Summer 8-2018

Creating a Better Technological Piano Practice Aid with Knowledge Tracing

Max Feldkamp
Max.Feldkamp@Colorado.EDU

Follow this and additional works at: https://scholar.colorado.edu/mkey_gradetds
Part of the Applied Statistics Commons, Music Pedagogy Commons, and the Software Engineering Commons

Recommended Citation
Feldkamp, Max, "Creating a Better Technological Piano Practice Aid with Knowledge Tracing" (2018). Keyboard Graduate Theses & Dissertations. 3.
https://scholar.colorado.edu/mkey_gradetds/3

This Thesis is brought to you for free and open access by Keyboard at CU Scholar. It has been accepted for inclusion in Keyboard Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Creating a Better Technological Piano Practice Aid with Knowledge Tracing

by

Max Joseph Feldkamp

B.A./B.M., Lawrence University, 2014

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Master of Music

College of Music

2018
This thesis entitled:  
Creating a Better Technological Piano Practice Aid with Knowledge Tracing  
written by Max Joseph Feldkamp  
has been approved for the College of Music  

Alejandro Cremaschi  

Andrew Cooperstock  

Alexandra Nguyen  

Date 5/9/2018  

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
Abstract

Feldkamp, Max Joseph (M.M., Music)

Creating a Better Technological Piano Practice Aid with Knowledge Tracing

Thesis directed by Associate Professor Alejandro Cremaschi

Modern music tutoring software and mobile instructional applications have great potential to help students practice at home effectively. They can offer extensive feedback on what the student is getting right and wrong and have adopted a gamified design with levels, badges, and other game-like elements to help gain wider appeal among students. Despite their advantages for motivating students and creating a safe practice environment, no current music instruction software demonstrates any knowledge about a student’s level of mastery. This can lead to awkward pedagogy and user frustration. Applying Bayesian Knowledge Tracing to tutoring systems provides an ideal way to track and predict student knowledge and skills. This thesis explains how to utilize Bayesian Knowledge Tracing in music practice software and discusses the benefits over existing pedagogical software.
# Table of Contents

Abstract ........................................................................................................................................ iii

List of Figures .............................................................................................................................. vi

List of Tables ............................................................................................................................... vii

Introduction ................................................................................................................................... 1

Objective and Scope ...................................................................................................................... 2

Literature Review ........................................................................................................................ 3

Practicing and Effecting Change ................................................................................................. 3

  Theoretical Framework and Research on Practicing and Learning ............................................ 4

  Rehearsal Frames: A Model for Effecting Change .................................................................. 9

  Practice Strategies for Improving Performance Ability .......................................................... 12

Motivation and Flow ...................................................................................................................... 14

State of Practice Software and Apps .......................................................................................... 19

  Gamification and Cognitive Feedback .................................................................................... 19

  Analyzing the Apps .................................................................................................................. 22

  Space for Improvement ............................................................................................................. 29

Bayesian Knowledge Tracing ...................................................................................................... 29

  BKT Equations .......................................................................................................................... 35

Methodology ................................................................................................................................ 36

  Bayesian Knowledge Tracing in Music .................................................................................... 36

  Encoding Music for Bayesian Knowledge Tracing ............................................................... 37

  Reactive Context and Goal Picking ......................................................................................... 39

  Assumptions .............................................................................................................................. 41

Demonstration ............................................................................................................................... 42

  Exercise Example: G Major Scale ......................................................................................... 42

  Musical Example: Horse-Drawn Carriage .............................................................................. 50

Combining Cognitive Feedback .................................................................................................... 53

Quantifying Practice Strategy Effectiveness ............................................................................... 54

Discussion .................................................................................................................................... 56

Validity .......................................................................................................................................... 56

Consequences for Motivation and Achievement ........................................................................ 56
Extensibility and Using Big Data ................................................................. 58
Human Pedagogical Applications ................................................................. 60
Confidence in Student Repeatability ............................................................... 60
Knowledge Versus Skill ............................................................................... 61
Further Study ................................................................................................. 63
Conclusion ..................................................................................................... 64
References ..................................................................................................... 67
List of Figures

Figure 1 – Different psychological states based on skill level and challenge level............. 16

Figure 2 – A three-star score in Piano Maestro ................................................................. 23

Figure 3 – Piano Practice with Wolfie .............................................................................. 24

Figure 4 – Smart Music demonstrated trouble hearing dense textures and harsh grading. ........................................................................................................................................ 26

Figure 5 – Annotation of Skills Involved in Playing GM Ascending in the Right Hand ..... 42

Figure 6 – Sample exercise to contextualize an F Sharp ................................................... 43

Figure 7 – A test for checking the thumb under/turn skill ................................................. 44

Figure 8 – The probability of mastery P(Lt) and probability of correctly answering the next prompt P(Ct) at each trial t for the skill of finding sharps correctly................................. 49

Figure 9 – “Horse-Drawn Carriage” from Piano Adventures by Faber. Reprinted under Fair Use. ........................................................................................................................................... 50

Figure 10 – The P(Lt) and P(Ct) at each trial t for playing staccato correctly. The rolling average is the mean of the last five trials................................................................. 52

Figure 11 – The P(Lt) and P(Ct) at each trial t for playing staccato correctly. The rolling average is the mean of the last five trials................................................................. 62
List of Tables

Table 1 – Rules for task picking and completion ................................................................. 40
Table 2 – Initial probability of mastery for finding and playing sharps .............................. 45
Table 3 – Initial probability of mastery for thumb tuck tasks ............................................. 45
Table 4 – Initial probability of mastery for playing G Major ................................................ 46
Table 5 – BKT Parameters for finding sharps ..................................................................... 47
Table 6 – Questions and student answers for finding sharps .............................................. 47
Table 7 – BKT Parameters for playing staccato ................................................................. 51
Table 8 – Student attempts at playing staccato ................................................................. 52
Table 9 – BKT parameters for finding F sharp ................................................................. 61
Table 10 – Student trials for finding F sharp ................................................................ .... 62
Introduction

While teachers play an essential role in setting standards and developing discrimination in their pupils, they cannot be present during every practice session and cannot provide the continuous, moment-to-moment feedback that determines the microstructure of practice with its constant starts, stops, and repetitions (Chaffin & Lemieux, 2004, p. 27)

The quality and quantity of student practice makes up the largest factors of achievement and is also one of the factors that teachers have the most difficulty controlling. Students frequently have trouble identifying problems, but when they do, they may not know how to, or want to, address their problem spots. Even students who can correctly identify correct practice techniques and patterns frequently do not actually apply the techniques correctly (Christensen, 2010). In recent years, applications for tablets and phones, called apps, and computer programs have started to provide a technological solution to some of these problems by providing instant feedback and basic goal structures.

Following the trend of gamification by incorporating badges, achievements, avatars, and other elements typically found in games, apps can provide powerful extrinsic motivation, but also present a tight cycle of trial and feedback. Internal motivation is a key ingredient for continued success, but for those just starting to practice regularly external motivation is often required before success is internalized (Chaffin & Lemieux, 2004, p. 31; Christensen, 2010). These apps, through effective use of game elements, are able to induce “flow” in users, a state of total absorption in the task at hand, even creating a loss of self-awareness (Csikszentmihalyi, 2004). Researchers consider flow to be an intrinsically rewarding experience, which further
makes apps an excellent tool to help students experience both intrinsic and extrinsic motivation.

The technology used by modern practice apps for listening and grading is extremely impressive and continues to evolve rapidly. However, no app has a system designed to help students effectively practice moment-to-moment, beyond giving right-or-wrong feedback. Other educational domains have started adopting various techniques to create intelligent tutoring systems, which are designed to monitor, predict, and guide student learning. A commonly used technique is Bayesian Knowledge Tracing (BKT), which tracks student knowledge and ability to apply concepts in a probabilistic fashion. An app could then use this additional information to give personalized assistance and guide the student to work on concepts that challenge them most. Further, relatively recent research has found that a technique called Cognitive Feedback has been successful in helping musicians improve the expressive quality of their playing (Arrais & Rodrigues, 2007). This thesis proposes adapting Bayesian Knowledge Tracing to determine when a student is on or off task, which practice strategies are helpful for them, and to give a student feedback on which skills or sections of pieces are least reliable via a Cognitive Feedback system.

Objective and Scope

The primary goal of this thesis is to identify major components of effective practicing behaviors, and break these skills down into the smallest possible units for application in BKT. The second goal is to illustrate how BKT can promote progress on a single piece, allowing us to identify problem spots easily, and encourage the student to focus their efforts there. As
such, we will be primarily focused on how to create change on the scale of a single practice session and how to train students to do the same. Therefore, we will not look at long-range progression concerns like repertoire sequencing and piece selection, though we will comment on how information from a system like BKT might support such concerns as well. There are a number of skills that are essential to development of practicing and musical ability that cannot be measured, at least easily, by a computer, and thus fall outside the scope of this thesis. Primarily, mental practice is a whole class of techniques considered quite effective by many teachers and researchers but is rather difficult to monitor (Jørgensen, 2004, p. 92). Also, monitoring physical technique is outside the bounds of current technological state of the art and is thus not addressed.

**Literature Review**

**Practicing and Effecting Change**

Students need to learn to study and practice effectively and independently, but many have not yet learned to do so (Duke R. A., 2005, p. 62).

In order to better examine the efficacy of practice apps, we must first examine the theory and application of practice and how expert teachers effect change in student performances. We will also address the concept of “rehearsal frames,” a pedagogical construct used in the analysis of instruction leading to short-term change within the framework of a music lesson. Common themes and predominant ideas will later be used to take a fine-grained look at the moment-to-moment interactions in apps and understand how they do or do not align with good teaching principles.
Theoretical Framework and Research on Practicing and Learning

Barry, Nancie and Hallam call systematic rehearsal the “repeated performance or systematic exercise for the purpose of learning or acquiring proficiency” (Practice, 2002, p. 151). Another characterization of practice is the “work needed to produce improvement.” (Chaffin & Lemieux, 2004, p. 23). When working towards a performance, the intent of practice is to enable complex physical, cognitive, and musical skills to be performed fluently with relatively little conscious control (ibid., p. 155). Lehmann, Sloboda, and Woody liken the activity of practice to how athletes train, with the stated goals of lowering performance variability, and to exceed one’s prior limits (2007). The act of practicing in this way results in physical, physiological, and psychological changes relating to the instrument. Another way to consider the process is the creation of “procedural memories” (Duke & Bruckner, 2009, p. 18). In this process, the practicer has a chance to get feedback on her or his playing and discover gaps between intent and reality, and then work to resolve those discrepancies.

The learning process may be in two phases: an introduction phase and a refinement phase. A new concept or technique is typically introduced by the teacher in the weeks ahead of a piece that uses the new idea (Jacobson, 2006, p. 26). This involves incremental preparation and introduction in isolation of the piece, before a systematic introduction to the piece starting with rhythms, moving to the intervals and so forth. This mode of operation most commonly typifies teaching young learners, but also appears at high levels of instruction, for example with technical execution. However, once introduced, a piece typically enters the follow-through and refinement phase (Jacobson, 2006, p. 30), which is principally the interest of most research, and of this paper. This phase is what is traditionally considered as practice.
Practice is absolutely essential to developing a broad and well-rounded set of musical skills, with multiple pieces of literature giving suggestions as to how much time is needed: some say 10 years of focused learning, while another says 10,000 hours are requisite for becoming expert (Chaffin & Lemieux, 2004, p. 20). The concept of ‘raw talent’ is apparently dubious at best, and the same authors tell us that “good evidence of a genetic basis for intellectual ability is still lacking.” (ibid., p. 19). Another debatable point in the literature is the degree to which the number of hours contribute to mastery. While cumulative lifetime hours do predict one’s overall expertise, quantity doesn’t seem to be very significant to achievement of short-term goals (Lehmann, Sloboda, & Woody, 2007, p. 152; Duke, Simmons, & Cash, 2009; Oare, 2011). Correspondingly, total hours spent preparing for a particular event don’t predict performance quality (Barry & Hallam, 2002, p. 152). Chaffin and Lemieux add that high achievement requires that practice time is well spent, not just that one spends a lot of time practicing (General perspectives on achieving musical excellence, 2004, p. 20).

There is good evidence for better predictors of achievement. One particularly interesting realization comes from a study by McPherson and Zimmerman, who say that achievement during the first 9 months of music lessons was determined less by the amount of practice than by commitment to the instrument (Chaffin & Lemieux, 2004, p. 23; see Motivation and Flow for more). Bonneville-Roussy and Bouffard have created a particularly interesting framework that integrates practice time with self-regulation, goal-orientation, and use of practice strategies that has successfully improved upon prior models of achievement (2015, p. 697). This body of research shows there are a number of effective practicing behaviors that are more relevant to improving performance quality. Chaffin and Lemieux give a
useful, if tentative, picture of five fundamental characteristics of practicing: Concentration, goal-setting, self-evaluation, strategies, and the big picture (2004, p. 24). Practice strategies will receive its own section, but we discuss the other elements here.

**Concentration**

Lehmann, Sloboda, and Woody break apart effective practicing into two broad categories: Macro-scale environmental influences and micro-scale patterns of activities in practice sessions (2007). The “macro-scale” influences are important, systemic and procedural concerns for establishing concentration. Being such a mentally focused task, many external factors can impact the quality of a practice session, ranging from wakefulness to bodily needs like food. Unfortunately, performers’ willingness to share their feelings toward practice and their daily habits varies widely, and they often report loving or hating practice in equal proportions. One common theme among musicians is seeking solitude to practice, and that practice can often include ancillary activities like reading, listening, and analyzing music. Even time of day can have an impact, and Sloboda et al. found that students who reported practicing scales in the morning, while the mind is fresh, were more accomplished than those who did so in the evening (1996; Chaffin & Lemieux, 2004, p. 24).

Mindless practice or practicing without listening is seen as a sort of daydreaming while playing, and something to be avoided (Chaffin & Lemieux, 2004, p. 24). Instead, there should be continuous and urgent activity in a continual cycle of pausing to think, immediate adjustment, and trial. With that level of exertion, the duration of practice also has an impact on both the type and quality of learning accomplished. Most often, shorter sessions are more effective, especially for simpler tasks (Barry & Hallam, 2002, p. 153; Chaffin & Lemieux, 2004, p.
Longer practice sessions are generally only more productive for more experienced and more motivated musicians but are limited by mental and physical fatigue. Generally, distributed practice is probably best for building skills, but massed practice can be helpful for skilled individuals preparing for a specific event. Barry and Hallam further corroborate this, and they say teachers recommend setting specific practice goals for each practice and having two or more short sessions (2002, p. 157). Jørgensen elaborates further, recommending that practice sessions are long enough for improvement without fatigue, and close enough together to minimize forgetting progress (2004, p. 25).

**Goal-Setting and Self-Evaluation**

Metacognition, or self-awareness (self-evaluation) and self-guidance (goal-setting), is identified as central to effective practicing (Barry & Hallam, 2002, p. 154) Novices typically have a great deal of difficulty correctly identifying a specific error or error-prone section, and often work to correct it by playing through the piece, whereas an expert is acutely aware of when, where and why errors occur. Chaffin and Lemieux echo this, saying self-evaluation and self-directed exploration may be sufficient for those who are adequately skilled, but students will require the help of a teacher to provide the necessary feedback, and to suggest productive strategies. In particular, the ability to accurately identify via critical self-listening, rehearse, and correct the source of every error is a crucial ability (Oare, 2011, p. 46; Duke, Simmons, & Cash, 2009). Some specific courses of action are recommended for cultivating this skill throughout the literature. For example, taking notes on desired outcomes, tips, or writing in reminders of these issues can help support effective practicing (ibid.). Teachers should help students quickly diagnose and problem-solve repeat issues, and further to help them access the core reason for
their mistake: Was the music misread or misunderstood, or was the mistake not heard (Jacobson, 2006, pp. 212-213)? Not allowing mis-learned pitches or rhythms to become habit is key to quicker learning, which supports the impact that metacognitive awareness can have (Oare, 2011, p. 42).

The expert pattern of practicing can be broken into work and runs (Chaffin & Lemieux, 2004, p. 27). Work is where a passage is isolated and repeated several times, with the goal of improving some of its aspects. Runs is the process of re-integrating the passage into the piece and evaluating progress towards the intended goal. Frequent assessments provide opportunities to obtain valuable information to both the student and the teacher (in the case of lessons) and to improve the performance of the skill (Duke, 2005, p. 84). The ability to slow a piece down and work it in sections is a characteristic of skilled practicers (Barry & Hallam, 2002 157; Duke, Simmons, & Cash, 2009). This requires prior analysis and typically marking the music before playing. Sections for practice should be created according to the structure of the music, as structurally guided practice has been shown to be highly effective. The sooner such practice is begun, the better (Chaffin & Lemieux, 2004, p. 22). Students are urged to sectionalize their practice, because it limits the number of problems that need to be dealt with and eases the burden of learning a passage (ibid., p. 27). During slow, sectional practice, the student must imagine each upcoming note and all its qualities to achieve progress (Lehmann, Sloboda, & Woody, 2007, Chaffin & Lemieux, 2004, p. 23).

Teachers and students should focus on difficult spots, and should aim to reduce broad goals to small, manageable tasks in a logical sequence (Jacobson, 2006, p. 214). Goals may be specifically technical, or expressive in nature in order to achieve an end effect (Jørgensen, 2004,
Goals set by teachers should be as specific as possible, involving concrete or physical instructions (Duke, 2005, p. 36). In addition, teachers need to relate goals to the quality of the performance such as tone, intonation, rhythmic precision, and expressiveness (ibid., p. 26).

The Big Picture

Forming the “artistic image” of a piece early on is a central tenet of effective practicing (Chaffin, Imreh, Lemieux, & Chen, 2003; Chaffin & Lemieux, 2004, p. 28). Not only is this an essential behavior, but also makes music-making more engaging (Duke, Simmons, & Cash, 2009). This begins at the first reading of a piece with something of a rough sketch, and eventually turns into a performance plan. This musical understanding informs the creation of sections, as described above, and enlightens the structure of practice as we have already seen. Sections can be chosen based on formal structure, technical difficulty, visual layout, or even harmonic progressions (Jørgensen, 2004, p. 93). Later in the process of refining a piece, the “big picture” starts to override the sections and is used to create fluent transitions between these sections (Chaffin & Lemieux, 2004, p. 29). However, the sooner an understanding of the musical structure of a piece, the quicker a student will grasp the piece, and the resulting performance will be significantly better because of the early development of that musical understanding.

Rehearsal Frames: A Model for Effecting Change

Rehearsal frames are tools that facilitate the analysis of teacher and student behavior and interaction during a lesson. Introduced in 1994 by Robert Duke (Duke, 2005), they are defined as “periods of concentrated attention and effort that are clearly directed toward the
accomplishment of specific proximal (near-term) goals.” Each period is composed of repeated cycles of trial and instruction, each allowing for an opportunity to assess how close the student is to accomplishing the goal at hand. Each frame is dedicated to a single objective and encompasses everything the student and teacher work on to get to that goal. The purpose of this analysis is to examine how and why incremental change happens in a lesson. Paired with studying how master teachers work with students, we can create one relatively accurate picture of how immediate change in performance quality can be achieved. One admitted issue with this perspective is that it fails to capture long-term learning and skill acquisition, but that would fall outside the purview of modern apps already.

A rehearsal frame begins at the identification of a goal or task. The frame ends when the goal is attained, when a new one is defined, or simply when the stated goal is no longer being pursued. Student and teacher work toward the attainment of the proximal goal through feedback, modeling and repeated trials. The course of the music and student errors dictate where these frames happen in the music (Duke & Simmons, 2006, p. 13). A rehearsal frame can either be successful or unsuccessful in achieving its objective, or the frame might be left unevaluated from a lack of student trials. One thing that is clear is that effective teachers leave far fewer frames unevaluated. Unevaluated frames are problematic because there is no evidence as to whether or not the student has acquired an understanding of the change to be made.

The use of this tool in pedagogical research suggests that frequent and succinct student trials are necessary to create change (Duke & Simmons, 2006, p. 13; Duke & Bruckner, 2009, p. 21-22). In fact, teaching effectiveness increases when student play-time makes up the majority
of the rehearsal frame, and when the student is immediately stopped to correct technical and
note errors. The caveat to this approach is that, after several trials, most teachers would stop
and move on to avoid frustration and exhaustion, giving the student ‘intuitively timed’ breaks
(Duke & Simmons, 2006, p. 13).

The type of feedback and modelling employed by the teacher can also determine the
success of a frame. Effective teachers utilize negative feedback that is frequent, direct, and
highly specific (Duke & Simmons, 2006, p. 15). Positive feedback, on the other hand, is given
considerably more infrequently, but is typically much more extensive. It is most commonly
shared after the student achieves a proximal goal (Duke & Chapman, 2011, p. 37).

Other types of critically important interactions include directives, modeling, questions
and information. (Duke & Bruckner, 2009, p. 24). Directives are defined as anticipatory
information, sometimes given mid-course, to remind students of an upcoming problem spot, or
an instruction of how to address a technical problem. Modeling is a particularly valuable kind
of directive, giving the student an aural model for a successful trial (Duke & Simmons, 2006, p.
15). Questions are the least frequent form of teacher-student interaction but can be very
effective for redirecting the internal thinking of more expert students (ibid.; Duke & Chapman,
2011, p. 36). Information, given somewhat more frequently than questions, is commentary on
general ideas to focus on, or sometimes background information to help guide stylistic or
technical work.

There are other effective teaching practices this analysis brings to light. For example,
high quality modeling is a key ability for teachers, and it should be employed relatively
frequently to provide targets to students (Duke & Simmons, 2006, p. 15). Further, teachers
retain and employ a clear mental image of how the piece should sound and use that to guide the lesson. Also, teachers maintain knowledge of student’s prior work, and use this to both plan proximal goals, but also as a point of comparison to give feedback on items that have improved or worsened over many lessons.

Rehearsal frames provide a useful way to examine app functioning. If apps can create environments with a specific, immediate, and attainable goal, then we can measure if the software provides useful feedback, directives, modeling, questions, or information in the manner a teacher would. Apps that successfully provide rehearsal frames, work and runs, goal-setting, self-evaluation, and concentration are more likely to promote better practice, and thus higher rates of short-term goal achievement.

**Practice Strategies for Improving Performance Ability**

Advanced students and professional musicians often hold idiosyncratic views as to what constitutes effective practice (Jørgensen, 2004, p. 85). Indeed, “no longitudinal data on how different styles of practice lead to better performance are available” – it seems that professional musicians can get equally good with a wide variety of approaches to practicing (Barry & Hallam, 2002, p. 152). It is therefore important to know what strategies are available, and which ones work well for one’s self. When practice strategies are taken out of context, authors often disagree on their effectiveness; for example, Barry and Hallam state that most teachers often recommend beginning a piece with slow practice, then gradually increasing practice tempo (2002, p. 157; Duke, Simmons & Cash, 2009). On the other hand, Jørgensen points out that in some situations slow practice first can be detrimental to building a musical
image of the piece, and full tempo practice will be more effective for developing the necessary physical and musical motions (2004, p. 94).

Jørgensen breaks strategic practice into three phases: Forethought, performance/volitional control, and self-reflection (2004, p. 85). In most ways, the self-reflection aspect has already been discussed in terms of metacognition and self-evaluation. Forethought includes goal-setting, but also selection of approaches on how to achieve a change in performance. This involves careful decision making based on what’s worked in the past, but also what hasn’t been tried recently. Routine can often be helpful but should be broken when it no longer produces the expected gains. For example: Warm-up exercises at the beginning of practice are very common, and address the needs of the particular instrument, but they may not suit the needs of every practice session of every performer.

Practice strategies often surround error-correction (undesirable expressive effects are likely considered errors as well). To reduce complexity and improve the attainability of a passage, different strategies can be categorized as a limited performance, a modified performance, a parallel passage, or a related exercise (Duke & Bruckner, 2009, p. 166). A limited performance is one that simply focuses in on a single note, two notes or small section in isolation. Modified performances go further, with warped rhythms, dynamics, tempi, or other features that fundamentally alter the music with the intent of facilitating performance. A parallel passage is a passage elsewhere in the music that is similar but may be more attainable. Finally, a related exercise eschews the original music altogether for an exercise that tries to build up the requisite ability to achieve the passage. After going through any of these
processes, re-contextualization is key to correctly integrate the worked passage back into the big picture.

Regardless, the most important feature of practice strategies is that musicians *think* about them. Most teachers report always or almost always emphasizing practice, but 40% of students entering conservatory report that their teachers put “very little” or “no” emphasis on practice behavior (Jørgensen, 2004, p. 98). This simple fact highlights the necessity that software not only includes options for strategic practice, but make the process of planning, self-evaluations, and trial a key feature. We intend to examine which strategies are available in each app, and the degree to which they are a central element of the instruction, and what features, if any, exist that encourage students to think about the strategies they are using to achieve their goals.

**Motivation and Flow**

In the preceding section, motivation is clearly an important element of developing skill in practicing, and helping students use better practice strategies. Motivation is so important, particularly early on in the learning process, that commitment to and interest in the instrument has been found to be a better predictor of achievement than the amount of time spent practicing (Chaffin & Lemieux, 2004, p. 23). This section will attempt to briefly unpack the current state of research on motivation from a very high-level perspective. Currently, there are five well-developed areas of research concerning motivation: Extrinsic versus Intrinsic motivation, Goal Orientation theory, Expectancy-Value theory, Self-conception/Self-efficacy, and Attribution theory. In addition, flow theory, which explains one specific process that leads
to intrinsic motivation, is particularly relevant in informing the mechanics of apps. Gamification, feedback systems, and a risk-free environment to improve motivation are key elements to successful practicing as well.

Motivation is typically defined as either being internal or intrinsic, or external or extrinsic. Extrinsic motivation is any external factor that give a person a reason to do something: Parental expectations, money, points, or the means to achieve another goal. Specific external motivators can fall into one of three categories: (1) external regulation, where an individual seeks reward or avoid punishment, (2) introjected regulation where an individual seeks social acceptance or to avoid guilt, or (3) identified regulation where the activity is not engaging, but it achieves a goal that is valuable for another purpose (Austin, Renwick, & McPherson, 2006, p. 225). Apps frequently attempt to increase external motivating factors through gamified design, a topic to be discussed later (see Gamification and Cognitive Feedback).

Intrinsic motivation, on the other hand, occurs when an activity is inherently rewarding or pleasurable. It is both more stable, and a good predictor of long-term skill acquisition (Chaffin & Lemieux, 2004, p. 30). Achievement, skill, feelings of ability and the ‘rage to master’ are major contributors to intrinsic motivation. Another key is flow, which is described as an inherently pleasurable and motivating experience (Csikszentmihalyi, 2004; O’Neill & McPherson, 2002, p. 35-36; Austin, Renwick, & McPherson, 2006, p. 217; Chaffin & Lemieux, 2004, p. 31). Conversely, external motivators such as too much parental supervision can impede development of intrinsic motivation, and growth overall.
Flow is a mental state where a person is completely and totally immersed in the task at hand. This state, something many people call being ‘in the zone,’ is associated with someone being so involved and engaged that they are unaware of their surroundings and the passage of time. Flow requires no distractions, a clear goal, immediate feedback, and a match between skill and task difficulty. The most basic requirement is limited outside distractions, since flow is easily interrupted or prevented if a person’s attention is forced away from the immediate goal, for example by another person or phone. The second requirement for flow, a clear task or goal, sets up the framework in which the feedback occurs. Feedback is required to be immediate, so that a student has knowledge of what is working and what needs to get fixed at all times.

The final ingredient to flow is equal matching of task difficulty and ability, which should be set so an activity is engaging without being overwhelming. If the task is easy and skill is low, apathy or boredom is the most likely outcome. Increased difficulty without the requisite skill results in anxiety, whereas the reverse situation might be relaxing but not particularly engaging. Flow requires that a person’s skill level ‘rises to the challenge.’ Learning software is potentially well-suited to encourage flow by potentially modulating difficulty based on the perceived skill of the student.

Challenge also comes into play in Goal Orientation theory, which describes a student as mastery-oriented, performance-oriented, or helpless. A student who is mastery-oriented is
driven by the challenge of learning and learns for the sake of learning. A performance-oriented student will seek simply to ‘get the grade,’ and do better than their peers, or at least not underperform (Austin, Renwick, & McPherson, 2006, p. 213). Where a mastery-oriented student will typically seek challenge and accept failure as a teacher, a performance-oriented student will seek easy tasks simply to do well and avoid failure as much as possible (O’Neill & McPherson, 2002, p. 38). Not only do mastery-oriented students tend to stick with a difficult task more readily, but they also tend to be more successful practicers. Children who exhibit a helpless orientation were found to practice twice as much to reach the same level of achievement as mastery-oriented children (Chaffin & Lemieux, 2004, p. 22). Apps that offer an environment where it is “safe to fail” can help shift a student’s outlook to be more mastery-oriented.

Expectancy-Value theory also deals with tenacity and engagement, but by looking at the social and personal values of the learner. Four factors weigh on the expectancy-value. Attainment value describes how important it is for a student to have success in a given subject. Intrinsic value is the innate appeal of a subject to the student. Finally, Utility value is the usefulness of a skill. These three values are weighed against the perceived cost of obtaining the skill, which could be measured in time lost pursuing other hobbies or perhaps a social cost (Austin, Renwick, & McPherson, 2006, p. 226). Low expectancy-value is understood as leading to attrition (O’Neill & McPherson, 2002, p. 32). While it is difficult for apps to directly affect the user’s Expectancy and Value, they can lessen the perceived difficulty of tasks through appropriate goal setting and decontextualization. Additionally, they can make the music more appealing with clever visual design and backing accompaniment tracks.
Success can strongly affect internal motivation and increase Expectancy-Value. One’s *Self-concept*, or the idea of how good one is at a subject, is a very important and self-reinforcing aspect of motivation (Austin, Renwick, & McPherson, 2006, p. 222). A related concept is the conception of skill as either mutable or a fixed trait of one’s being. A particularly damaging self-view is where one believes themselves both unskilled and incapable of growth. However, the far more powerful element of the research regarding self-concept is *Self-efficacy*, the belief in one’s future ability to achieve, or the capacity for attainment. High Self-efficacy is a particularly strong predictor of students’ long-term achievement (ibid., p. 223; O’Neill & McPherson, 2002, p. 34; Chaffin & Lemieux, 2004, p. 30). Apps that show a student’s progress over time may help improve a self-concept, and hopefully help demonstrate growth potential.

Attribution theory focuses on the student’s view of how successes and failures come about, and how these impact motivations. Attribution runs on two axes: Locus and stability (O’Neill & McPherson, 2002, p. 36-37; Austin, Renwick, & McPherson, 2006, p. 227-228; Chaffin & Lemieux, 2004, p. 30). A student may attribute success and failure to either internal causes or external causes, and the cause may either be stable or unstable. For example, an unstable and external attribution would be luck, but a stable, external attribution would be the ease of the task or teacher aid. Someone with an internal, unstable attribution believes outcomes are due to factors like effort. Alternatively, a stable attribution would be to consider success or failure because of one’s skill. Perhaps unsurprisingly, external attributions are largely unhelpful, but more surprisingly attributing failure to lack of effort is not particularly helpful (Austin, Renwick, & McPherson, 2006, p. 228). Instead, focusing on developing better strategies for next time appears to be more productive, and in general an internal, stable
 attribution is most beneficial. Music software has the potential to help learners focus on developing better practice strategies. Games, for example, do not call a player ‘just bad,’ but instead invite experimentation with strategies for growth, and learning software takes advantage of this more liberating mentality.

To sum up: The most enduring and effective motivation occurs when a student (1) is self-motivated rather than externally, (2) enjoys the learning process, (3) values the skill being learned, (4) believes in his or her potential, and (5) attributes success to controllable, internal factors. Motivation can, and typically does, come from external sources, but it should be quickly internalized through success and flow experiences. In the next section, we will explore how software exploits multiple aspects of motivation theories to involve students and encourage them to keep going.

**State of Practice Software and Apps**

This section takes a high-level look at several major modern practice apps, how they use feedback and motivation, and how they guide students through their practice session. We will also discuss some historic uses of technology in music, and its efficacy. Modern practice does not always help achievement in the way teachers and students might hope, but some design models are currently striving to change that.

**Gamification and Cognitive Feedback**

Gamification is a recent trend in educational apps, and particularly in music apps. The move towards gamified design has been inspired by the dramatic rise in popularity of video games in recent decades. Gamification can mean the introduction of any number of different
game-inspired modes of interaction, such as badges, experience points, levels, avatars, and achievements with the goal of harvesting the flow-inducing features of immediate feedback to create motivating cycles of expertise – i.e. boosting self-concept. While some of the research body surrounding gaming suffers from poorly defined terminology and objectives, inappropriate methodology and contradictory results (Fels & Seaborn, 2015), some of the findings are still worthwhile to mention.

Birch (2013) found that a game designed to improve student musical technique had a positive effect on the number of technical elements students mastered, and on their attitude toward practicing them, while self-efficacy levels were not affected. Despite the fact that this particular study had a relatively small sample size, it is encouraging to see that the use of game design can be an effective way to encourage learners to engage in an otherwise tedious task. Another, much larger and more recent study suggested that badges could increase user activity and participation in a system (Hamari, 2017). Additionally, a fairly comprehensive review of the literature found that in a majority of cases both in and out of music, gamification has resulted in positive effects and benefits in the areas of engagement and enjoyment (Hamari, Koivisto, & Harria, 2014).

On the other hand, there are also several examples of gamification studies with negative results. One of them found that, while students scored better overall and on practical assignments, they did poorly on written activities and participated less in class (Domínguez, et al., 2013). The researchers did, however, discover that there was an initial boost to student motivation. Another study found that students in a gamified course showed less motivation,
satisfaction, empowerment, and lower final exam scores over time than those in the non-gamified class (Fox & Hanus, 2015).

This somewhat muddy picture of the impact of gamified education is echoed in research in technology in music. For many years, the problems of studying the interaction between technology and educational outcomes have been largely the same: the ever-changing nature of technology combined with small samples and other experimental problems confound most findings (Webster & Hickey, 2006, p. 386). Despite these shortcomings, the research suggests that educational technology can at least improve attitudes and engagement in the activity at hand, which can be a useful tool for tedious task like practicing.

Another software design paradigm called **cognitive feedback** has shown much more definitive capacity to help musicians (Juslin, Friberg, Schoonderwaldt, & Karlsson, 2004; Arrais & Rodrigues, 2007). The research around cognitive feedback has involved software that provides information on what type of emotion a performance of a segment of music is most likely to evoke in a normative audience. Additionally, the software gives advice to adjust aspects of the performance like tempo, dynamic, timbre, articulation, and several other parameters to achieve a desired emotive impact. The study found that the system was quite effective in comparison to an unassisted practice, and even compared favorably against teacher assistance. While an interesting result by itself, this suggest that the pattern of simply giving concrete, directly actionable instructions for progressing towards a user-stated goal can be highly effective. With this in mind, we will now examine how apps apply the principles of gamification and cognitive feedback, along with other aspects effective practice we have discussed above.
Analyzing the Apps

There are several modern apps available that try to encourage practice and improve practice behaviors. This thesis will look at Piano Marvel from the company of the same name, Piano Maestro by JoyTunes, Piano Practice with Wolfie by Tonara, Percebe Music by Percebe Music, and SmartMusic by MakeMusic as prime examples of practice apps that give feedback and guidance. Each exhibit particular strengths and weaknesses in their ability to help students learn, but more importantly all share three common motivational features that make them appeal to users: Increased extrinsic rewards like badging and levels, immediate feedback (which is beneficial for flow), and decreased cost of failure.

Apps are primed to provide two major flow conditions: Immediate feedback and the matching of challenge to skill. They all give some form of note-by-note right or wrong feedback, helping students to hear and focus on problem spots. Additionally, they all use a large repertoire of materials that have been graded into a number of levels, providing natural progression and accessibility to students of many skill levels by giving them the right material at the right time. Secondly, each app leverages gamified systems in an attempt to offer more external motivators and opportunities to recognize achievements. Finally, they all offer unlimited low-risk, judgement-free time to rehearse, promoting a mastery-oriented attitude and reducing the fear of failure. However, we will also see that these apps still lack in moment-to-moment reactivity and goal adaptation required for the successful tight feedback loops described in rehearsal frames, flow theory, and practice literature as a whole.

The way the practice apps in question have been designed to include external motivators is fairly apparent from their user interface. Both Piano Marvel and Piano Maestro
give a zero to three-star rating on a student’s performance as well as a more refined percentage score. They are both organized into levels similar to modern piano methods, but much more fine-grained than their print counterparts. *Piano Maestro* includes the feature to “unlock” upcoming levels as the user completes the easier ones, thus providing a powerful intrinsic goal, and a strong incentive to try more advanced levels as soon as they are unlocked. They both offer ways to ‘test into’ an appropriate level of difficulty. *Piano Maestro* gives each student an ‘avatar,’ or in-game persona, that levels up in technique and reading categories as they prove themselves on harder and harder pieces. *Piano Practice with Wolfie* awards badges for spending a certain amount of time practicing a piece, and for completing it. All these techniques work well for providing external motivation to a variety of students.

All of these apps provide immediate performance feedback, usually in the form of an immediate right-or-wrong mark on each note played. Each app seems to have a different capacity for nuance in their grading. *Piano Marvel* requires a MIDI keyboard plugged into a computer to do its grading. It compares the student performance to a ‘correct’ performance, presumably as recorded by *Piano Marvel*’s pedagogical personnel. The student plays along with a backing track to help keep the time and pulse. Grading evaluates each note as either right or wrong: Correct rhythm with incorrect notes and correct notes with incorrect rhythm both result in a zero grade. It does not take into account note duration or dynamics, even though velocity
(dynamic information) is available through MIDI. This results in a somewhat harsh and basic grading system, without much ‘intuition’ as to what underlying issues and problems.

Piano Maestro, an iPad app capable of “listening” to an acoustic piano, features a more sophisticated approach to assessment. The app is capable of tolerating background noise and various piano timbres, and it is tolerant enough to interpret vocal singing as correct playing. Like Piano Marvel, Piano Maestro provides an accompaniment track. The grading feature can understand rhythm and pitch separately: It is possible to play all the right notes, but in poor time and receive a one hundred percent pitch score but zero percent rhythm score. However, the primary flaw seems that it is more willing to forgive lateness than earliness. This does provide an advantage in that it allows the student a chance to correct a misread note, and improve their score, but this also means that if a note is miscounted and the next played too early (a common problem for young learners), it may be scored incorrectly. A positive side effect of this is that ‘extra’ notes are graded fairly leniently. Piano Maestro does not grade note duration/articulation or dynamics either, and also struggles with thick chords and textures due to the limitations of its acoustic listening feature.
Unlike the apps above, *Piano Practice with Wolfie* does not provide immediate feedback, but instead gives a grade after listening to the entire performance. It is able to listen to an acoustic piano the same way as *Piano Maestro*. The app distinguishes beat problems from rhythm problems and grades them separately. It also gives a grade for “flow,” which is a measure of backtracking or skipping. The biggest issue is that the feedback comes in the form of letter grades, with little specific information about problematic sections of the piece, or how to improve them.

*Percebe Music*, an app still “in progress,” is perhaps the most promising software of the group. At the time of writing, *Percebe*’s website appears to be having technical difficulty, and the app seems non-functional. That said, the features demonstrated on video by the author gives an excellent vision of where the app succeeds. *Percebe* uses a MIDI connection to listen to the performance and uses the most information provided by the digital instrument, including dynamics. The most unique feature of the app is that, while it has a metronome, it highly discourages the use of it. It dynamically tracks tempo fluctuation (rubato), and judges it using an unspecified ‘artificial intelligence’ algorithm against verified performances. It also evaluates not just dynamics, but also voicing, and marks notes in an accompaniment pattern when they stick out over the melody line. It is capable of picking up exactly where the user starts playing without any further instruction and gives a grade as soon as the user stops playing. Grading is lettered, with an “uh oh” replacing F.
Similar to *Piano Marvel*, *Smart Music* is a web-based system, but unlike the former, it features music for a multitude of different instruments. The variety and amount of music available for practice is quite impressive. Once in a piece, the experience is akin to *Piano Marvel*, with a metronome click track and accompaniment track playing along as it scores. It is capable of listening to live playing over the microphone, much like *Piano Maestro*, and it seems to share the limitations to texture density. The grading, particularly related to timing and rhythm, seems to be quite severe. It allows very little leeway, and, to make matters worse, expects the user to adjust to unmarked tempo fluctuations. For example, the accompaniment and metronome contained an unmarked ritardando in the last measure of the first line in the figure above, which resulted in this particular performance getting graded wrong in that measure. These difficulties somewhat lessen the value of the feedback in this particular case since it is frequently difficult to tell if a problem arises from the performance or the software. This may not be the case for other instruments, but seems to be the case at least for piano. Of all the systems reviewed, *Smart Music* seems to be the least gamified: It lists each of the trials.
you have attempted, and the percent you achieved on that trial. Since it seems to be geared
toward adult musicians, this is probably an intentional move to make the software seem more professional and attractive to that audience.

The apps offer benefits to practicing other than motivation and feedback; they also offer better practice techniques in the form of unique modes of interaction that are designed to increase the rate of learning. *Piano Marvel* has a feature called “Practice Mode.” The mode provides the piece subdivided into three levels: “Minced,” “Chopped,” and “Whole” – each of which is given at slow, medium and full tempi. For example, the “Minced” level might break the piece down into two or four measure chunks to facilitate learning. The biggest downside to this particular feature is that students need to be told how to find and use this feature by an instructor or peer, and they need to be driven enough and willing to work on smaller sections at a slower tempo to seek the feature out. *Piano Marvel* have also recently introduced a “Prepare Mode,” which waits for the correct note to be played before it allows the music to move on and gives the red/green markings for incorrect and correct notes played.

*Piano Maestro* includes a “Learn” mode that breaks apart the hands and slows down the tempo to aid learning. It also includes a “Hold it” mode that pauses the music on every note until the user plays the correct note. This mode is particularly helpful for learning the notes of a piece and encouraging pitch reading. The largest downfall is that a clever student could potentially play every single relevant note to bypass it and get the song to continue. However, the idea is strong, and *Piano Maestro* automatically takes a student to the next part of the learning process when one part has been mastered, making it somewhat better in that regard to establishing and following through on proximal goals.
Piano Practice with Wolfie has the greatest variety of features intended to help practice. Perhaps the most interesting feature, if not extremely innovative, is the ability to draw and write on the score and take notes in this fashion. Also, the app is capable of recording your performance and playing it back to you, which is useful for cultivating self-listening and critical thinking. The app also links to various web resources such as recordings and other materials relevant to the piece, giving a more musically comprehensive picture of how the piece should sound and providing an important modeling aspect for the app.

No particular features aimed at practice were demonstrated in the videos of Percebe, but there were a number of highly valuable features. The ability to start anywhere and only practice a small portion of the music lines up well with research regarding sectioned work. Also, the strong focus on musical playing will not only be motivating for the student but is good for viewing pieces from an expressive perspective. The visual feedback about tempo control and rhythmic accuracy also align with cognitive feedback principles.

Smart Music has a couple of features that are useful for learning musicians to help make the most of practice time. A particularly interesting one was the inclusion of a ‘loop’ mode, which allows a practicer to highlight a section of music, and have continuous, looped evaluation of that section in isolation. This is a feature shared with Piano Marvel, and it is probably the single most intriguing tool of any of the apps reviewed, but it still relies on the student’s self-awareness and planning to use the tool. Another useful feature is the ability to play back a performance. Given the difficulty Smart Music has grading, this might be a helpful tool for students to confirm errors, to self-assess their dynamics, tone, and timing in a more
realistic way. The tool doesn’t have any support for critiquing dynamics or deviations from timing.

**Space for Improvement**

*What teaching machines then and now lack is the capacity to accurately assess students’ knowledge, skill, and attitudes, and to incorporate this information into the ongoing process of instruction in the moment, modifying instruction tasks in ways that lead learners to successful conclusions* (Duke, 2005, p. 56).

The quote from Duke nicely summarizes the most apparent unaddressed space for improvement in modern apps: A lack of reactivity to a student’s needs and current ability. In all the apps, students are assumed to seek out and think about successful strategies for improvement based on the feedback they are given, which is known to be untrue. What is needed is a system that can understand where the student is at. Then, it should help the student identify proximal goals, and guide the student to build a series of steps that takes them to success in a dynamic, real-time manner. In short, a better digital tutor will help students create rehearsal frames and teach the student how to self-guide through the process. In the following section, we discuss a method for assessing the knowledge and skill of a student so that we can apply that to just such a system.

**Bayesian Knowledge Tracing**

Timely and relevant feedback are key elements to providing quality education, but the matter of achieving that is complicated, and often involves a great deal of domain knowledge: In this case, piano pedagogy. Even the most sophisticated systems are predominantly human-machine hybrid systems, meaning the role a teacher plays in the development of students is
expected to remain as important and essential as ever. However, algorithms have advanced a long way in recent years: For example, computer algorithms have been found to grade the writing portion of the SAT test as accurately as human graders – a rather surprising and tremendously helpful tool (Klein, 2008). In particular, that demonstration of machines successfully grading student responses in a difficult domain (for machines) is promising for music. One particularly promising field of research is in providing learners with the best material at the ideal time, so that they are focusing on material and concepts that are most difficult for them, and not spending time on concepts they already know. Such systems are broadly called Intelligent Tutoring Systems, and this is the type of system that is a central focus of this thesis.

One particularly common technique used in production systems today for achieving such a task is Bayesian Knowledge Tracing (BKT). BKT is a member of the Bayesian family of machine learning techniques that resulted in technology like spam filters and predictive spell checking (it’s how your iPhone knows which words might come next) and inherits the ease-of-use and good performance characteristics shared by the family. These traits, along with its ability to leverage a deep domain knowledge, and its popularity and frequency in literature make it a good choice for a first application of machine learning to a new domain. As of writing, the author is not aware of any application of BKT or similar techniques to music – proposed or trialed. The most common application of knowledge tracing in general appears to be mathematics, where questions can be answered quickly and are purely right or wrong.

BKT is dependent on having a set of production rules, or focused steps to achieve an answer, or in the case of music, achieve improvement. An example could be that a student
needs to practice the two notes of a large leap out of rhythm before expanding their practice to be in rhythm, and then moving to the surrounding context. An alternative way to represent a production rule is as a *buggy rule*, or as a common mistake that a student might make. In the case of the large leap, perhaps overshooting or undershooting the destination pitch, or reaching the correct pitch but out of rhythm would be excellent examples of common mistakes students make. The tricky part of these rules is that it can be difficult to ascertain the root cause of a mistake, or that there may be multiple production rules that are active all at once. Typically, the most likely cause is selected, but with applying some additional knowledge regarding practice, it might be possible to pick target practice techniques and sections more easily. The selection algorithm paired with the production rules is called a *domain model*, and this is the essential focus of this thesis – applying pedagogical knowledge in a computational environment to enrich the capability of apps. A domain model may be considered as a formalized computer version of the knowledge, procedures and skills relevant to piano practice.

The most characteristic piece of BKT, though, is how it treats the *student model*. The student model is how BKT is capable of tracking the student’s acquisition of skills and knowledge over time. An application can track the student’s ability to recognize and apply production rules over time to build an understanding of their skill level. Even in normal lessons, a teacher must come to understand the student’s ability level and knowledge and track progression from week to week. BKT allows us to chart a student’s knowledge based on their answers to questions or performance of a piece of music (in our case) by employing a “Hidden Markov Model,” which attempts to ascertain if the student understands a concept through examination. The technique treats the student’s true ability as a latent, or hidden, variable –
this is much like how a teacher cannot know precisely what a student’s level is but may build up
an understanding over time with careful evaluation. For example, Duke recognizes this when
he writes “what students know is invisible to teachers – until we express that knowledge in
some way, by speaking, writing, or otherwise acting in a way that indicates what we know.”
(2005, p. 34)

Bayesian Knowledge Tracing uses four probabilities for its student model. The first
value called the probability of mastery, \( P(L) \), is the likelihood that a student has already
learned a skill; this value updates each time a skill is used and assessed. The second, fixed value
is the chance that a student could achieve a correct answer or performance by chance or
guessing: The \textit{probability of guess} \( P(G) \). Sometimes this value is straightforward, for example,
when there is a one in seven chance of correctly guessing the name of a white note, but other
times this value may be more difficult to ascertain. The third, also fixed value is how likely it is
that a student’s mistake is the result of a slip-up, rather than a failure of understanding: The
\textit{probability of slip} \( P(S) \). The last value, \textit{probability of transit} \( P(T) \), describes the likelihood of
the student having learned a skill in the process of solving the problem or working a
passage. These values can be taken together to form a reasonable estimate of a student’s
mastery using a set of equations. Typically, the probability of transit and the \textit{initial} probability
is per-skill, but the model may also be individualized, so that these values are both per-skill and
per-individual: This individualization can make a significant improvement to the efficacy of the
technique (Yudelson, Koedinger, & Gordon, 2013).

The model works by updating the probability of student skill mastery after each
trial. If the student answers question correctly then there are two possibilities: (1) The student
already understands the skill and they didn’t slip, or (2) the student doesn’t already understand, and they guessed. If the student makes a mistake there are also two possibilities: (1) The student doesn’t have mastery of the skill and they didn’t guess correctly, or (2) the student does know how to do the skill and simply slipped. Regardless, the two possibilities are combined to arrive at a new probability of mastery prior to the current trial. This probability is then combined with the possibility that the student didn’t have mastery prior to the test but learned the skill in the process of the attempt. This new combined value becomes the updated probability of student skill mastery. Further, we can estimate how likely the student is to get the next question right by looking at (1) the probability that the student has mastery of the skill and doesn’t slip, combined with (2) the probability that the student hasn’t learned the skill and guesses correctly. To see how these rules are implemented in math, see BKT

**Equations.** Further examples of the BKT process are given in the demonstration section.

A further addition to these two pieces is the concept of presenting material best suited for the student given their current assessed skill level and adjust the ordering and timing of new skills to match. In the computing and machine learning space, this is called adaptive sequencing. This can involve not only monitoring a single student, but also looking at the broader population of students, and working to identify common pitfalls by looking at where students pause before answering or attempting a solution or identifying other key metrics. Fortunately, as piano teachers we have a large body of experience and literature on how to pre-select many of these pitfalls and issues, and intelligently sequencing material and music is a topic discussed at great length in literature and schools.
BKT does have some drawbacks, chiefly that it treats all variables as purely binary. Specifically, it makes the assumption that an answer can be categorized as either entirely correct or completely incorrect. Further, it assumes that a student has either totally learned a concept or not at all. These issues may present some difficulties in the realm of music. We will discuss some naïve approaches in an attempt to solve these issues, but the key idea is to break down concepts as much as possible to the smallest possible units. In relation to music, Duke tells us it is highly unlikely that a student will know what teacher means by telling a student to “sing more expressively,” and the skill will need to be broken down into its component pieces to be effectively taught (2005, p. 36). Such a task can sometimes involve something called *Cognitive Task Analysis*, where domain experts sit with learners and have students talk through their thought process while solving problems. Secondly BKT, like any machine learning approach, requires a sufficiently large dataset of student data to train on to generate meaningful values and settings, such as the probability of slip for a given skill. We hope to expedite this process and create representative values for these properties based on the available literature, and intuition where we lack evidence for a reasonable value.

BKT is a comparatively understandable technique, and has a number of available extensions, making it a sensible choice to apply to music. Further, it can leverage a substantial body of expert knowledge in music, making it easier to apply. Other techniques, like *Deep Knowledge Tracing*, could also be used for a musical intelligent tutoring system that have advantages in certain areas. For example, the algorithm does not require any special foreknowledge of the structure concepts and sub-concepts in the way BKT does, but it does need vastly more data than BKT in order to ‘learn’ the structure of knowledge. Even a
moderately successful application of BKT to piano instruction would imply that the domain can be broken down into machine-understandable concepts that can be tracked and trained against, and then that success can be built upon further by bringing more data and more flexible algorithms to bear.

**BKT Equations**

The probability that a student has learned a particular skill is given by

\[ P(\text{learned}_{\text{now}}) = P(\text{learned}) + P(\text{transit}) \times (1 - P(\text{learned})) \]

Where \( P(\text{learned}) \) depends on whether a question was correctly answered, or passage correctly played. If the answer is correct, then:

\[ P(\text{learned}) = \frac{P(\text{learned}_{\text{prior}}) \times (1 - P(\text{slip}))}{P(\text{learned}_{\text{prior}}) \times (1 - P(\text{slip})) + (1 - P(\text{learned}_{\text{prior}})) \times P(\text{guess})} \]

On an incorrect answer or performance, then:

\[ P(\text{learned}) = \frac{P(\text{learned}_{\text{prior}}) \times P(\text{slip})}{P(\text{learned}_{\text{prior}}) \times P(\text{slip}) + (1 - P(\text{learned}_{\text{prior}})) \times (1 - P(\text{guess}))} \]

The probability of the student’s next trial being correct is given by:

\[ P(\text{correct}_{\text{next}}) = P(\text{learned}) \times (1 - P(\text{slip})) + (1 - P(\text{learned})) \times P(\text{guess}) \]
Methodology

Bayesian Knowledge Tracing in Music

To date, there is no evidence of Bayesian Knowledge Tracing (BKT) being used in educational music software. However, the idea of using Hidden Markov Models in music is not entirely new: There has been substantial success using the technique for score-following (Pardo & Birmingham, 2005). This suggests that the class of techniques is generally suitable for application to music. The major problem facing the application of BKT to music education is that there is no large, systematically collected set of data on which to train a machine learning technique. Further, even if such data existed, there is no immediately apparent method to assess what constitutes a skill for the purpose of BKT. Here we give a demonstration of how to create activities useful for BKT and illustrate how BKT can be useful despite the lack of data.

In order to demonstrate the potential for BKT to improve educational music software, we will bypass the ordinary data collection and model-fitting steps, and manually supply reasonable parameter values to drive the technique forward. We will make assumptions as to general features of a skill as easy or difficult to master, and about a student’s prior contact with particular skills. For example, we will assume a slightly higher initial probability of having learned a skill, since we expect any tutoring software to be used in conjunction with typical private lessons. The primary interest here is to illustrate how BKT will guide student practice through different challenges (hard versus easy, familiar versus foreign, etc.). With that intent, it should be less important if we characterize skills totally correctly, and more important that we illustrate how the workflow handles different classes of skills.
BKT can easily be demonstrated in spreadsheet software, such as Microsoft Excel, when the parameter values are already known. We will show how exercises and pieces can be encoded to be divided into discrete technical, musical, reading, and other skills to drive BKT. Coupled with a simple heuristic to adjust task difficulty and set goals, our demonstration will show how a student struggling with two or three different skills will be led through more effective practice behaviors to quicker success. This will show how we can create an environment that may be better suited to creating flow experiences, lowering the perceived risk and cost of failure, increasing self-efficacy, and improving practice efficiency.

**Encoding Music for Bayesian Knowledge Tracing**

The fundamental requirement for BKT is that all skills must be evaluated with exclusively ‘correct’ or ‘incorrect’ trials – there is no grey area. We assert that every exercise and piece can be broken down into a set of constituent skills that can be trained with simplified activities. Thus, a student who has correctly learned each constituent skill correctly should be able to play the exercise in a manner that is correct. In some cases, it might be obvious when something is correct: For example, if an F sharp is played instead of a G sharp, there is no doubt that the trial was incorrect. Some skills might require extra care to define correctness in a way that is usable for BKT. The most problematic example skill is likely timing. Percebe demonstrates a good example of how to monitor and track timing, and the research by Arrais and Juslin further demonstrates ways to accurately listen to timing aspects of a performance. In Percebe, a chart reports back the real-time tempo at each moment of the piece. Such a system makes it possible to detect inappropriate fluctuations in tempo, as well as a performance that doesn’t have any timing flexibility.
In terms of critique, some sort of probabilistic threshold will need to be adopted. For example, we could propose a system that grades a performance as correct if it predicts that 95% of teachers would accept the timing in the performance. This sort of setup would allow flexibility for pieces that require a great deal of timing flexibility as well as those that are extremely precise. The system will have to apply a ‘mostly right’ threshold to timing, dynamics, articulation, tone quality, and generally any activity where a line has to be drawn at good enough. To apply these evaluation techniques to a system using BKT, there needs to be a way to decompose pieces and exercises into activities or sections that train the sub-skills necessary to complete the piece successfully.

It is abundantly clear from literature that sections of pieces should be bounded by musical and technical considerations. Further, literature indicates that tasks should be created as concretely as to make evaluation clearer, which works well for the binary nature of BKT (Duke, 2005, p. 36; Jacobson, 2006, p. 214). This makes it clear that each activity and piece of music needs to be decomposed into every physical movement and every mental step. We give two examples of this sort of analysis here: A G Major Scale, and “Horse-Drawn Carriage” by Faber from level 2B of Piano Adventures (Hal Leonard). In particular, the “Horse-Drawn Carriage” is a valuable illustration, since it makes heavy use of the G Major scale. This provides evidence that skills can live inside larger contexts, reducing the burden of laboriously analyzing each and every single note of every piece by reusing prior analysis.

The primary chosen aspects in these examples are visual, rhythmic, physical, dynamic, and articulation elements. We attempt to annotate as many challenges of each variety in both examples as possible, so that the system has as many smaller skills to track as possible. In this
way, there is as much resolution as possible when attempting to ascertain the root cause of an issue. For example, we must ascertain that a note is read and understood correctly before we can attempt to determine if a physical issue is at fault. The physical aspect is particularly challenging, since there is no way to be certain the pianist is playing with the right fingering or with good technique generally. We will rely on ruling out other causes of mis-playing to infer a physical cause of an issue, and will tackle these types of physical causes by providing some creative exercises (see the G Major example). Otherwise, the technological capacity to record and evaluate rhythm, reading, dynamics, timing, and virtually every other musical facet have been demonstrated by modern software and research. This is essential to taking the next step toward developing an intelligent tutoring system for music practice.

**Reactive Context and Goal Picking**

One consistent theme found in both the literature on practice and BKT is that a large concept (or piece) needs to be broken down into its constituent elements. For even slightly sophisticated pieces, it will be useful to avoid the need for practicing each pair of notes in isolation or clapping each measure of the rhythm in isolation. For the sake of maintaining motivation, interest and flow, it will be necessary very early on to change context to sections that need the most work. However, even the most intelligent system (or teacher, for that matter) is not likely to always set the context correctly, immediately, and with perfect

\[ \text{Creative use of technological advances may allow for better capture of such issues: For example, perhaps there is something to be gained by using speech recognition and asking a student to recite their fingering or their counting while they are playing to gain additional information.} \]
consistency. We can utilize BKT to help the system realize when it has set the context too small (i.e. \( P(\text{learned}) \) is already very high, or becomes very high quickly), but more importantly we can utilize BKT guidance to realize when the context is too large. If we notice multiple consecutive drops in \( P(\text{learned}) \), we can use that observation as an indication that the scoping too is broad.

We define some basic rules of operation to help define fundamental behavior of the system – choosing what to work on first, and how to know when to switch to an easier or harder context.

**Table 1 – Rules for task picking and completion**

<table>
<thead>
<tr>
<th>Guiding Principle</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pick task</td>
<td>Zone of Proximal Development: Choose a skill close to mastery</td>
</tr>
<tr>
<td></td>
<td>Choose skill with highest ( P(L) ), such that ( P(L) &lt; 0.95 )</td>
</tr>
<tr>
<td>Switch to easier task</td>
<td>Too many incorrect trials indicate a task difficulty too high</td>
</tr>
<tr>
<td></td>
<td>Switch when percentage of trials correct is lower than probability of guessing for 5 consecutive trials</td>
</tr>
<tr>
<td>Task complete: Pick next task</td>
<td>We need ample evidence to know the student has acquired the skill</td>
</tr>
<tr>
<td></td>
<td>Finish when ( P(L) &gt; 0.95 ) for two trials or more</td>
</tr>
</tbody>
</table>

Of these three tasks, ‘switching to an easier task’ is the most complicated and poorly defined, since the rule must additionally decide *which* task to switch to. In theory, when a skill has several finer-grain tasks, the student will have demonstrated mastery on each sub-skill, so identifying the root cause can be difficult. Consider a situation where a piece has a tricky chord change coupled with a syncopated rhythm. If a young student who has proven to be adept with both rhythmic challenges and reading chords demonstrates difficulty with the combination of the two in the form of an awkward delay, one might ask whether the trouble is because they
are struggling to feel the rhythm, or because the chord change has not been properly learned. As much as possible, the system will use locality and temporality (i.e. where the issue is and when it is) to ascertain the root issue. Beyond that, the system should focus on more fundamental skills first. For example, rhythm will be isolated first in the example above, since the combined skill cannot happen at all without rhythmic correctness. Beyond that, the combination of the two tasks may still present difficulty, which is where alternate contextualized activities are helpful – such as a version with a substantially reduced tempo, which can be useful to simplify the task and aid mastery.

**Assumptions**

For the sake of our demonstration, we will assume that we have an accurate listening system that is capable of listening to as many notes simultaneously as necessary, and has a conception of rhythmic flex, much like *Percebe Music* has. In real life, this is probably achieved with a MIDI listening software package, coupled with some sort of intelligent score-tracking technology (e.g., as in Birmingham and Pardo). Additionally, we assume that we have access to the ability to reliably grade musical aspects of performance such as articulation, timing, and dynamics with an acceptability thresholding technique as described earlier. Further, we will make some further assumptions that our imaginary student has already demonstrated mastery of the prerequisite skills to start work on the skills we present here. Thus, we do not build every skill from the ground up; rather we leave the process of generating exercises that test each individual skill as an exercise for the reader.
Demonstration

Exercise Example: G Major Scale

A one-octave ascending G major scale, right hand alone at a moderate steady tempo will be our first demonstration. The primary skills our imaginary student will be guided through are the physical motion of the thumb under finger three, and accurately finding F# on the keyboard. We will assume that the student has already learned certain fundamental skills such as how to execute a G Major five-finger pattern and has demonstrated good reading.

Encoding

![Diagram of G Major Scale](image)

*Figure 5 – Annotation of Skills Involved in Playing GM Ascending in the Right Hand*

Here, we will consider an ascending G Major scale produced by the right hand as having three fundamental motions that guide our annotation of the skill: The initial three steps (filled third), the thumb under finger three (followed by the hand going over the thumb), and the stepwise filled fifth (See Fig. 5). Given there are only quarter notes in this example, we consider no particular rhythmic skills. Visually and physically, there is additionally the skill of finding and playing the F sharp (F#) consistently. We present no articulation or dynamics here, but an easy
extension of this exercise would be to layer those considerations on top of the exercise. However, this clearly illustrates how even relatively mundane exercises can be broken down into very specific skills, certainly too many to try to apply BKT to the entire task. Instead, we must develop tasks that work on the development of each sub-skill. Compared to a five-finger pattern exercise, which we assume has already been learned, there are only two issues novel to the exercise: The thumb turn, and the F sharp. We will develop exercises to ascertain whether each skill has been learned or not.

The F Sharp

![Figure 6 – Sample exercise to contextualize an F Sharp](image)

The simplest way of testing if the student can recognize and play an F# is to simply display an F# and ask the student to play the note they see. This particular test is unfortunately quite useless since it is testing for knowledge rather than a skill (See Skills versus Knowledge). A more useful initial activity is to train the skill for identification of sharps in general. However, putting that knowledge to use in a scale will be a significantly more difficult task, since it requires the student to prepare her or his fourth finger to play the F sharp. Fig. 6 represents a minimal context that we believe provides reasonable evidence that a student has mastered playing F sharp in context.
Thumb Under/Turn Tests

As stated earlier, one of the more difficult tasks when only electronic listening is available, is to be certain that correct fingering and technique are being used. This particular exercise is designed so that there is no easy way to successfully execute the task without performing with the correct fingering. However, even here, there are four sub-skills: Playing a harmonic third, turning the thumb under, and a harmonic fifth are relatively obvious, but we must remember the skill of moving the hand over the thumb is implicit here. Having the student play the upper G gives some level of guarantee that this motion is occurring correctly. We can further break down this task into two sub-tasks: the first three notes of the task (G-B-C), and the last three notes (B-C-G). These two sub-tasks illustrate how to separate the thumb tuck from the movement of the hand over the thumb, but first we must necessarily assume that the player is using the correct fingering with no further information available. Eventually, however, the availability of all these different modes of assessment should guide the student to success.
Initial State

The BKT process requires information about the likelihood of the student already knowing how to perform the tasks being asked of him or her. The initial $P(L)$ parameter represents this likelihood, and we break it into the various groups as we describe above.

Recall that the probability of two events occurring simultaneously is the product of the probability of each event happening in isolation.\(^2\) For finding sharps in general, and then applying that skill in playing, assume the following initial probabilities of mastery:

\[\text{Table 2 – Initial probability of mastery for finding and playing sharps}\]

<table>
<thead>
<tr>
<th>Skill (name)</th>
<th>Initial $P(L)$</th>
<th>Reasoning/Student History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find various #s out of context (p-find-#)</td>
<td>0.25</td>
<td>There has been no prior work on this skill. Give the student a 1-in-4 chance they already know how to (consistently) find any sharp on the keyboard from lessons.</td>
</tr>
<tr>
<td>Play F# in context (p-play-f#)</td>
<td>$0.2 \times p\text{-find-}$ # = 0.05</td>
<td>There has been no prior work on this skill. A 1-in-20 chance (0.05) they'll play the F# correctly, and they can find an F# at all.</td>
</tr>
</tbody>
</table>

For the ability to execute the thumb turn, we assume that the student can already play C major, so assume:

\[\text{Table 3 – Initial probability of mastery for thumb tuck tasks}\]

<table>
<thead>
<tr>
<th>Skill (name)</th>
<th>Initial $P(L)$</th>
<th>Reasoning/Student History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thumb Tuck (p-tuck)</td>
<td>0.98</td>
<td>The student has demonstrated the skill in C Major</td>
</tr>
<tr>
<td>Hand Over Thumb (p-over)</td>
<td>0.9995</td>
<td>The student has demonstrated the skill in C Major</td>
</tr>
<tr>
<td>Tuck and over in G, i.e. fig. 3</td>
<td>p-tuck * p-over * 0.2</td>
<td>There has been no prior work on this skill.</td>
</tr>
</tbody>
</table>

\(^2\) E.g., there is a 0.5 probability of getting heads on a coin flip once, but $0.5 \times 0.5 = 0.25$ probability of getting two in a row.
Finally, we give the cumulative skill the following initial probability of mastery:

Table 4 – Initial probability of mastery for playing G Major

<table>
<thead>
<tr>
<th>Skill (name)</th>
<th>Initial $P(L)$</th>
<th>Reasoning/Student History</th>
</tr>
</thead>
<tbody>
<tr>
<td>G Major Scale, in steady tempo</td>
<td>$p$-tuck-over * $p$-play-f# * 0.1 =0.00098</td>
<td>There has been no prior work on this skill. Just shy of a 1-in-1,000 chance the student knows how to do this, given our model of their other knowledge.</td>
</tr>
</tbody>
</table>

As stated in the methodology, these numbers are entirely fabricated, but are structured here to represent mostly relative likelihoods that the student already knows how to execute the skills. Some of the values are exaggerated for the purposes of our illustration: In reality, we would expect somewhat higher probability of success coming from a private lesson. To continue our illustration, we now pick the skill with the highest $P(L0)$ less than 0.95. The student has already demonstrated mastery of some of the basic physical aspects in learning C Major (i.e. $P(L0) > 0.95$ for those skills) but has never demonstrated being able to find sharps or play them in context. The skill with the highest probability of already having been learned is finding sharps out of context, so that is the first skill to be worked on.

**Sharp Finder Activity**

The first activity simply involves displaying a random sharp note on the staff and asking the student to play the note they see on the keyboard. Much like a flash card, the correct answer is revealed after the trial, so that there’s an opportunity for feedback and adjustment. Each piece of trial feedback should take advantage of giving information or a directive, much
like Duke discusses for his rehearsal frames. Consider the following BKT parameters for the activity:

Table 5 – BKT Parameters for finding sharps

<table>
<thead>
<tr>
<th>Parameter</th>
<th>p(L0)</th>
<th>p(Transit)</th>
<th>p(Slip)</th>
<th>p(Guess)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; p &lt; 1</td>
<td>0.25</td>
<td>0.15</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The initial probability of having learned the skill is given above as $p(L0) = 0.25$. The likelihood of figuring out how to consistently identify a sharp on any given trial, (also known as the transition to mastery probability), is 0.15. The number given here can be interpreted as meaning the skill is not very difficult to master, but not trivial either. We give a very low probability of slip, primarily because the only real source of slip will simply be inattention to the marking. The probability of the student being able to guess any note is 1 in 12 (0.08), and the probability of guessing a black key is 1 in 5.

With these parameters, the student is given the following challenges, and responds with these answers:

Table 6 – Questions and student answers for finding sharps

<table>
<thead>
<tr>
<th>Attempt</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenge</td>
<td>F#</td>
<td>G#</td>
<td>C#</td>
<td>A#</td>
<td>B#</td>
<td>D#</td>
<td>F#</td>
<td>A#</td>
<td>G#</td>
<td>E#</td>
</tr>
<tr>
<td>Answer</td>
<td>F</td>
<td>F#</td>
<td>C#</td>
<td>A#</td>
<td>A#</td>
<td>C#</td>
<td>F#</td>
<td>A#</td>
<td>G#</td>
<td>E</td>
</tr>
<tr>
<td>Right?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Attempt</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>F#</td>
<td>G#</td>
<td>C#</td>
<td>E#</td>
<td>B#</td>
<td>D#</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Answer</td>
<td>F#</td>
<td>G#</td>
<td>C#</td>
<td>E#</td>
<td>B#</td>
<td>D#</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After 16 questions and answers, the student finally completes two trials where the probability of mastery is above 0.95 (see plot in figure 8 below). In the answers above, the student initially struggles to remember what exactly sharps do, and after some comparison with supplied feedback starts to get on the right track, but struggles with white-note sharps, and still occasionally confuses sharps for flats. In the plot below, $P(L_t)$ is the probability of mastery after $t$ attempts, and $P(C_t)$ is the probability that the student will correctly answer on their next attempt. These probability values have been calculated based on the challenges and answers from table 6.
Figure 8 – The probability of mastery $P(L_t)$ and probability of correctly answering the next prompt $P(C_t)$ at each trial $t$ for the skill of finding sharps correctly.

**Flow Onward**

After completing the sharp finder activity (i.e. once the probability of mastery is high after attempt 15 in the plot above), the system will choose the next activity: Playing the F sharp in context as we described earlier. Once mastery is demonstrated with this new activity, the system will skip training for the sub-tasks of a “thumb turn” and “hand over” in isolation, since mastery of those skills has already been demonstrated (see assumptions above). However, it will move on to try the combined activities as shown in Table 4. Once completed, the student can finally work on the complete G major scale. At any stage above, if the student starts demonstrating difficulty, the software has the option to back-track. For example, if the student is not playing the F sharp once re-introduced in the full scale, she could easily be taken back to
the skill of playing the F sharp in context as a ‘reminder’ activity. Once the overarching activity is complete, the student is free to move on to other exercises or pieces, for example the piece “Horse-Drawn Carriage,” addressed in the next section.

**Musical Example: Horse-Drawn Carriage**

**Encoding**

![Figure 9 – “Horse-Drawn Carriage” from Piano Adventures by Faber. Reprinted under Fair Use.](image)

Here we illustrate one possible way to break down and encode a real piece with multiple overlapping technical, cognitive, and musical complications for use with BKT. Each element is chosen as an item that could be trialed in isolation, or at least within a greatly reduced context. There are too many skills in this example to enumerate possible activities for each one, but simplified activities can easily be created by reducing to a single measure of a single hand, or a couple of notes hands together. Also important here is that complex skills like the G major scale in the second measure can be coded as a single item if they have already been broken into manageable components, as demonstrated in the prior section.
Process

Instead of populating a complete initial state, we will create a student that is in the middle of practicing. The student has been demonstrating troubles with transitioning to staccato articulation in measure four when playing the entire excerpt in context and has already failed five trials. This triggers the rule to reduce context. Since there is additional locational information about the error, the algorithm can set the context to the right hand alone in the third and fourth measures of the piece, as this has the best chance of success. Suppose that in this case, the student is struggling to reliably execute a short enough staccato sound, and the underlying issue is a fundamentally technical one.

Take the following parameters for the BKT skill:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$p(L_0)$</th>
<th>$p(\text{Transit})$</th>
<th>$p(\text{Slip})$</th>
<th>$p(\text{Guess})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt; p &lt; 1$</td>
<td>0.1</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Playing good staccato is undoubtedly a much harder skill to master than finding sharps regularly, so we present a lower probability of initial mastery, and a much lower probability of learning the skill on any given trial. However, the chances of slipping up and not getting it quite right are also higher, while guessing is somewhat less of a factor. Digital evaluation of staccato is a considerably more difficult matter. Using MIDI, the basic measure is taken as the duration of a note in order to determine if a note was played staccato or not. Even so, the system needs to be able to answer the question, “how short is short enough?” For this, we rely on the acceptability thresholding technique to judge each trial as either right or wrong, as described in the methodology:
Table 8 – Student attempts at playing staccato.

<table>
<thead>
<tr>
<th>Attempt</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Attempt</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Right?</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Figure 10 – The P(Lt) and P(Ct) at each trial t for playing staccato correctly. The rolling average is the mean of the last five trials.

In Figure 10, a plot of the average success rate over the last five trials (the rolling average plot) is added, in addition to the plots of the probability of mastery and the probability of a correct trial on the next attempt. This additional plot illustrates that, despite the difficulty the student demonstrates with the skill at first, she or he is still always doing better than simply guessing, allowing the activity to continue. Particularly note trial thirteen in the plot, where the
student has had three consecutive correct trials (trials 11, 12, 13 in the table), and the moving average hits 80 percent. Despite what seem like indicators of success, BKT only rates the student as having slightly better than a one-in-two probability of having mastered the task. Indeed, the student then fails a handful of more trials before going on to truly demonstrate success.

**Flow Onward**

After demonstrating isolated mastery of the staccato skill, the student would be led to re-include the left hand, which complicates matters. Once reliable mastery is shown there, the student can finally move back to a fully contextualized performance. After that point, the student might demonstrate difficulty with another skill, on which the system will focus next. It is important, however, to successfully reintegrate the skill just learned (staccato) into the overall context first. Having examined how this process can work in practice, we now examine some possible extensions of the core idea, and the consequences of using a guided practice tool like this one.

**Combining Cognitive Feedback**

One particularly prominent feature for effective practicing in the literature is the ability to self-assess and self-guide practice. In our examples above, we have simply forced the student to do what our set of rules deemed best for them at that given time. However, a better design would be to allow the learner to choose their own path through practicing the piece. One interesting and important piece of information reported by BKT is a prediction of how likely the next trial is to be successful. This could be relayed back to the user as a type of
reliability metric, and then allow the user to self-guide through sections of music that haven’t been learned as thoroughly, or perhaps are more prone to slip in the first place. In addition to the reliability of a piece, the estimated level of mastery could be reported back to the student or reported on request so that the student could test their level of self-understanding. This way, there can be feedback not only about which parts of a piece are learned or not, but also about the student’s ability to find these sections, allowing for a powerful secondary level of practice skills enhancement beyond the immediate activity being addressed.

**Quantifying Practice Strategy Effectiveness**

In the literature, there is extremely little information on what specific practice strategies are effective, or what kinds of situations they should be used in, or which kinds of learning styles and level of learners benefit the most from them. In fact, we saw that a strategy that is particularly useful for one student may not be for another. An exciting output from a complete BKT analysis is the four-parameter probability set that permits evaluation of the characteristics of an activity or strategy. One of those parameters, the probability of transit, $P(T)$, turns out to be extremely useful for understanding how helpful a specific learning strategy is for students. $P(T)$ can be interpreted as how quickly students can learn the skill with the particular activity under test. Consider two different activities that train the same skill: One has a $P(T) = 0.05$, and the other has a $P(T) = 0.2$. That means that on any particular trial, a student using the second technique is nearly four times more likely to acquire the skill at each trial. In practice, a difference in $P(T)$ that large might mean the difference between four trials or sixteen trials before the student is likely to have reached mastery. If a teacher knew this about an activity,
that information would mean choosing the second activity to train the skill rather than the first, and a quicker learning curve for the student. Systematically applied, such a way of thinking might make entire classes of skills dramatically easier to learn. This is one potentially very interesting path or research that could be pursued using BKT.

Consider further the idea of an individualized $P(T)$. As mentioned, the literature has highlighted that practice strategies vary greatly in effectiveness from person to person, so imagine if it were possible to discover which strategies are most effective for a given skill, for a given person. This kind of information would have the power to dramatically transform the way teachers, human or digital, plan activities or even lessons, allowing them to personalize a curriculum to an individual's strengths and weakness. With access to even simple data collection, this information could be easily computed dynamically. A full-production app could easily track a whole range of different practice strategies and track the per-student efficacy of these strategies via the probability of transit parameter.

In fact, there is space to apply BKT even without the availability of an automatic grading system. If a student could be relied upon to correctly self-evaluate (perhaps the rare college student), then BKT could be used as a powerful practice aid to understand how quickly different activities bring students to mastery. The hardest element for a human to apply the concept would be identifying discrete practice activities, since human practice doesn’t usually follow clearly defined activities at all times. In this case, the correctness of each trial gets judged by a trusted individual, and after enough attempts and progress has been recorded, the efficacy of the activity could be reported back to the student. This is only one of several ways this type of
thinking can benefit musicians today. We will be exploring several more after taking a look at the validity and consequences of applying a knowledge tracing algorithm to music practice.

**Discussion**

**Validity**

There are several interesting characteristic traits associated with each of the BKT charts. When the student first begins a task, notice that the probability of the next trial being correct is always higher than the probability of skill mastery. This “beginner’s luck” gap is primarily due to the probability of guessing being the predominant factor in getting answers correct. At the end of the learning process, the probability of the next trial being correct is never as high as the probability of mastery. This “human error” gap is characterized by slippage being the predominant reason for incorrect trials. One particularly interesting feature of BKT is the crossover point (see Fig. 8, attempts 12 to 13; Fig. 10, attempt 13 and 19) where the probability of mastery is higher than the probability of getting the next trial correct. This point could be interpreted as the moment at which the student is more likely to get answers correct because of her or his mastery than because of a good guess. Whether or not these elements are reasonable in the activity, they are still good indicators that application of BKT to the skill makes sense. In the demonstration here, the flow through the activities guided by BKT fit well with what has been described in literature. Thus, even with contrived values, using a system like this may work very well to help students practice more effectively.

**Consequences for Motivation and Achievement**
There are several major advantages of BKT-driven practice sessions like the ones presented here, over traditional practice app workflow. The foremost mode of operation of all of the apps analyzed was to run pieces and exercise from start to finish, often with a merciless click track or perhaps a pleasant accompaniment to play with. While they may offer ways to work on a particular section, or reduce to a single hand, and most often offer reduced tempi, only Piano Maestro starts in this mode (referred to as their “Learn” mode). This is the first possible difference with BKT: Software driven by the BKT can start a student where they are most likely to succeed from the very first moment, because BKT gives the software an understanding of the student’s prior knowledge. Not only does this make good pedagogical sense (see any literature about goal-setting), this will ultimately mean that students have overall fewer failed trials, which should have a positive impact on intrinsic motivation due to success and higher self-efficacy. Fewer failed trials also will reduce the perceived risk failure during practice since there should simply be less of it.

A further benefit of being able to evaluate a student’s level of mastery is having a way to judge when to move to the next task or move back to a simpler task. The flexibility to do this is a key element of flow. Recall that presenting tasks that are neither too difficult nor too easy is one of the essential prerequisites for flow to occur. If an app simply moves on after a single correct trial, a student could easily find themselves quickly out of their depth, having not truly learned the essential skills to succeed at a higher level. Not only is this bad from a perspective of flow and motivation, it will not build good practice habits that develop confidence and consistency. Using BKT to guide progression and task context greatly enhances the capacity to control difficulty and encourage success and cultivate flow conditions.
None of the apps and software evaluated here seem to communicate a *true* sense of task mastery. They demonstrate a surface understanding of how correct a piece is, and most frequently give an “A,” three stars, or a badge on a single correct run of the piece and move on. This ignores the aspect of actually having deeply learned a piece. As any teacher knows, a piece demonstrated successfully once in a lesson can easily crumble the next time, or (worse) in a performance. Having a model of student ability is key to developing real mastery and reliability. The cycle of trial, evaluation, and assessment of goal achievement closely mimics Duke’s idea of rehearsal frames, and also echoes the idea of *work* and *runs* found in practice literature. The continuous cycle of repetition given by BKT with many opportunities for feedback, information, and directives to modify the next performance fits very well with the literature presented here, and is a novel capability not demonstrated elsewhere.

**Extensibility and Using Big Data**

There are many ways to extend the BKT-based tutoring algorithm by incorporating elements of other BKT research. The most immediate requirement to apply BKT to music instruction is data collection and fitting the data to the BKT model, preferably using the input of pedagogical experts. Secondly, there are additional parameters that could be added to the BKT formulation to make it more accurate for music, such as individualized BKT parameters of the probability of forgetting a skill.

The most challenging hurdle to actually applying BKT to music practice is the lack of trial data. In order to make such data available would be a fairly substantial undertaking, since there would be a need to encode trials of particular sections and activities, and most apps don’t focus on smaller levels of detail in that way. However, should sufficient data become available,
there are several different methods available to ascertain the Probability of Transit $P(T)$, Probability of Guess $P(G)$, Probability of Slip $P(S)$, and Initial Probability of Mastery $P(L0)$ parameters required to drive the type of system described here. One possible method is Expectation Maximization, which draws on modern optimization work to derive the parameters. This is the most commonly found approach to finding the BKT values in literature, but it does have some drawbacks (Hawkins, Heffernan and Baker). For example, there may be several different sets of parameters that fit the data well, making interpretation difficult. Equally problematic is the fact that the method can find degenerate sets of parameters, or parameters that make little pedagogical sense in context.

Another method, Empirical Probabilities, draws upon values that are set a priori, much as they are here to guide the machine learning process by attempting to estimate when the learner actually acquired mastery of the skill. The empirical probabilities technique, or one similar to it, could draw on estimated values like the ones given here to greatly improve the quality of the results, which validates this our approach despite the lack of a large data set.

In general, basing an intelligent tutoring system on BKT allows for enhancements based off of the technology. For example, recent research into individualization of parameters (e.g. Gordon, Koedinger, and Yudelson) could be used to create a powerful tutoring system that is able to track individual strengths and weaknesses, and report to the teacher what worked or did not for a specific student. Another fascinating extension of BKT that could find application in music is inclusion of a probability of forgetting, as a counterpart to the probability of transit. The additional parameter tracks the possibility of a student actually losing mastery on any given trial. The argument against the inclusion of the probability of forgetting is that a student who
stops giving correct trials may have never learned the skill correctly in the first place. However, the flexibility to include such extensions demonstrates the power and capability of utilizing a technique like BKT to help understand student mastery and help guide students to success more consistently and more quickly.

Human Pedagogical Applications

Even with no immediate way to apply BKT in software, using the algorithm as a way to think about teaching on a day-to-day basis can be extremely fruitful. Here we look at lessons learned from applying BKT to music teaching and learning, and practical ways to use a more quantitative approach to thinking about music pedagogy.

Confidence in Student Repeatability

One particularly fascinating element of using the BKT process is its ability to determine how many trials are necessary before the system is convinced that the student has true mastery of the task. Particularly with difficult-to-master skills, or skills that have a high probability of guess, the number of correct repetitions, and especially consecutive correct repetitions, are quite high before the algorithm can justify moving on. Duke notes that teachers often don’t require enough trials from their students to ascertain if the student truly knows how to accomplish the task at hand (2009). Perhaps thinking more about the task at hand and really analyzing how likely a student is to succeed the next time they are asked to play a section or exercise can help to guide lessons toward more lasting achievement.

For example, looking back at the staccato example is particularly educational. Suppose we adopted a more intuitive, heuristic rule that simply said we consider a task learned after
three correct trials in a row. Had we done this, the process would have moved on and then the staccato skill would fail again later on, since the student did not have enough opportunities to solidify the ability to play staccato in isolation. The idea that practice is the process of making tasks easy appears frequently in the literature reviewed earlier. A good reason for the high level of repetition shown is that the goal of practice should be not only that the student is aware of, and capable of, a task, but should also have mastery of a task and be able to reliably demonstrate the skill in a variety of contexts.

**Knowledge Versus Skill**

BKT can also be used to help recognize the difference between knowledge and skill. For example, consider a student that has gone through an activity that presents F sharp in several different octaves and is asked to play the given note, instead of working on finding and playing sharp notes in general.

For the sake of illustration, we give the following parameters to the BKT algorithm to trace if the student can reliably find F sharp:

**Table 9 – BKT parameters for finding F sharp**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( P(LO) )</th>
<th>( P(Transit) )</th>
<th>( P(Slip) )</th>
<th>( P(\text{Guess}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; ( p &lt; 1 )</td>
<td>0.25</td>
<td>0.2</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notice that while \( P(Slip) \) is assigned a reasonable value, \( p(\text{Guess}) \) is set unusually low. Even if a student were to play a note totally at random, there’s a 1-in-12 chance of getting it right, much higher than the 1-in-20 chance we give here. These settings are not set in a particularly appropriate or realistic way. However, more realistically assigned parameters
produce a much more onerous and impractical activity. With the parameters as they are, the following table is a possible interaction before the system decides mastery is acceptable (i.e. when two consecutive trials with $p(L) > 0.95$ have happened).

Table 10 – Student trials for finding F sharp

<table>
<thead>
<tr>
<th>Attempt</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

If $p(\text{Guess})$ were raised to 0.15, the student would have needed to execute a further three correct trials in order to move on in the system. No doubt, even nine trials of just playing different F sharps over and over would feel monotonous, let alone twelve or more. In contrast, there were sixteen trials for training the sharps skill as a whole, despite fewer correct answers.
in the first part of the activity. There is no easily quantifiable, mathematical metric that can explain exactly what makes this activity ineffective, but a pedagogical analysis can easily reveal the activity is testing rote knowledge, rather than a transferable skill.

In the BKT literature, algebra is a common subject of applied research. If we think of the kinds of activities found in algebra, often there will be many problems that necessitate the same skill set; for example, solving for $x$ in a linear equation (of the form $mx + b = c$). The idea of having a student solve the same problem over and over is absurd, which would be the analogy of finding F sharp repeatedly. In algebra, solving a generic linear equation is a skill, but knowing the value of $x$ in a specific equation is knowledge. Thus, it would follow that finding F sharp is knowledge, but being able to figure out how to find the sharp of any note is a skill. Thinking in this way can help us realize which activities and tasks are most useful to students and as pedagogical tools.

**Further Study**

There currently exist no rigorous studies of the effectiveness of any practice apps, or even individual features of music tutoring software. Without randomly assigned treatment and control groups, preferably monitored in a longitudinal study, there is no way to concretely ascertain the motivational and achievement impacts of music apps. In the case of music, there would need to be additional controls for teacher instruction and support of the technology, since a teacher who is particularly supportive, or unsupportive of technology could have an undue impact on study outcomes. In the case of a new technological approach as proposed here, there would need to be further research comparing the proposal to both extant software
approaches and unassisted outcomes. There is little doubt that this kind of research is needed, but in the absence of such studies, we can draw some conclusions based on the research we do have. Given that a BKT-based approach allows an app to encourage many of the good practice behaviors seen in literature and creates an environment with more opportunities for intrinsic motivation, we believe it is reasonable to conclude that the inclusion of the technology in practice assistance will be a great help for students.

Conclusion

In the course of this thesis, we looked at research about effective practice, and the impact of motivation on the quality of student practice. Then we looked at the recent trend of gamification in apps and analyzed the ways in which apps strive to create an atmosphere that is motivating and helps student fix problems more readily. The major drawback to all modern music instruction software is the lack of understanding when a student has learned a skill or not, something that we propose Bayesian Knowledge Tracing is capable of addressing. The demonstration clearly shows that BKT has tremendous potential to set goals and drive more productive student practicing behavior. The more focused nature of a BKT-driven tutoring system would have numerous additional benefits to intrinsic motivation, self-efficacy, and achievement.

Using BKT in the context of practice tutoring software gives a computer enough information about skill mastery to start making better, pedagogically-sound decisions. These decisions can help guide students towards working on what challenges them most while staying within what they can achieve. This type of goal-choosing and critical evaluation is one major
component to effective practicing, but is also important to flow theory, which requires that
tasks challenge the student, but not overwhelm them. If a BKT-driven design can create more
flow experiences, and expedite success, then the system will be far more intrinsically rewarding
and longer lasting than a gamified system.

Piano Maestro ranked in the top 100 downloads in the education category on the Apple
store on the first of April 2018 despite being available since at least July 2014 (App Annie: Piano
Maestro by JoyTunes, 2018). This demonstrates the popularity of these kinds of tools and
underscores the potential impact improvements could have for students using these tools.
With technology becoming only more and more prevalent in daily life, tutoring systems for
music will continue to flourish and probably grow even further. Creating systems that
encourage both short and long-term growth in the app space will become an increasingly useful
tool for many students as well as teachers.

Software that includes BKT-driven decision making has tremendous applicability to a
wide variety of situations. Young students could benefit the most from such a system, since
their repertoire is the most easily structured by an algorithmic system. However, with enough
technological advancement, there is plenty of potential to help advanced students who need
help practicing. There is a risk that some students will view a technological practice aid as an
alternative to a teacher rather than a supplement, but perhaps that will be offset by making
some level of piano education more accessible to potential students. Alternatively, when used
as part of a teacher’s studio, there is the potential to provide instructors with additional
information regarding the skills their students struggle with each week. Ideally, using a practice
tutoring system is not a permanent situation, since research literature shows that self-guided,
internally motivated practice is superior to enforced practice. With that in mind, the ultimate goal of any software should be to produce self-sufficient, effective practicers.

In addition to further research into the real-world efficacy of intelligent tutoring systems, like the one proposed here, another important follow-up topic is quantifying practice strategy effectiveness. BKT offers a way of quantifying effectiveness of tasks for teaching skills that has never been explored in the context in music. Individualization of the BKT model is particularly useful, since it enables examination of how different kinds of students learn skills via different tasks at different rates. This sort of study could open up an entire new way to think about learning styles and learning personality, and ultimately inform teachers how to deal better with a wider variety of students by giving an understanding of each student’s needs.

This type of follow-up suggests that, while BKT is foundationally a mathematical and computer science idea, the benefits of using it in piano practice and the most interesting research to be done with regard to BKT is pedagogical and psychological in nature. Which activities are most beneficial for a student trying to learn a new piece? Are there correlations between the efficacy of certain groups of activities and a student learning style? When is a passage ‘done’? BKT can help answer questions like these in a way that will help teachers address students’ problems.
References


67


Hawkins, W. J., Heffernan, N. T., & Baker, R. S. (n.d.). Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. In S. Trausan-Matu,


https://www.smartmusic.com/


https://www.pianomarvel.com/


