, reproducibility, and critical appraisal.

1. Methodological aspects

1.1. Clarity and detail of computational methods

The sections on aspect-specific difficulty (3.1.1) and the cognitive approach based on information content (3.1.2) introduce interesting and well-motivated features and modeling choices. The operational definitions of fingering displacement, playing speed, and polyphony are clearly stated and grounded in the structure of the digital scores. Likewise, the use of the Mikrokosmos-difficulty and CIPI datasets and the Henle 9-class difficulty scale is appropriate and well justified.
However, once the paper moves from conceptual description to reporting concrete performance metrics (balanced classification accuracy, MAE, distribution plots, significance tests), the methodological description becomes rather terse. For example, the XGBoost model is introduced as a "Gradient Boost model (XGBoost)" used to predict Henle difficulty from three features, and performance is reported (balanced accuracy 29.48, MAE 0.9040) without specifying:

- The size of the datasets actually used in training and testing,
- How the data were split (train/test, cross-validation scheme, random seeds),
- Which hyperparameters were chosen for XGBoost (e.g., number of trees, maximum depth, learning rate),
- Whether any feature normalization or rescaling was applied, and

- Whether the reported metrics refer to a single split or are averaged over multiple runs.

Similarly, in the IDyOM-based analysis, the paper notes that models were trained on Mikrokosmos using three difficulty classes (Beginner, Intermediate, Professional), and that a Kruskal–Wallis test followed by Dunn's post-hoc comparisons was performed on the resulting information-content distributions. This is conceptually appropriate, but the statistical procedures would be easier to assess if the authors stated explicitly, for example:

- The number of pieces or events per difficulty level,
- How independence of observations was handled (per-event vs per-piece statistics),
- Which correction for multiple comparisons was applied in the post-hoc tests, and
- The effect sizes corresponding to the reported significance.

Given that the paper presents itself as an "essay" but nonetheless reports quantitative results that underpin several claims about correlation with expert difficulty ratings and the distinctiveness of information-content distributions across levels, I would encourage the authors to expand the methodological subsections sufficiently for readers to understand and, in principle, reproduce the analyses. This could be done concisely, perhaps in a short "Methods" box or an extended supplementary section.

1.2. Evaluation design and scope of the proof of concept

The description of the prototypes for aspect-specific score reduction and learner-specific generation is a valuable contribution, and the authors are commendably candid about their current limitations (e.g., lack of long-term musical structure in IDyOM-based generation, or the sometimes counter-intuitive results of the reduction heuristics).

However, the empirical evaluation of these prototypes is extremely limited:

- For the reduction system (3.2.1), the evaluation consists of a qualitative interview with a single domain expert, who judged only one of four reductions as practically useful. No further detail is provided on who the expert is, how s/he was recruited, what questions were asked, or how the qualitative feedback was analyzed.
- No evaluation is reported with actual learners, either in controlled settings or in exploratory pilot use.

For a paper that aims to argue for the pedagogical potential of these systems, it would strengthen the contribution to make the scope of the proof of concept very explicit: the reader should clearly understand that the current evidence is limited to correlations with expert difficulty labels and to the judgment of one expert on a very small number of examples, and that no claims are yet made about effectiveness with students. Even if full user studies are beyond the scope of this particular article, a clearer statement of these limitations would help set realistic expectations and guide future empirical work.

2. Use of the Zone of Proximal Development (ZPD)

The discussion of the Zone of Proximal Development (ZPD) in Section 1.1 provides a familiar educational-psychology framework for arguing why personalization and a smooth difficulty progression are desirable. The authors also use the ZPD to motivate the idea that difficulty is multidimensional, and that different learners may occupy different regions in a multidimensional "difficulty space". In this sense, the ZPD functions as a high-level conceptual justification for the subsequent technical work.

However, the ZPD is not operationalized in the methods or results:

- The computational models are trained on Henle difficulty labels and Mikrokosmos difficulty levels, not on any direct measure of a learner's current and potential development.
- None of the algorithms (XGBoost, IDyOM, heuristic reduction) explicitly encode or estimate a learner's ZPD; they estimate difficulty properties of pieces and fragments.

Thus, while the ZPD provides a useful conceptual anchor, it currently remains at the level of narrative framing rather than being integrated into the design or evaluation of the systems. The authors might consider either making this purely conceptual role explicit (e.g., “we draw on the ZPD as a motivating framework, although we do not directly model it in our experiments”), or sketching more concretely how future work might operationalize ZPD-related constructs (for example, by combining the proposed difficulty measures with longitudinal data on student performance).

3. Heuristic strategy for score reduction
In Section 3.2.1, the authors describe a “moving window” approach to identify locally difficult sections and then “a heuristic-based strategy” that removes one note at a time until the difficulty targets are met or stopping conditions apply. They then report that only one of four reductions was judged useful by the expert, and they reflect on the shortcomings of this approach.
Given that the reduction procedure is central to the prototype, I would encourage the authors to:

- Specify more precisely which heuristic is used (e.g., rules for choosing which note to remove in a given chord or passage, priority criteria between hands, treatment of melodic vs accompanimental notes),
- Clarify whether only one heuristic was implemented or whether multiple alternatives were considered, and, if possible
- Articulate why this particular heuristic was chosen as the first attempt.

At present, the phrase “a heuristic-based strategy” suggests a generic category rather than a specific, well-defined procedure. Providing more detail would help readers understand the results of the expert evaluation and the authors’ own diagnosis of why the current reduction method often fails to simplify playability.
4. Literature review and database coverage
The systematic review is described as drawing from four online repositories: Scopus, IEEE Xplore, ACM Digital Library, and Google Scholar, yielding 360 initial documents and 40 relevant papers after screening. This is a solid starting point and shows an awareness of both general bibliographic databases and discipline-specific outlets in computer science and engineering. However, given the educational and pedagogical orientation of the article, it is somewhat surprising that the Web of Science (WoS) Core Collection is not mentioned among the sources searched. WoS indexes many journals in music education, educational technology, and education more broadly, which are highly relevant to the topic of personalization and AI-supported learning. While Google Scholar can partially compensate for gaps in database coverage, it is less structured and more difficult to use for systematic screening. I would therefore suggest that the authors either expand the description of their search strategy to justify the chosen databases and clarify whether WoS was considered, or consider including WoS in a revised search, at least to check whether important work in music education and educational technology may have been missed. This would strengthen the claim that the review offers a comprehensive picture of generative AI for music education.
5. Language and typographical issues
The manuscript is generally readable and well written, but there are a number of minor typographical and linguistic issues that should be addressed. Some specific examples (non-exhaustive) include:

- Abstract: “We agree with that assessment.”
  Do the author mean they agree with the prior "statement"?
- Section 1, introduction: “(e.g., computer science, mathematics, language,
  and medical education. 3,4”

Here, the closing parenthesis is missing. It should be "(e.g., computer science, mathematics, language, and medical education). 3,4".
- Section 1.1: "zone of proximal develpment" appears in a figure caption and/or text; "development" is misspelled.
- Section 3.3: "we can improve the its generative abilities"
Either "the" or "its" should be removed. Possible corrections are: "we can improve its generative abilities" or "we can improve the generative abilities".
- Inconsistency in dataset naming:
The dataset is referred to as "Mikrokosmos-difficulty (MKD)" in one place and "Mikrocosmos-difficulty" in another. It would be advisable to standardize the spelling throughout.
- Contractions and punctuation:
In several instances, apostrophes in contractions appear to be missing (e.g., "dont" instead of "don't"), possibly due to typesetting or conversion issues. A thorough proof-reading should ensure consistency.

A careful language edit would remove these small distractions and improve the overall presentation quality, especially since the paper is otherwise conceptually clear.

These revisions would, in my view, significantly improve the clarity, methodological transparency, and impact of an already promising and thought-provoking manuscript.

Is the topic of the essay discussed accurately in the context of the current literature?
Partly

Is the work clearly and cogently presented?
Partly

Is the argument persuasive and supported by appropriate evidence?
Partly

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Music education. Curriculum. Teacher education. Educational Policy Studies. Global Studies in Education.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 30 April 2026

https://doi.org/10.21956/openreseurope.25198.r73200

© 2026 Argyriou M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Maria Argyriou
University of the Aegean, Mytilene, Greece

The article offers an original, timely, and intellectually stimulating contribution to contemporary scholarship on artificial intelligence in music education by exploring how generative AI may enhance personalized learning through the adaptive creation of learner-specific musical exercises. The manuscript is conceptually ambitious and clearly structured, presenting a coherent argument that current adaptive systems in music education remain pedagogically limited because they rely primarily on fixed repositories of instructional material, predefined learner models, and recommendation-based personalization. In contrast, the authors propose a more dynamic model in which AI systems generate exercises ad hoc, tailored to the evolving technical and cognitive profile of individual learners, thereby enabling a more refined form of pedagogical personalization. This central argument is theoretically grounded in educational concepts such as Bloom's "two-sigma effect" and Vygotsky's zone of proximal development, which are appropriately mobilized to support the claim that musical learning benefits from carefully calibrated instructional challenge aligned with learner readiness.

A major strength of the article lies in its clear and cogent presentation. The progression from theoretical framing to systematic literature review to the presentation of exploratory computational prototypes is logical, well developed, and accessible, while the use of concrete musical examples effectively illustrates the multidimensional nature of technical difficulty in piano learning. Particularly valuable is the authors' reconceptualization of musical difficulty as a multidimensional construct, including parameters such as fingering displacement, speed, and polyphonic complexity, rather than a singular scalar metric. This framework opens meaningful possibilities for more nuanced learner modeling and adaptive pedagogical design. The manuscript also contributes positively to the broader cultural and social understanding of the field by positioning AI not as a substitute for music teachers but as a human-centered tool capable of supporting differentiated instruction, widening access to individualized learning, and enhancing educational responsiveness.

Nevertheless, several aspects require further development in order to strengthen the manuscript's scientific robustness. Although the technological and computational literature review is substantial, the pedagogical framing remains somewhat narrow. The concept of personalization is discussed primarily in technical terms of difficulty adjustment and skill calibration, with limited engagement with broader scholarship in music pedagogy, embodied musical cognition, creativity, learner agency, and socio-cultural theories of music learning. Incorporating perspectives from contemporary music education research would deepen the educational relevance of the paper and broaden its conceptualization of musical development beyond technical proficiency. Furthermore, while the proposed computational models are innovative and promising, the empirical basis for claims regarding educational effectiveness remains preliminary. The presented systems function primarily as proof-of-concept prototypes, and the manuscript would benefit from clearer acknowledgment of their exploratory status, together with a more explicit discussion of how future classroom validation, teacher feedback, and learner-centered pilot studies might

assess pedagogical impact in authentic educational settings. Finally, given the article's advocacy for human-centered AI, a more explicit consideration of ethical dimensions (including algorithmic bias, equity of access, data privacy, and the preservation of teacher agency) would significantly strengthen its scholarly completeness.

Overall, this is a thoughtful and innovative article that identifies an important and underexplored direction for interdisciplinary research at the intersection of music education, pedagogy, and artificial intelligence. Its conceptual originality, strong organization, and future-oriented vision make it a valuable contribution to the field. To strengthen the manuscript's scientific soundness, the authors should address the following points: (1) broaden the pedagogical framework by engaging more substantially with contemporary music education scholarship beyond technical skill acquisition; (2) clarify the exploratory nature of the proposed computational prototypes and moderate claims regarding educational effectiveness until empirical validation is available; (3) outline a clearer framework for future classroom-based evaluation involving both teachers and learners; and (4) expand discussion of ethical considerations related to AI-driven personalization, particularly in relation to equity, bias, privacy, and teacher agency. With these revisions, the manuscript would make a strong and meaningful scholarly contribution to the field.

Is the topic of the essay discussed accurately in the context of the current literature?
Yes

Is the work clearly and cogently presented?
Partly

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Music Education; Artificial Intelligence in Education; Educational Technology; Music Pedagogy; Early Childhood Music Education; Digital Learning; Interdisciplinary Arts Education

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.