5-2023

Music Mentor

Jacob Webb

Music Mentor

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science
Computer Science

---

by
Jacob Webb
May 2023

Supervised by:
Dr. Bart Knijnenburg
Dr. Carlos Toxtli-Hernandez

Abstract

Extra-curricular learning is on the rise, and many are interested in expanding their current knowledge by utilizing the recent increase in educational technology. While many forms of educational technology exist, there are few interactive and engaging platforms that teach music theory. Apps such as Perfect Ear and MyMusicTheory are great for becoming familiar with reading music and recognizing pitches, however, they often become dry with repetition and repeated tasks. By combining existing technologies that can complete real time conversions from raw audio to MIDI, our goal was to gather information such as harmonies, key and compatible chords from the user's input. Using this data we aimed to create dynamic lesson plans based on user input, rather than using the same repetitive prompts from overused question pools. We were successfully able to generate these lesson plans, however, the lesson plans that we were able to create are somewhat limited. Given time restraints, we struggled to implement the pruning of audio input to match the desired lesson plans, as the recorded notes must match the correct format to generate a successful plan. Furthermore, we were not able to train a reliable voice model to recognize notes before the project was due. Though not fully complete, we successfully created a prototype dynamic lesson plan that can potentially engage users and assist in the learning of music theory if implemented in future technology.

Table of Contents

Chapter 1

Introduction

With educational apps like Duolingo having over 56.5 million monthly users, it is clear that educational technology is becoming an increasingly popular medium for learning [1]. With this rise in the use of educational technology, the efficacy of any new technology that is able to push the bounds of learning is drastically increased. Therefore, it is our responsibility to make learning as effective as possible. Seeing as how music theory apps do not appear in the top 50 apps in the App Store when searching for education, it is clear that improvements in the current music theory technology can be made. Upon investigation of such music theory apps we found lesson plans that were much more basic than that of sophisticated and popular apps like Duolingo. Where Duolingo adjusts lesson plans so that the difficulty matches the experience of the user, music theory apps like Perfect Ear ask extremely similar and basic prompts with every question regardless of the user's success rate. Duolingo attempts to gain as much information on each user as possible so that it can design lesson plans suited for each individual [2]. While the artificial intelligence that Duolingo uses to track its users' experiences is likely beyond our scope of ability for this paper, our goal was to create a similarly dynamic and targeted learning experience to increase the user's enjoyment, engagement and knowledge by basing the lesson plan and its corresponding difficulty on the user's own input.

Most apps are currently limited by restrictions of what can be taught. For example, in Duolingo you can select topics to learn, such as words related to eating or words related to the park. However, that is the limit of control the user has on the subject. By prompting the user for musical input in the form of humming or singing prior to the generation of their lesson plan, we are able to extract the notes they sing and create lessons based purely on the notes that they created themselves. What makes this project unique is that it designs the lesson plans around the user's audio input. Since the user is the creator of their own lessons, they can determine the difficulty of any given lesson by changing the complexity of their vocal patterns to correspond with their desired difficulty. This gives the user more control over the lesson plan, and provides a new and exciting way to learn. Not only will they be able to hear the changes that they make with each new lesson, but they can also visualize the sheet music containing the notes they input to see firsthand what different notes look like when written. We aim to promote a curriculum of self-actualization by allowing the user to set their own difficulty, by promoting as much experimentation as possible, and by implementing a 'guess and check' type experience for each question. The goal of self-actualization is to put the user in an environment that is suited for their own individual learning [6]. Rather than seeing the user or student as a checkbox that is ticked upon completion of the lesson plan, we aim to create a sandbox-style learning environment in which the user can learn at their own pace, dependent only upon their own curiosity. In this way, the curriculum is focused more so upon the will of the user, rather than a strict, pre-defined lesson plan. By creating lessons this way, we hope to not only give the user a great experience, but to help them learn in a new and interesting way. We hope that the ability to change the lesson plan increases interest and experimentation among users. We hope that this potential newfound interest in music can create an environment that is more conducive to joyful and effective learning.

Chapter 2

Related Work

Much research has been done in the field of music generation, alteration, and the teaching of music theory. Generating models from sample audio input to assess genre similarity, compatible chord progressions and other important information has been done in many research papers [5] [12] [9] [8]. Papers have also aimed to teach users about music in an interesting and engaging manner [5] [9]. Prior studies have been published regarding the effectiveness of dynamic learning in the classroom [17]. Other studies have been done that discuss best 'static' teaching methods [11] [4] [6]. Research has also been done in the realm of the effectiveness of multimodal learning in music theory [7] [10]. Finally, DuoLingo implements advanced algorithms that would be quite useful in this project [15] [14] [16]. Our goal is to expand upon all of these previous ideas and to not only generate models based on user input, but to ask users unique questions based on their input that can effectively teach them music theory.

In the paper titled, "Musical and Conversational Artificial Intelligence," the researchers made an AI that assists in the creation of music through voice-based interaction in the hopes that anyone can create music, without having to have an extensive background and knowledge in cutting edge digital audio workstations [5]. Using the Short-Time Fourier Transform to analyze vocal input from a user, in conjunction with a conversational intelligence, the project was able to successfully create a platform that is able to assist in basic music composition. Music Mentor builds upon the idea of helping those who are new to music, allowing them to learn more. However, we aim to give the user an easy way to learn music theory, rather than creating music directly.

A paper titled, "A Functional Taxonomy of Music Generation Systems," discusses the components needed to compose a full song based on a set number of inputs [8]. By assessing the successful and failed components of existing mechanisms that were designed to create music, they created a taxonomy of what they believe to be a successful music composing technology. One of the largest components of Music Mentor is the manipulation and generation of notes and chord sequences from input in the process of teaching the user. While we do not compose music directly, the creation of chords and rhythm will be needed in future iterations of this project. Creating chords based on a monophonic input are already required for this iteration, and this taxonomy can serve as a good example of system design.

Another paper focuses on the composition of music based on a monophonic melody [9]. This project uses functional scaffolding for musical composition (FSMC) to generate music. The key point of the FSMC approach is to keep every part of the composition related to one another. By doing so, the result is hoped to be one of synchronicity and harmony. This paper is similar to ours, since we too are attempting to generate music from a monophonic melody. However, as previously mentioned, instead of generating an entire composition, we are simply finding harmonies and chords from the monophonic input and attempting to use it as a medium to learn.

Another paper, "Data-Driven Exploration of Musical Chord Sequences," was made with the goal of using databases of music to generate interfaces that show what the most popular chord progressions are [12]. By using complex equations, they created a polygonal slider that allows you to mix and match genres of music to generate chord sequences corresponding to your selected genres and assigned weights. Similarly, our project aims to generate chord sequences from an input. The paper we are examining uses a large database of chord progressions and their corresponding genre to assign valid progressions for the user. Implementing a similar technology would be very useful for scoring questions based on chord progressions. If we could build a database of previous answers, we could then base accuracy of the current answer on its similarity to previous chord progressions of similar inputs.

A paper was written about the use of a self-actualization curriculum in the classroom and its benefits when introduced to ninth grade English students [11]. The teacher who implemented this style of curriculum in the paper believed that self-actualization was beneficial to students who were not going to attend college. Since there was no goal of getting a college degree amongst these students, there was no need to learn what seemed like arbitrary knowledge that carried little effect in the students' lives. As a result, the teacher adopted the strategy of self-actualization. By letting the students learn in their own way, he believed they were able to envelop themselves in the course material with more enthusiasm and engagement. Similarly, our project aims to engage the user with the content. We aim to teach through self-actualization by giving the user a variety of ways to complete any given task. By giving them this freedom, they can experiment and discover things on their own. Hopefully it will increase engagement and enjoyment just like it did for the ninth graders.

In the paper, "Cooperative Learning In The Music Theory Classroom," the authors attempt to change music theory learning habits by splitting students into groups and having them cooperatively discuss, learn and teach lessons together [17]. Furthermore, groups would discuss and critique one another on larger assignments. They believed that learning in this dynamic way created an attitude that was conducive to learning, rather than forced. This approach is quite similar to the self-actualization technique mentioned prior. Once again, we aim to give the user as much free-reign over their lessons as possible. In this way, we implement a similar attitude change that we hope will be beneficial.

Research has been done to find the impact of learning keyboard skills using software, rather than a one on one session with a teacher [4]. The study points out that it is difficult for a single teacher to instruct many students in learning the keyboard, as normal piano lessons are one on one. In this study, they used software connected to their keyboards, called "SmartMusic," to complete lesson plans. Though apprehensive at first, the vast majority of students found learning with software was extremely beneficial. This reinforces the ability Music Mentor has to teach music theory. It has been proven in the past that using software to learn music theory is possible, and even preferred by some.

"Popular Music Education in and For Itself, and For 'Other' Music: Current Research in the Classroom" discusses the quality of music theory curriculum as it applies to type the of music being taught [7]. The paper suggests that restrictions placed upon music being taught to students drastically impairs their musical fervor later in life. The paper believes that an emphasis on learning both classical music, and popular music will have a better outcome for the student and their future in music. Music Mentor allows for any vocal input, which satisfies this suggestion.

One study discusses how multimodality can contribute to learning languages [10]. The paper argues that using multimodal instructions can increase the efficacy of the material. They used an example of Swedish poetry that was being taught to students. Instead of simply writing out the words and explaining what they meant, the students were shown a line of poetry, accompanies by the audio of someone reading it, and a video describing the words in the poem. By adding these many modes of learning, the paper believes understanding can increase. Similarly, Music Mentor utilizes multimodal teaching for music theory. By letting the user not only place the note, but also hear the note they are placing, we introduce multimodal teaching. Also, the user can hear what they have sang, and even sing for some questions.

DuoLingo has multiple studies that apply to our research. Two of these papers discuss the way in which DuoLingo selects questions to suit the user as they complete the lesson. In one paper, they discuss the use of machine learning and natural language processing to accurately gauge the proficiency of a user [15]. The other paper generates learning curves from data that can be used to accurately predict error probability of future questions [14]. These advanced methods and algorithms used to determine future question choices and likelihood of success would be extremely useful in an application like Music Mentor. Though we are not at this point of development, the fully functional version of Music Mentor would use a combination of DuoLingo-style algorithms to successfully recommend valid questions at the user's current level of music theory competence.

Chapter 3

Methodology

In this section we discuss an overview of our methods. Major areas of the project include the creation of a web application and its subsequent components (Section 3.1), integration of existing music transcription software in our project (Section 3.2), and the design of basic lesson plans for the user to complete (Section 3.3). To meet all of our goals, we set out to tweak and combine multiple existing technologies.

3.1 Web Application

Our first step was to create an application on which we could test the technology. Using the React Native JavaScript framework, we created a web application that we could begin testing on. Using React Native, we created a component that mimics sheet music and allows the user to manipulate notes as if they were writing music in person. By populating this component with the transcribed notes from the user's input (Section 3.2), we could then prompt the user with questions related to their melodies. We allow the user to select the position on sheet music they would like to place a note. When they hover over a position, it displays an opaque note that is darkened when clicked. If the user wants to select a sharp note, they can drag the mouse from the original position slightly to the right and it will produce a sharp note. Similarly, flat notes can be created by moving the mouse slightly to the left from the original position. If the user wants to create a half note or whole note, they can click the desired note and drag their mouse to the right. If they drag it a medium distance and release, it will change the original note to a half note. If they click and drag

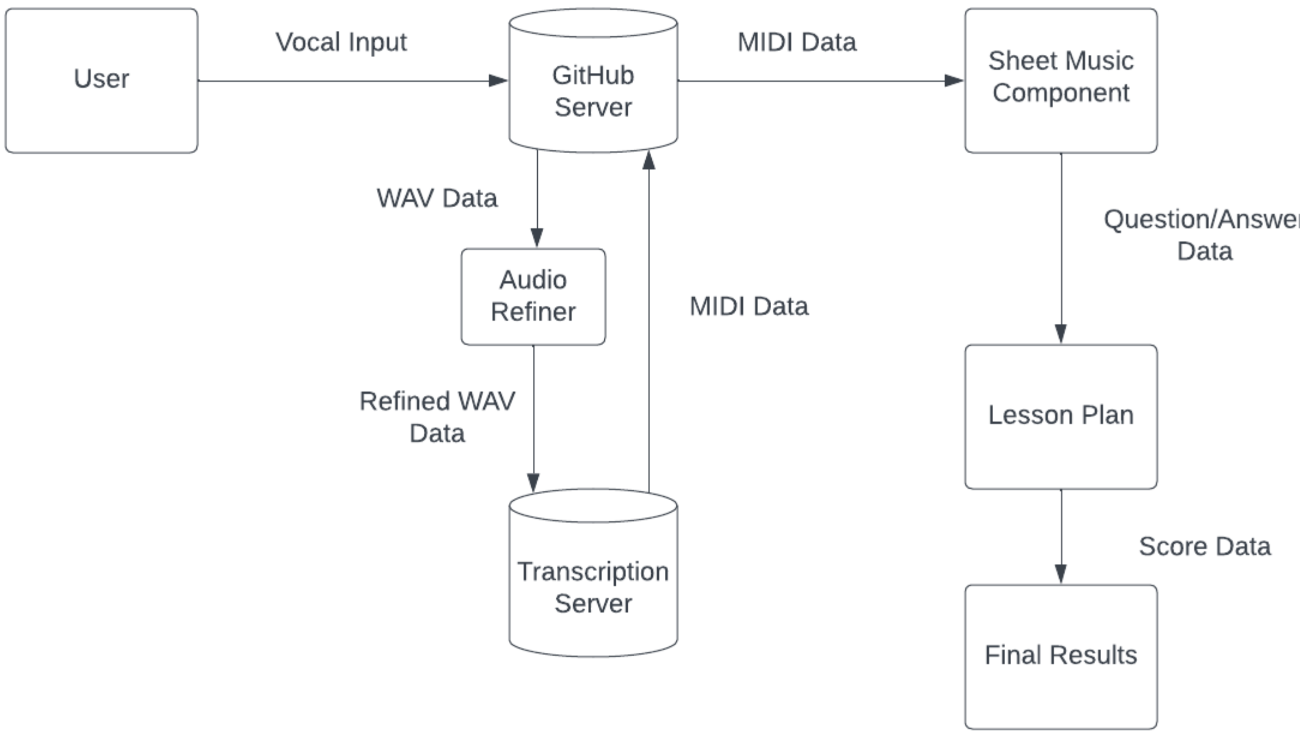

Figure 3.1: System diagram of completed project

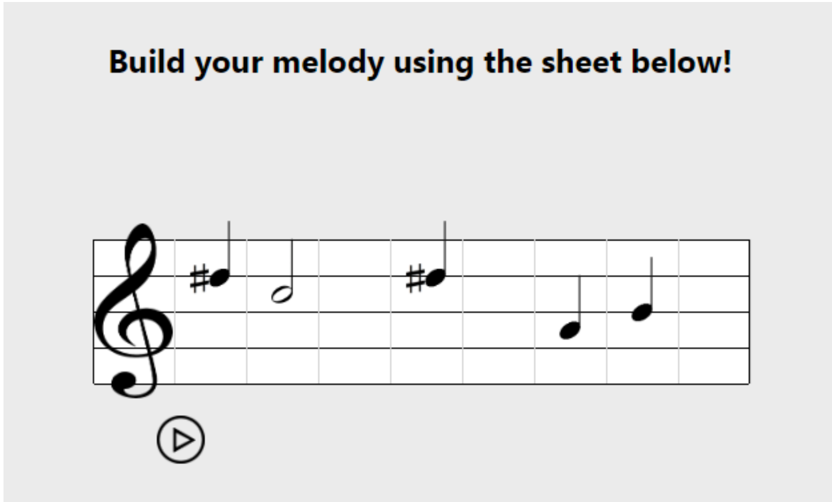

Figure 3.2: Example question from basic lesson plan

the mouse from the original position a far distance to the right, it will create a whole note when they release. In this way, the user can create any melody in the sheet that they want. The melodies that they create are made to answer questions asked, such as questions regarding harmonies and chord progressions related to their vocal input. When they are satisfied with their answer, they can continue. They will be told if they were correct or not, and they will be shown the correct answer.

3.2 Transcription

To accomplish the transcription of the user's musical input, we needed to find a way to interpret audio input and use it in our lesson plan. We decided to use transcription software that can convert WAV file data into MIDI data that can then be used to generate lesson plans. In order to accomplish this task we utilized Magenta's open source MT3: Multi-Task Multitrack Music Transcription software (MT3) [3]. MT3 uses models that were trained with Google Research's T5X framework to transcribe musical input. T5X is a JAX-based library for training, evaluating and inferring models using advanced neural networks [13]. MT3 has two pre-trained models, but it does not have a model trained to transcribe voice. As a result, the MIDI output is skewed slightly when met with complicated vocal input. The transcription software resides on a separate server that communicates with our GitHub Pages server via HTTP request, passing back and forth the music data. This software is extremely helpful since it can accomplish our goal of transcription; however, it does this transcription very slowly. MT3 is designed to be used with power supplied by a CPU or GPU. We had access to a server with large amounts of memory, so we opted to use the GPU version for this project.

3.3 Lesson Plans

Finally, our last major objective was to give the user a valuable lesson plan that could keep them engaged and teach them effectively. Originally we planned on designing multiple lesson plans to give the user variety, but due to time constraints we were not able to accomplish this. Rather, we created a single "proof of concept" lesson plan that is designed to teach the user basic transposition and sheet music recognition skills. This lesson plan consists of five questions that ask the user a range of prompts regarding their melody. We tried to grade each question open-ended when we could. If the question involves sheet music writing, we grade questions on percentage of correct notes. If the question is multiple choice, then we give a percentage of one hundred percent for a correct answer and zero percent for an incorrect answer. Finally, we show the user the correct sheet music after they finalize their answer so they can learn which notes were answered incorrectly. For all of the questions in this lesson plan, difficulty of the question corresponds with the complexity of the vocal input. As a result, an experienced user can create a more difficult lesson plan by making their melody extremely complex, whereas beginners can create a simple melody for an easy plan, and slowly ramp up in difficulty.

The first question prompts the user with a blank sheet music component and asks them to build their melody using the note builder. When the user clicks on a location in the sheet music, the corresponding note is sounded. The user can then also listen to their raw recorded audio to compare the notes one by one until they are satisfied with their answer. This 'guess and check' technique is promoted in hopes that it can give the user a curriculum of self-actualization, where the user is given more control over what and how they learn. This lesson aims to help the user learn how to read and write sheet music. The second question asks the user to select what key their melody is in based on what they learned from the sheet music in the prior question. By letting the user listen to their raw audio and audio queues from each multiple choice answer, they are able to select their answer. This lesson helps the user learn to recognize key by ear. The third question populates the sheet music component with the melody that they sang and asks the user to draw the third harmony of the melody. This lesson aims to help the user learn how to make chords, as third harmonies are elements of chords. The fourth question asks the user to transpose their melody to a different key, changing notes to sharp or flat with the key change as needed. This lesson plan aims to teach the user transposition between different keys. Finally, the fifth question asks the user to write chords in a given format to complement their melody. This lesson is the final question, and it gives the user plenty of flexibility with their answer. It not only teaches the user about building chords, but it allows for creativity in answers, as there are many valid chord progressions that satisfy this question.

Chapter 4

Vision

This project has much potential, and with further development it could be seen all the way to the end. Currently there are many restraints to the functionality of the app. While it works on the surface level, there are many things that need to be accomplished for the app to run smoothly and function without error. To complete the project, changes must be made to the sheet music component, the transcription software and the lesson plans.

4.1 Limitations

While we made it quite far into the project, not all goals were met and some functionality was not completed to fruition. The fastest note that we implemented in the program is a quarter note, meaning if the user sings something too fast, the input will not be parsed accurately. The audio input can also be skewed if the user sings out of time. To prevent this, we began implementing a metronome to count beats while singing, but it was not completed in time. Similarly, the sheet music component is limited to a fixed number of bars, meaning the input is limited by length. The sheet music component also does not support time signatures or tempo currently. Furthermore, due to restrictions by the speed of the transcription, larger input would drastically increase the time of lesson generation, causing an unsatisfactory result. The feedback mechanism was not completed in time, and accuracy scores are not given at the completion of each question as a result. With only one user, it currently takes approximately seven seconds to generate MIDI output. With more users, it would take much longer to accomplish this task. For that reason, an alternative method to generate transcription would be ideal.

4.2 Future Work

In further development we will aim to add more functions to the sheet music component. Seeing as how the sheet music component is what connects the user with most of the lessons, it is vital that it takes maximum precedence. In order for it to be usable, the sheet music component will need to allow for all types of notes, ranging from thirty-second notes to whole notes. It should include rests, dynamics such as crescendo, different clefs, time signatures and the ability to change key signatures. In a perfect world, the sheet music component would allow for much longer inputs so the user can create longer melodies that are much harder to complete than the short, simple melodies.

The next step in future development would be fixing the transcription software. There are currently only a handful of ways to do this. Firstly, we need to train a model to recognize voice. This would make the recognition of vocal input much better. This would then allow for the complicated input that we strive for. The only other thing we can aim to improve is the speed at which the audio is transcribed. By dedicating more memory to the transcription software we can speed it up.

In the final version of this app, there will be a multitude of lesson plans available to the user. There are many possible lesson plan categories, such as plans that focus on transposing music, recognizing and designing harmonies, creating and recognizing chords and more. Eventually, we aim to get to the point where we can automatically generate lesson plans using similar algorithms to DuoLingo. Using machine learning and databases of collected information could yield endless question choices. By creating basic templates for lessons, we could use these advanced techniques to generate an endless amount of potential questions to ask users of any experience level. While lesson plan difficulty is meant to be dependent upon the difficulty of the musical input, we can also scale the difficulty by changing the question. Questions such as advanced harmonies have a higher baseline difficulty than questions such as transposition. Therefore, adjustments in question difficulty can still occur even for beginner users who prefer basic melodies. The opportunities for lessons are exciting and endless. However, they are not able to be implemented well until the prior two steps are complete. After that is done, development can begin on new lesson plans and this app can reach its full potential.

Chapter 5

Conclusions and Discussion

In the future, with further development we hope to see this project come to fruition. By increasing maximum input length, better pruning of user input and adding more variety to the sheet component, we will allow for more complicated and interesting melody generation. Furthermore, with a properly trained vocal model, transcription software will be able to process user input quicker and more accurately, which will increase usability and engagement. Once all of that is completed, we hope to create a multitude of lesson plans and design an interesting learning experience for anyone interested in learning more about music. With continued development, we will be able to fix many of the design flaws and failures of this project. Although not all of our goals were met, our software still serves as a useful prototype for the future of dynamic music theory lesson creation.

Bibliography

[1] Duolingo announces 51% DAU growth and record bookings and revenue in third quarter 2022; raises full year guidance - Duolingo, Inc. Accessed April 29, 2023.

[2] How Duolingo uses AI in every part of its app, August 2020.

[3] MT3: Multi-Task Multitrack Music Transcription, May 2023. original-date: 2021-11-03T22:15:44Z.

[4] Michael R. Callahan. Teaching and Learning Undergraduate Music Theory at the Keyboard: Challenges, Solutions, and Impacts. *Music Theory Online*, 21(3), September 2015.

[5] Fabio Catania, Giorgio De Luca, Nicola Bombaci, Erica Colombo, Pietro Crovari, Eleonora Beccaluva, and Franca Garzotto. Musical and conversational artificial intelligence. 03 2020.

[6] Elliot W. Eisner. *Conflicting conceptions of curriculum, edited by Elliot W. Eisner and Elizabeth Vallance.* Series on contemporary educational issues. McCutchan Pub. Corp., Berkeley, Calif, 1974.

[7] Lucy Green. Popular music education in and for itself, and for 'other' music: current research in the classroom. *International Journal of Music Education*, 24(2):101–118, 2006.

[8] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. A functional taxonomy of music generation systems. *ACM Computing Surveys*, 50, 09 2017.

[9] Amy K Hoover, Paul A Szerlip, Marie E Norton, Trevor A Brindle, Zachary Merritt, and Kenneth O Stanley. Generating a Complete Multipart Musical Composition from a Single Monophonic Melody with Functional Scaffolding.

[10] Petra Magnusson and Anna-Lena Godhe. Multimodality in Language Education – Implications for Teaching. *Designs for Learning*, 11(1):127–137, August 2019.

[11] George E. Newell. The emerging self: A curriculum of self-actualization. *The English Journal*, 66(8):32–34, 1977.

[12] Eric Nichols, Dan Morris, and Sumit Basu. Data-driven exploration of musical chord sequences. 02 2009.

[13] Adam Roberts, Hyung Won Chung, Anselm Levskaya, Gaurav Mishra, James Bradbury, Daniel Andor, Sharan Narang, Brian Lester, Colin Gaffney, Afroz Mohiuddin, Curtis Hawthorne, Aitor Lewkowycz, Alex Salcianu, Marc van Zee, Jacob Austin, Sebastian Goodman, Livio Baldini Soares, Haitang Hu, Sasha Tsvyashchenko, Aakanksha Chowdhery, Jasmijn Bastings, Jannis Bulian, Xavier Garcia, Jianmo Ni, Andrew Chen, Kathleen Kenealy, Jonathan H. Clark, Stephan Lee, Dan Garrette, James Lee-Thorp, Colin Raffel, Noam Shazeer, Marvin

Ritter, Maarten Bosma, Alexandre Passos, Jeremy Maitin-Shepard, Noah Fiedel, Mark Omernick, Brennan Saeta, Ryan Sepassi, Alexander Spiridonov, Joshua Newlan, and Andrea Gesmundo. Scaling Up Models and Data with $\texttt{t5x}$ and $\texttt{seqio}$, March 2022. arXiv:2203.17189 [cs].

[14] Burr Settles and Brendan Meeder. A Trainable Spaced Repetition Model for Language Learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1848–1858, Berlin, Germany, 2016. Association for Computational Linguistics.

[15] Burr Settles, Geoffrey T. LaFlair, and Masato Hagiwara. Machine Learning–Driven Language Assessment. *Transactions of the Association for Computational Linguistics*, 8:247–263, December 2020.

[16] Matthew Streeter. Mixture Modeling of Individual Learning Curves.

[17] Lawrence Zbikowski and Charles Long. Cooperative Learning In The Music Theory Classroom. January 1994.