2004-03-18

An Implementation of a Drill and Practice System to Assist in the Teaching of Basic Music Theory

Todd C. Wilson
Brigham Young University - Provo

AN IMPLEMENTATION OF A DRILL AND PRACTICE SYSTEM TO ASSIST IN THE TEACHING OF BASIC MUSIC THEORY

by

Todd C. Wilson

A project submitted to the faculty of

Brigham Young University

In partial fulfillment of the requirements for the degree of

Master of Science

Department of Instructional Psychology and Technology

Brigham Young University

March 2004

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Todd C. Wilson

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Paul Merrill, Chair
Date	David Williams
Date	Stephen Jones

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Todd C. Wilson in its final form and have found that (1) its formats, citations, and bibliographical style are consistent and acceptable and fulfill university and department style and requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date	Paul Merrill Chair, Graduate Committee

Accepted for the Department

Andrew Gibbons
Department Chair

Accepted for the College

Barbara Culatta
Associate Dean, David O. McKay School of Education

ABSTRACT

AN IMPLEMENTATION OF A DRILL AND PRACTICE SYSTEM TO ASSIST IN THE TEACHING OF BASIC MUSIC THEORY

Todd C. Wilson

Department of Instructional Psychology and Technology

Master of Science

In order to help a group of introductory-level music students achieve a foundational understanding of music theory, a series of computer-based instructional modules were developed using a system called Technology-Assisted Language Learning (TALL). This system, though primarily intended to be used in developing natural language-based instruction, was designed to be flexible enough to handle a broad range of academic subjects.

The design of the instruction was largely accomplished via formative evaluation, where student and expert reviews of prototypes played a significant role. Students reviewed three separate prototypes of the instruction, and experts in instructional design and evaluation were also asked to provide feedback. The instructional approach of the learning modules consisted of drill and practice exercises, which included remedial

feedback. Activities were sequenced such that review was required until a specified level of mastery had been achieved.

The implementation of the software was less than perfect as numerous software bugs were present throughout. This caused frustration on the part of students and resulted in inaccuracies in the data collected by the system. A formal evaluation of the software and implementation was conducted in order to answer specific questions generated by those identified as stakeholders in the experience.

TABLE OF CONTENTS

At BYU one of the first courses required of students entering the music program is Music 195: Music Theory 1. While most students arrive with a broad sense of music theory some do not have the basic theoretical knowledge proposed by this entry-level course. Naturally, students that come into the music program represent a great variety of music knowledge, and indeed enough come with a basic knowledge of music theory that it is inefficient for the professor teaching Music 195 to spend time reviewing such concepts and information. Those that don't possess this basic knowledge are expected to come up to speed quickly.

Context

The modules were developed using Technology-Assisted Language Learning (TALL), which, though originally designed to teach natural languages, is flexible enough to teach other subjects as well. To a certain extent, the development and deployment of the instruction were a test of TALL to see if it could indeed successfully supplement learning for a topic outside of language.

The technological context created by TALL provided an extensive foundation for building our instruction. Many of the aspects typically involved in developing computer-based instruction—such as activity sequencing, user authentication, data tracking, and content development—were already built into TALL. Moreover, TALL's content development tools allowed for quick and efficient changes to be made to activities and their sequences. TALL also provided a structured database for housing our content. For example, TALL is built on the notion of concepts (basically anything that can be learned), which can be categorized and have various types of relationships with each other. This provided a formal semantic structure in which we could fit ideas such as musical triads and scales.

I worked with Dr. Stephen Jones, a professor of music at Brigham Young University, in developing a series of instructional modules intended to assist students in gaining basic competency in select elements of music theory.

Purpose of the Instructional Materials

The instruction was primarily intended as a supplemental resource to allow students extra practice, to be accomplished on their own schedule and at their own pace. The software was also designed to supplement the students' textbook, providing richer feedback and interactivity than they would experience with a textbook. It is important to note that the instruction we developed within TALL was intended to provide only supplemental learning to students enrolled in an on-campus course. In addition to accessing the TALL instruction, students attended scheduled classes, made use of a textbook, and received personal assistance from Dr. Jones.

Project Background

This project was initiated by Stephen Jones who, as a BYU administrator helping to guide the TALL project, expressed an interest in using TALL in a basic music theory course he would be teaching Spring term 2003. Having a background in music, experience with TALL, and a need for a master's development project, I naturally offered to help Stephen bring the project to light.

Audience

Students using the instructional modules were members of Dr. Stephen Jones' Music 195 class taught Spring term 2003. As entry-level music majors they had previously passed a standard exam that assesses basic music competency. As such, we were able to presuppose at least a rudimentary knowledge of music, including the ability to read notes on a staff and familiarity with at least one musical instrument (including voice).

Literature Review

There are three principle areas of my literature review: (a) using computers to teach music theory, (b) drill and practice instruction, and (c) traditional music pedagogy. Another dimension of my literature review is examining existing software packages designed to teach music theory.

Computers in Teaching Music

The history of using computers in teaching music is naturally tied to the advancement of technology. Berz and Bowman (1995) specify four distinct "cycles/periods" of "educational computing in music education" (p. 17):

1. Development Period (before 1965)
2. Mainframe Period (1965-1978)
3. Microcomputer/Traditional Computer-Assisted Instruction (CAI) Period (1978-1989)
4. Emerging Technologies Period (1989-)

Around 1965 computer technology had developed to a point feasible for use in music education. Indeed, some of the first music education applications were created using early mainframe computers. For example, in a 1967 study described by Kuhn and Alvin (1967) music instruction was created on an IBM 1620 mainframe computer, with student input recorded via a microphone feeding into a 4-channel tape recorder. A loudspeaker and "typewriter-like terminal" were also used for inputs and outputs (p. 2).

While computer-assisted instruction was possible using mainframe computers, it was also expensive and cumbersome. The advent of the microcomputer helped to change that. Programming on microcomputers was not only simpler and less expensive, but their extreme cost-effectiveness over mainframes led to their proliferation, making them far more accessible to students. The rise of the microcomputer also saw the development of hypermedia technology. Berz and Bowman (1995) define hypermedia as “an interactive multimedia application in which discrete information units can be accessed in nonlinear ways” (p. 19). With hypermedia, the computer became more adaptive, and allowed for richer use of media; this, in turn, led to the “individualization of instruction” (Kuhn, 1974, p. 89). Computers were able to respond to student performance with precise, extensive, and immediate feedback.

In a 1995 study, Berz and Bowman identified the “Emerging Technologies Period” as a time where computer technology became increasingly associated with “complex, content-oriented learning environments” (1995, p. 18). “Improved delivery tools (power, speed, and flexibility)” and “newer instructional models” also significantly increased the computer’s effectiveness and application (Berz & Bowman, 1995, p. 18). This period includes the advent of Internet technology. All this has aided music education as students and teachers are better able to inter-communicate, and instructional material is more easily disseminated.

It is important to note that the use of computers in education does not necessarily enhance learning. Webster cites several studies that objectively question the effectiveness of computer-assisted instruction. Typical are the conclusions of one study that “showed no significant difference between the students’ scores on a music fundamentals test, demonstrating that out-of-class work with the custom software was as effective as in class teaching of music fundamentals” (Webster, 1992, p. 425). Webster also points out that the effectiveness of computers in music education will not only be determined by the nature of computer instruction itself, but also by the varying nature of the students involved. He notes that not all students "profit equally from technology, pointing to the fact that not all students are the same" (Webster, 1992, p. 423). Student familiarity and comfort level with computer-based instruction is undoubtedly one significant determining factor.

Despite the lack of evidence supporting the notion that computers in education will enhance learning, there are other side benefits that seem apparent. Parrish (1997), for example, conducted a study using two groups—one group receiving purely traditional classroom instruction, and the second receiving classroom instruction supplemented by computer-based instructional modules. Agreeing with Webster, this study found that "knowledge gained about music fundamentals remained constant" between the two groups involved; however, "music theory lecture time was reduced by using CAI," giving the computer-aided group a clear advantage (Parrish, 1997, pp. 98-99). Webster (1992) also cites a study by Goodson which found that "interactive hypermedia instruction required less instructional time in order to achieve equal or higher scores on a 22-item music listening test" (p. 424).

Although learning gains aren't a natural result of the application of technology in music education, a common trend seems to be that students often prefer to use computers to other forms of instruction. In one study cited by Webster (1992), data compiled on student attitudes "showed significant preferences for hypermedia formats as opposed to a control group, with definite preferences for laser-disc technology and the open-ended style of presentation" (p. 425). Making computer-based instruction available can also "save time for teachers, because students who progress at a slower rate than do others do not demand as much extra help from the teacher outside of class" (Parrish, 1997, p. 91).

As in any other area of education, the application of computers should be made judiciously. After an exhaustive review of current literature regarding computers in music education, Webster (1992, p. 435) concludes the following:

> So, is music technology effective and is it worth the trouble? On balance and on a very basic level, the answer to this question is yes. Does music technology hold the key for solving all our music teaching problems? Of course not. Are there abuses in its use? Absolutely. Does it always improve learning? No, much depends on the context—especially the teacher and its use instructionally. Is it worth the trouble to keep studying its role in music teaching and learning? Unconditionally, yes.

Drill and Practice

Learning activities labeled as "drill and practice" are often looked down upon because they only address low-level skills or knowledge. However, David Salisbury (1990) points out that "recent research on cognitive learning suggests that the role of drill and practice in learning may be more important than has previously been realized" (p. 23). What has traditionally been identified as fundamental units of knowledge and skill can often be broken down into still smaller units. It is in learning these "subskills" that a drill and practice approach seems to fit best. Merrill and Salisbury (1984, p. 19) give this example:

> Consider the musician learning a new piece of music. Once the mechanics of the piece have been mastered, the musician can then focus attention on interpretation.

> The implication of this research is that drill and practice can serve a very important role in bringing the learner to a level of "automaticity" on lower level subskills so that the learner can more readily perform some higher level complex skill.

While it's true that there are inappropriate uses of drill and practice, it would seem that this is one of the most appropriate areas of computer-assisted learning, since "[i]t performs best in well defined routines with predictable and correct reactions" (Decoo, 1994, p. 152).

A skill begins to reach "automaticity" as it "requires less and less attention and interferes less with other ongoing cognitive processes" (Merrill & Salisbury, 1984, p. 19). This notion of "automaticity" holds especially true in many aspects of music theory and performance. Concepts such as note names and positions, clefs, rhythm and meter indicators, and key signatures are all "building blocks" that are essential elements in mastery of theory and adept performance. This is especially true in the area of performance. Because music is a time-sensitive art form it is critical that musicians have mastered fundamentals in order to play pieces properly.

Vazquez-Abad and LaFleur (1990) give drill and practice educational applications as those "in which a learning task is broken into subtasks and then each of these is taken in turn, using feedback to reinforce mastering of each subtask as well as to correct failure to master" (p. 43). Most authors also include repetition, previous instruction, and feedback as important or necessary elements of drill and practice (see, for example, Vazquez-Abad & LaFleur, 1990, p. 43 and Merrill, 1986, p. 16). Other common elements include maintaining a history of learner performance (see Vazquez-Abad &

LaFleur, 1990, p. 44), determining a “subset” of items for a learner to work on at a time (see Salisbury, 1990, p.25), spacing out practice periods (see Salisbury, 1990, p. 25), and adjusting presentation based on student performance (see Salisbury, 1988, p. 23).

In terms of situations where drill and practice is most appropriate, Vazquez-Abad and LaFleur (1990) suggest the following: “Any time that a job or task calls for ‘learnable’ subtasks that have to be performed automatically, or when a skill has been targeted for instruction which must be brought in while performing a (more complex) task, we may then be dealing with prime candidates for drill and practice” (p. 44). Merrill and Salisbury (1984) suggest that “multiple discrimination skills” are often required to complete more complex tasks involving “classification, rule using, or problem solving” (p. 19). They define multiple discrimination skills as those that “involve distinguishing different stimuli from each other” (Merrill & Salisbury, 1984, p. 19). Salisbury (1988) makes a distinction between drills used to target “factual” information, and those intended to address “intellectual skills such as concept learning, rule using, problem solving, or using procedures” (p. 118). This is especially applicable in music theory, where identifying and constructing correct musical progression often involves adherence to a set of rules or procedures.

Feedback in drill and practice exercises is another aspect in which computers provide a distinct advantage over more traditional methods. When designed correctly, computer-based activities can provide immediate corrective and instructional feedback tailored to student responses. Vazquez-Abad and LaFleur (1990) suggest that because “[d]rill and practice is an instructional strategy rather than an evaluation procedure…special effort should be put into providing abundant, precise and quite specific feedback" (p. 46). While it's true that drill and practice exercises necessarily help to evaluate student performance, their principal purpose should be instructional. That is, they primarily provide a context for the student to receive course corrections and build automaticity of skills as they perform. Decoo (1994) summarizes well the importance of drill and practice in stating that "we may be putting too much of our R&D energy into low-impact CALL (computer-assisted language learning), while neglecting somehow the high-impact drill and practice of CALL" (p. 153). So while computers are often inappropriately used in education (for a variety of reasons), in the case of drill and practice exercises in music education, it would seem that they can be effectively utilized, saving the teacher from "the least exciting part of classroom teaching, namely the necessary drill and practice, and subsequent evaluation" (Decoo, 1994, p. 154).

Traditional Music Pedagogy

I think a useful starting point in summarizing traditional methods for teaching music theory would be to define what the term itself is. Interestingly enough, this turns out to be a difficult task. Opinions on what music theory is "often vary tremendously from school to school and sometimes from individual to individual within the same department" (Rogers, 1984, p.3). While music fundamentals (e.g., scales, chord progressions, and chromatic harmony) provide a critical basis for music theory, they are only a precursor to more advanced study. Duckworth and Brown (1978) describe this basic level as "various parts, or elements, of a piece of music and the ways in which these individual elements combine and interrelate to form a musical composition", that is, the "*systematic* study of how music works" (p. 2). The principle objective of our project was to assist students in obtaining foundational knowledge that they would later use as stepping-stones to more advanced topics.

Rogers (1984) emphasizes the interrelationship between thinking and listening in the study of music theory: “the more thinking that takes place, the more there is to hear; the more listening that takes place the more there is to ponder” (p. 8). He goes on to explain that musical analysis centers on this cyclical relationship, leading the student to question how and why a particular piece of music “works”, which further leads the student to query how it might have been composed, and, subsequently, how it should be performed, heard, or taught.

According to Rogers, pedagogical approaches to music theory can be categorized under four dichotomous sub-headings, which I will summarize below.

Integration vs. Separation

Music theory programs often differ in their degree to integrate written and aural skills. The integrated approach “mixes ear training, analysis, and composition within a single unified course each semester” (Rogers, 1984, p. 16). The rationale for approach is that musicians will understand not only the technical aspects related to manipulating and analyzing written music, but also how those manipulations and alterations will sound when played or sung.

Those that favor separating written and aural skills “argue that intellectual comprehension and hearing abilities develop at completely different rates” (Rogers, 1984, p. 16), making an integrated approach less effective, as the ear will often “lag behind” the eye and mind. The separated approach emphasizes the pace at which two distinct classes of skills should be learned, maximizing the efficiency and effectiveness of instruction and learning.

Rogers (1984) to proposes an alternate way of understanding an “integrated” approach, essentially changing its definition. In this “integrated” approach the study of principles is used to explain “how a competent listener makes sense out of a piece of music,” which are then used “as the basis of an ear-training program that emphasizes not merely accuracy of performance or response but also the understanding of what is heard along with the attempt to hear even more” (p. 17). This approach places less emphasis on the organization and structure of classes, and more on understanding relationships between thinking and listening.

Comprehensive Musicianship vs. Isolation

Rogers (1984) defines comprehensive musicianship as “a curricular arrangement that attempts to include and interrelate at least four subjects that otherwise might be taught more conventionally as isolated courses.” These courses are “music literature, harmony, counterpoint (and melody), and formal analysis” (p. 20). In contrast, the isolated approach would involve addressing each of these subjects as separate courses, essentially leaving it to the student to draw relationships between them.

While a unified understanding of music is a lofty goal, Rogers also points out that there are issues to take into consideration with such an approach. Comprehensive musicianship requires intense preparation and dedication on the part of both teachers and students. With so much material to cover there is often a tendency to emphasize breadth of study over depth. Also, “eventually dogma can set in and work against the spirit of innovation and experimentation that is at the heart of the CM (comprehensive musicianship) philosophy" (Rogers, 1984, p. 23).

In spite of the potential problems, Rogers (1984) proposes that aspects of a comprehensive approach can inform more conventional "isolated" teaching methods. For example, one teaching a class on harmony might interweave "aspects of texture, pacing of climax points, the changing historical evolution of a pitch or rhythmic language, [and] performance suggestions" (p. 23).

Historical vs. Astylistic Approaches

These two approaches contrast the advantages and disadvantages to emphasizing the chronological ordering and stylistic qualities of musical periods. Tracing music on a chronological continuum "encourages study of the developmental aspects of theoretical principles," which also "introduces additional topics such as the nature of musical style itself and style change" (Rogers, 1984, p. 25). A given musical movement is seen as being an outgrowth of prior work, much like scientific discoveries build upon one another.

While few would contest the validity of music's evolution through history, those supporting an "astylistic" approach argue that "the study of music should begin with examples that are most well known or most relevant for undergraduate students" (Rogers, 1984, p. 26). Some argue that study should instead begin with movements considered to be relatively simple, then moving to the more complex, or possibly starting with contemporary music since it would be most familiar to students. Others advocate learning centered on principles and ideas that cut across historical and stylistic movements. This approach appropriately compares and contrasts works from completely different time periods.

Concepts vs. Skills

The fourth approach contrasts the study of music in the abstract (conceptual) with one that emphasizes practical application and skills. A curriculum that stresses “reading, discussing, analyzing, and thinking about music” (Rogers, 1984, p. 27) can lead to a deep and informed understanding of music, but it can also lead to the pitfall of settling into a rigid approach that tends to “restrict the consideration of alternate questions and answers” (Rogers, 1984, p.28). Consequently, instruction of this type can become stagnant and shortsighted.

In contrast, the skills approach advocates learning by doing—including “singing intervals, aurally recognizing a chord, [and] transposing a scale” (Rogers, 1984, p. 28). The apparent assumption is that students will mature in their understanding of music by experiencing and creating it, as opposed to just pondering it.

Again, Rogers advises against taking an “either/or” approach. There exists a natural interplay between concepts and skills, as one will inform the other. Rogers (1984, p. 28) proposes that an “amalgamation of the two is the only sensible solution,” avoiding an imbalance.

> After discussing these four dichotomous approaches to music pedagogy, Rogers (1984) points out that, on the whole, most of these relationships are “not contradictory at all…but rather complementary” (p. 29). A teacher will inevitably carry biases and preferences in each camp, but ultimately should strive for a blending based on present needs. It is important, though to be aware of the

strengths and weaknesses of all sides of the issues, so as to know the appropriate approach to take for a given situation.

Music Theory Educational Software Packages

In this section I review several software packages designed to teach music theory. I provide an overview of each package, as well as a link to the web site of each so that more information can be obtained. Additionally, I'll briefly address strengths and weaknesses and how they compare to our project.

MiBAC Music Lessons I & II

The MiBAC packages, from MiBAC Music Software (http://www.mibac.com/Pages/Downloads.html), provide drill and practice exercises that give the learner extensive control over what is to be learned. Relatively few instructions are given for each of the activities, but the user interface is more or less intuitive. The Music Lessons I package covers the more basic aspects of music theory, and covers the following topics: note reading, circle of fifths, key signatures, major and minor scales, modes, jazz scales, scale degrees, intervals, note/rest durations, scales ear training, and intervals ear training. Music Lessons II addresses these more advanced topics: chord elements, triads, triads ear training, seventh chords, seventh chords ear training, and roman numerals. Additionally, Music Lessons II provides naming, writing, and playing variations on each of the activity types.

For each activity presented the software provides basic instructions that allow the student to attempt the activity, and subsequently check the correctness of their answer. Specifically, the learner is allowed to play the notes shown; and, if he or she gets stuck, they can click a "Show Me" button to see the correct answer. Figure 1 is a screenshot of a sample activity.

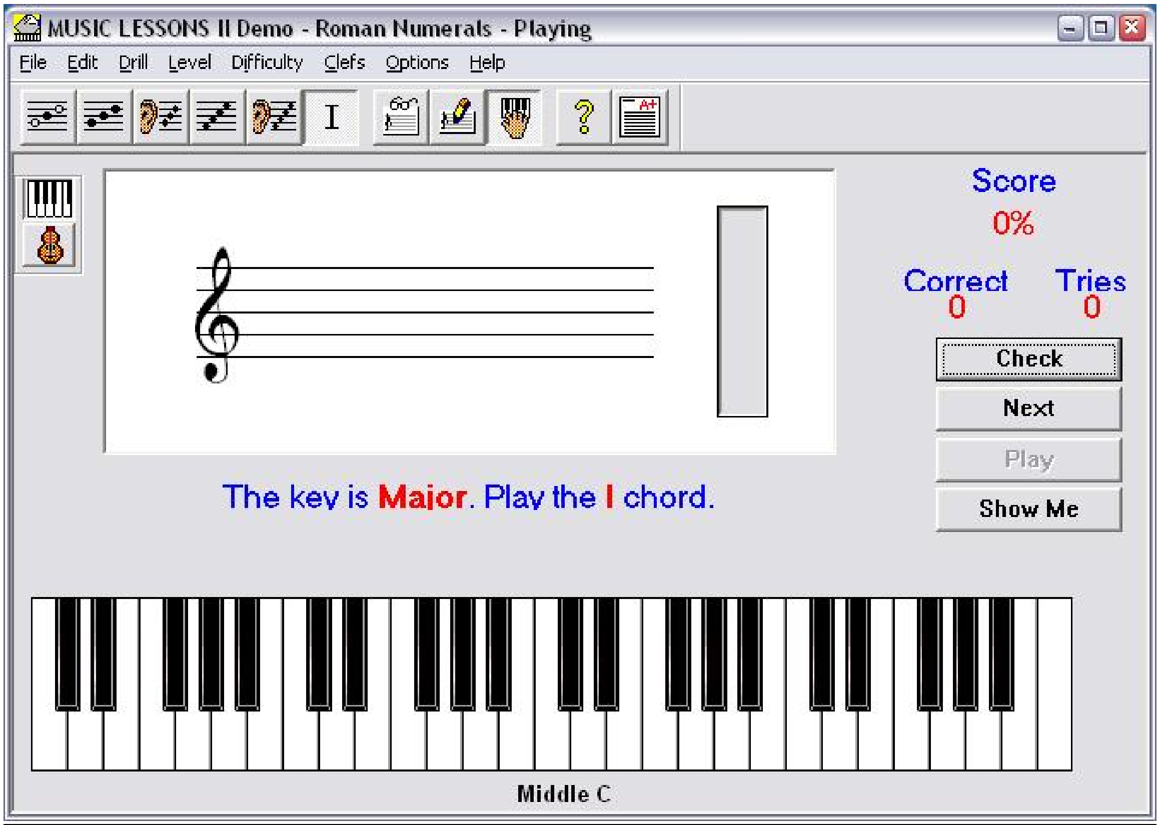


Figure 1. The MiBAC user interface.

Various elements of the activities can be controlled by the learner, such as the clef used (which can be randomized), the level of difficulty, as well as various visual preferences. MIDI support is also built in, so that an external device, such as an electronic keyboard, can be used with the application. Learner progress (or score) is tracked, and can be saved, and later loaded, to an external file.

The instruction that we developed, like the MiBAC software, focuses principally on the drill and practice of basic music theory and contains almost no instructional material. However, the user interface differs from MiBAC in that all notes are entered using a virtual keyboard embedded within the application. Given that the software emphasizes the practice of material, it would make a good supplement to music theory courses at beginning and intermediate levels.

Musition 2

Rising Software's Musition 2 (http://www.risingsoftware.com/musition2/) divides its topics into four categories: music reading, terms and symbols, key centers, and instruments. Within each topic are several sub-topics, each of which have accompanying activities and reference pages. The user interface is intuitive and easy to navigate, and gives the learner complete control over which topics will be studied when. The activities take a basic drill and practice approach. The software is designed to be used with multiple users and classrooms, and has administrative tools for these things built in. Security settings can also be adjusted, for example, to allow only certain users to alter groups of students and to edit other users.

After selecting a topic to study the learner is presented with individual activities in a sequence. Figure 2 gives a screenshot of a sample.

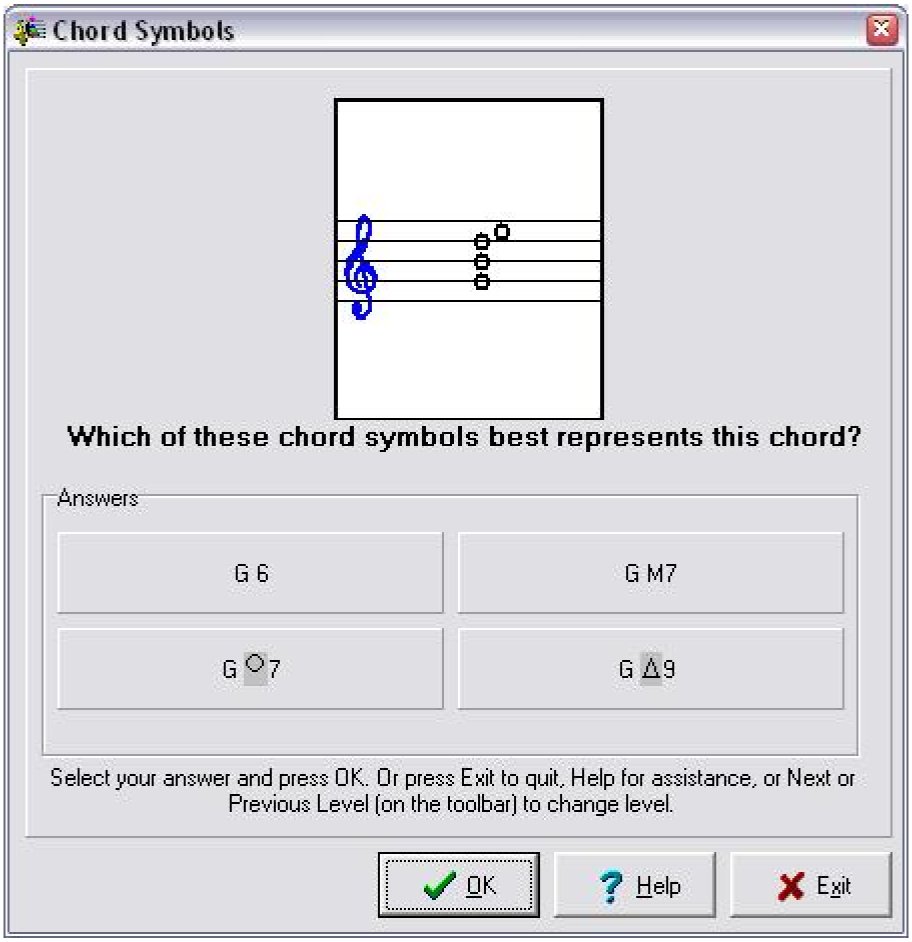


Figure 2. A sample Musition activity.

After making an attempt at an activity the learner is given a window that indicates whether or not the answer was correct and incorrect, and also allows the learner to elect to try again, play the correct answer, play the learner's answer, or advance to the next level.

Student performance is tracked, and can be later viewed by a teacher using a special reports interface. The software also contains a basic test creation and administration tool. Musition allows for instruction to be given in either German or English. MIDI can be used for both input and output.

The main strengths of Musition 2 lie in the simplicity of its interface, and the direct, supplemental correlation between its features and classroom instruction. In this regard it is similar to the learning activities we developed. The software could be greatly enhanced by more specific feedback, however, as the feedback it currently provides makes no attempt to provide specific suggestions to the learner.

Happy Note

Happy Note (http://www.happynote.com/music/learn.html) is designed principally for children, and consists of a series of games intended to provide practice on basic music concepts. The learner can control various visual and audio elements, such as background color and sound effects, as well as the level of difficulty for the instruction.

Figure 3 is a screenshot of an activity designed to give practice on identifying the names of notes on the staff.

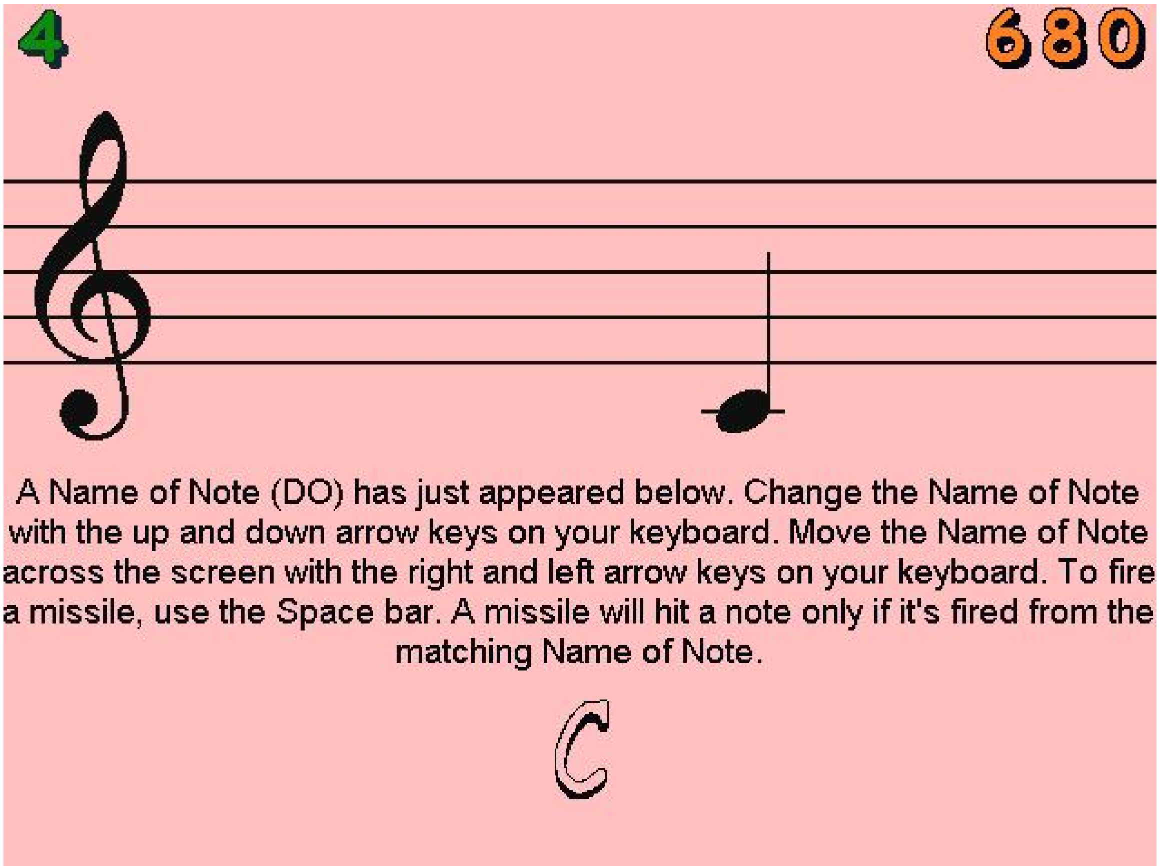


Figure 3. A screenshot of a Happy Note activity.

The design of Happy Note bears little resemblance to the instruction we developed principally because its target audience is so different. Its emphasis on visual appeal and ease of use make Happy Note ideal for younger learners focusing on rudimentary musical concepts.

GNU Solfege

GNU Solfege (http://solfege.sourceforge.net/) is a software application intended primarily for ear training. While the user interface is somewhat confusing, it does give quite a bit of control to the learner. Topics include 9 chords, harmonic intervals, melodic intervals, scales, 12 tones, and rhythm. Most of the activities have a similar setup, and generally involve playing a note, chord, or interval, then requiring the learner to identify what was played. The learner is allowed to repeat what was played as many times as is desired. If the learner gets stuck a "Give Up" button is often available to reveal the correct answer.

Figure 4 is a screenshot of a sample activity used to practice interval identification.

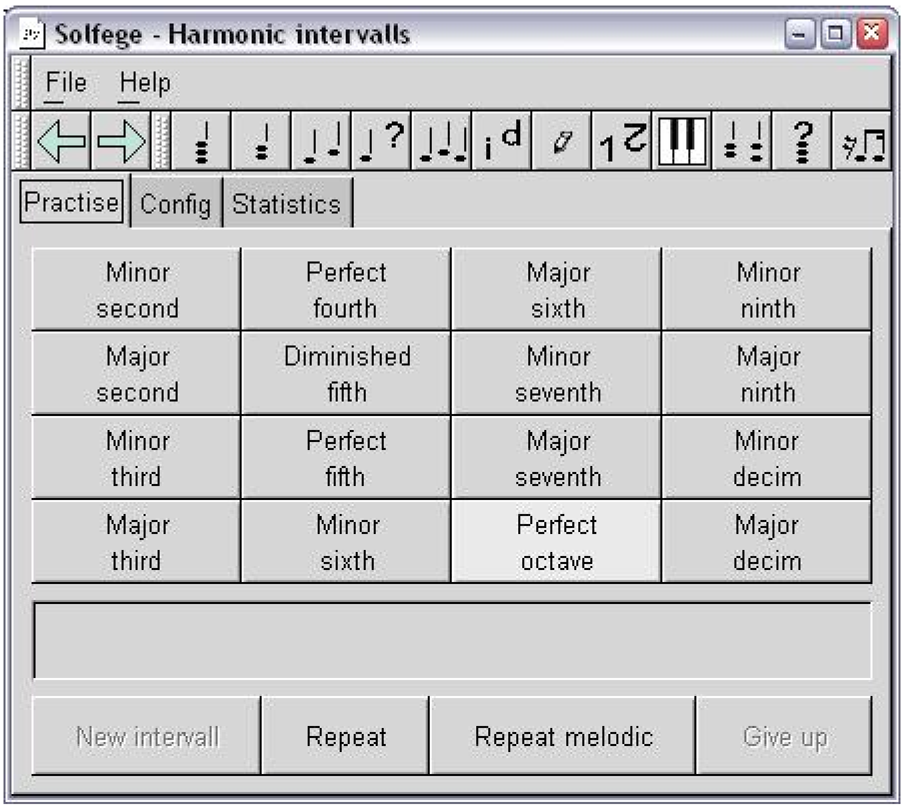


Figure 4. A Solfege activity on musical intervals.

Solfege tracks learner performance on an activity level, and reports are available that summarize the data on a session basis or over time.

In spite of its non-intuitive user interface, Solfege would likely appeal to those primarily interested in ear training. While it's relatively visually unappealing, but has definite practical value. Were we to expand the instructional material we developed to focus more on ear training it's likely that we would take a closer look at Solfege.

MacGAMUT

MacGAMUT (http://www.macgamut.com/) is intended primarily for ear training, though it covers many of the same topics that we did in our study. The user interface to the application is clean, but not particularly intuitive. It lacks some of the refinement of other programs, but is still very useful.

One of the strengths of MacGAMUT lies in the feedback it provides to the learner. Though the feedback on activities is primarily correctional (as opposed to instructional), the learner is given access to a range of data relative to their progress. The learner is able to view how many "levels" have been mastered, number of tries, number of correct answers, and time per activity.

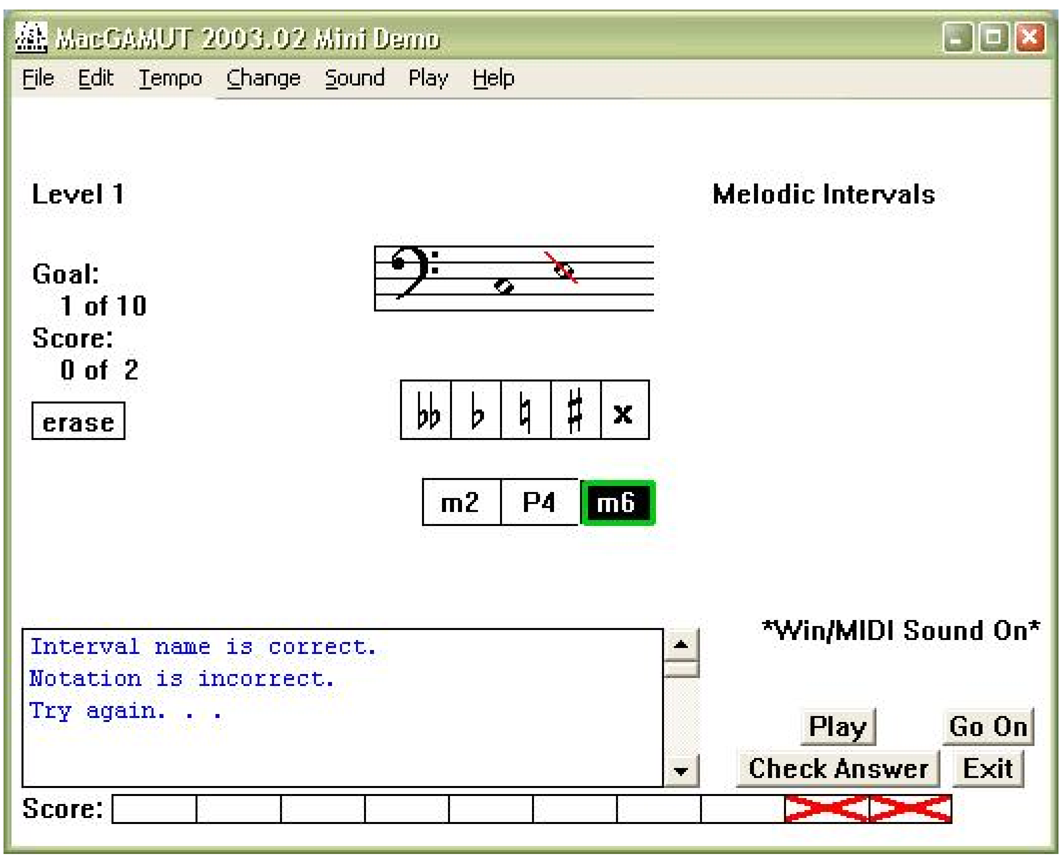


Figure 5. A MacGAMUT providing feedback on an activity.

Activities include aural training on intervals, scales, chords, and harmony. A student can also create his or her own drills by selecting specific concepts they'd like to review. The learner is also given basic control over such things as the tempo and instrument type. The application allows the learner to make multiple attempts at an activity, and tracks how many attempts have been made.

Teoria

Teoria (http://www.teoria.com/software/index.htm) is a flexible application that covers intervals, scales, and chords. The user interface is somewhat non-intuitive, but does give the learner ample control over topics to be learned. The learner can input answers with the graphical keyboard or using an electronic keyboard via a MIDI interface. For each learning topic teoria provides activities to be used for practice, as well as “construction” activities, which are more exploratory in nature. The varied types of activities supported by teoria sets it apart from the others. For example, the scale construction activity allows the learner to view and listen to various scale types (e.g., major, natural minor, pentatonic major), as shown in Figure 6.

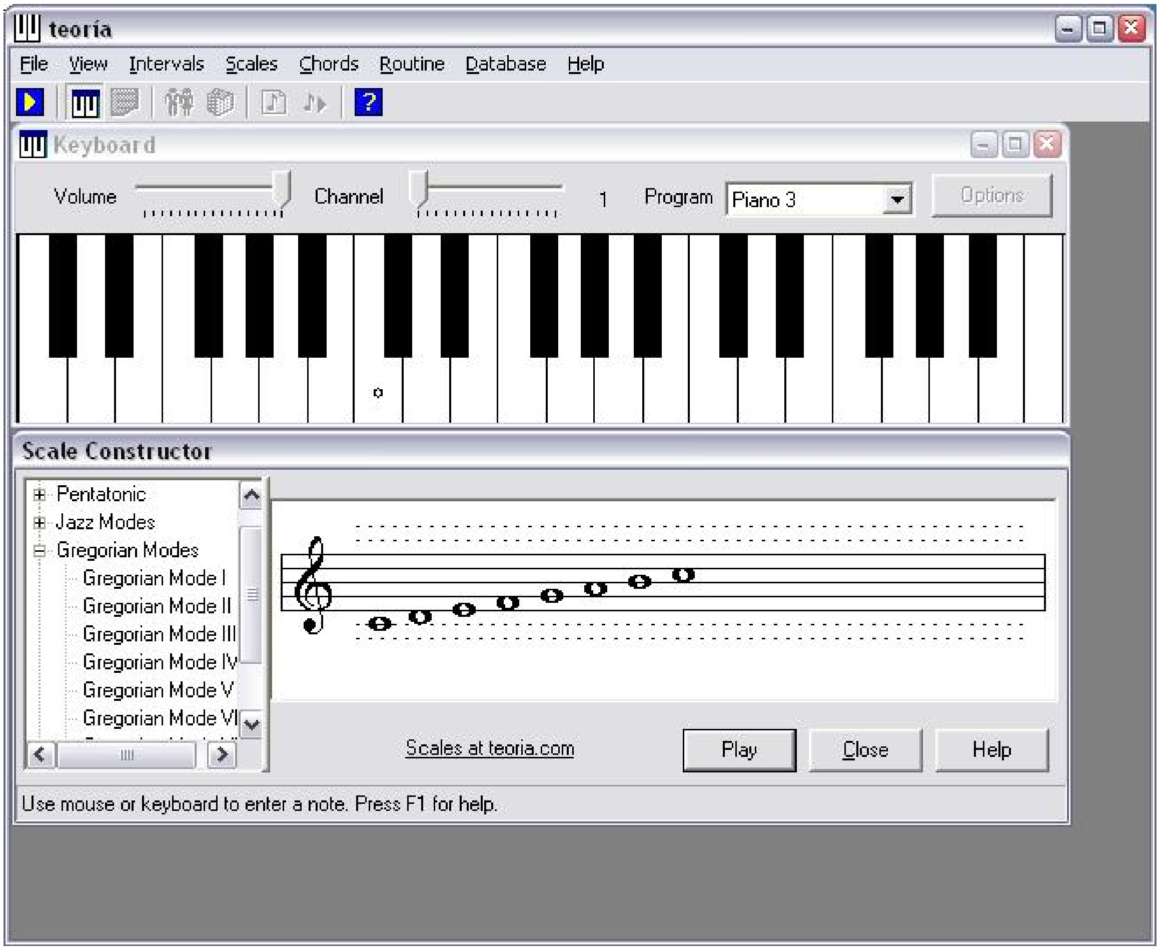


Figure 6. The teoria scale construction activity.

The teoria application also allows groups and users to be created, so that the software could be used in a classroom setting. Student scores are tracked, and can be viewed in a report format.

The flexibility of teoria seems to be one of its greatest strengths. In theory courses that emphasize intervals, scales, and chords (such as the course our project was designed for) teoria would make an excellent supplement.

Metronimo

Metronimo (http://www.metronimo.com/uk/index.htm) covers very basic aspects of music theory, and is primarily intended for children. Topics covered include notes, rests, inflections, clefs, and key signatures. A high emphasis is placed on visual appeal in the activities, which contain almost no instructions. The format of most of the activities is nearly identical—on the left side of the panel are a series of labels, with the right side containing musical symbols. The student is to drag each label from the panel on the left with it's corresponding symbol in the panel on the right. Figure 7 shows an example of an activity designed to drill key signatures. The consistency of the interface across the various activities makes for easy usage, especially for younger learners.

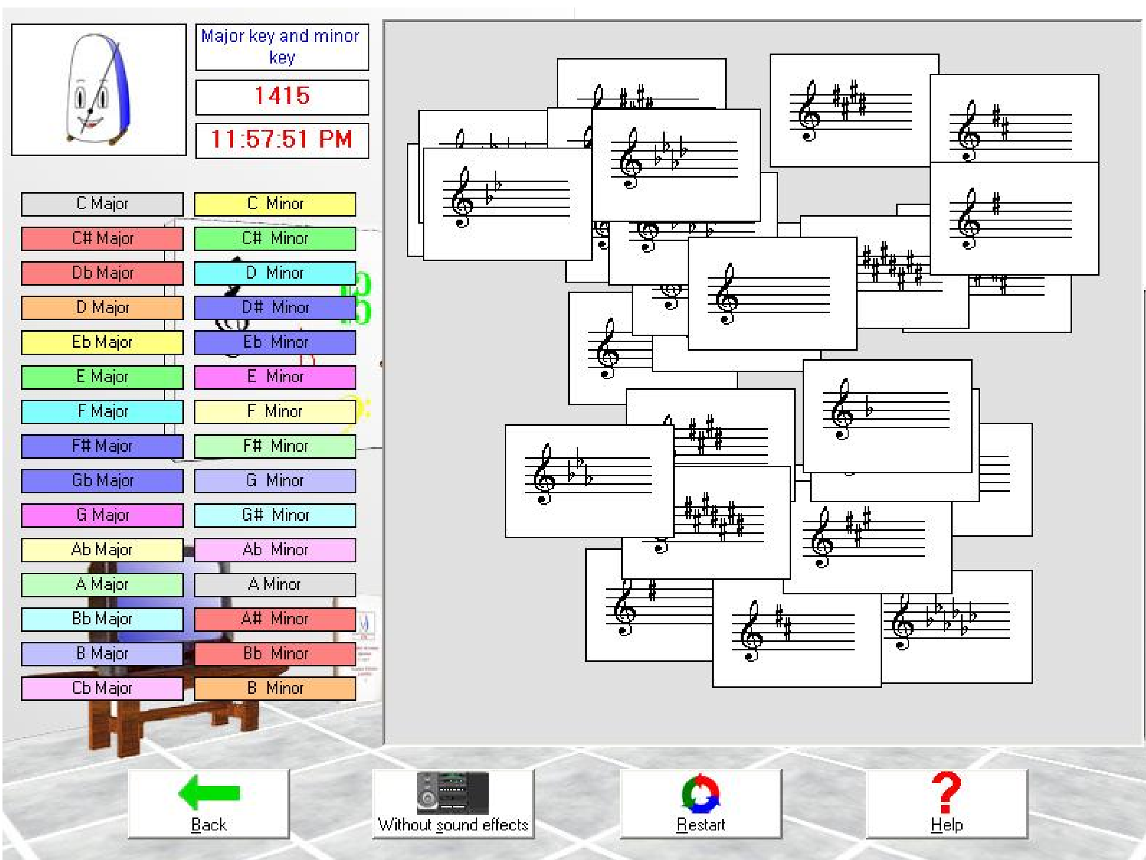


Figure 7. A Metronimo activity dealing with key signatures.

Metronimo apparently offers no ability to track or save learner performance data.

Metronimo is colorful and simple to use, which makes it especially suitable for young learners. The user interface is somewhat unique, so it may present a small learning curve at the outset.

Alfred's Essentials of Music Theory

The Essentials of Music Theory application (http://www.alfred.com/) provides in a single package both drill and practice exercises and instruction. Topics covered include staffs, notation symbols, note values, scales, key signatures, intervals, triads, and modes. The learning material is divided into a hierarchy of volumes, units, and lessons. Each unit also contains ear training exercises and a review section for each of the lessons found therein. After selecting a lesson to work on, the student is first shown an instructional animation explaining the concept. The learner is allowed to repeat and skip portions of the animation. After the instructional portion of the lesson finishes, the learner is allowed to practice using drill-type activities. If an incorrect answer is entered the student receives no feedback aside from an indication as to whether or not the activity was answered correctly; this is an area that could be improved upon. Figure 8 shows a sample activity dealing with intervals.

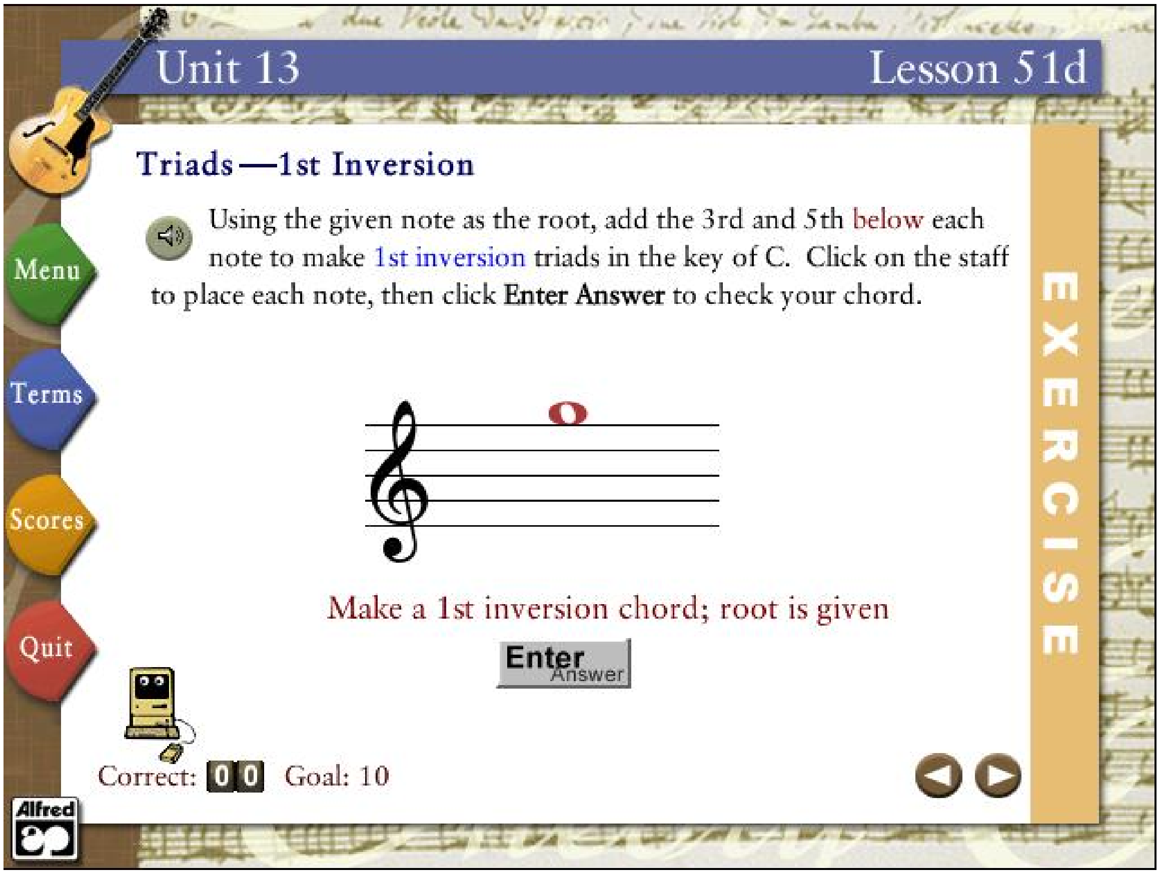


Figure 8. An activity involving intervals.

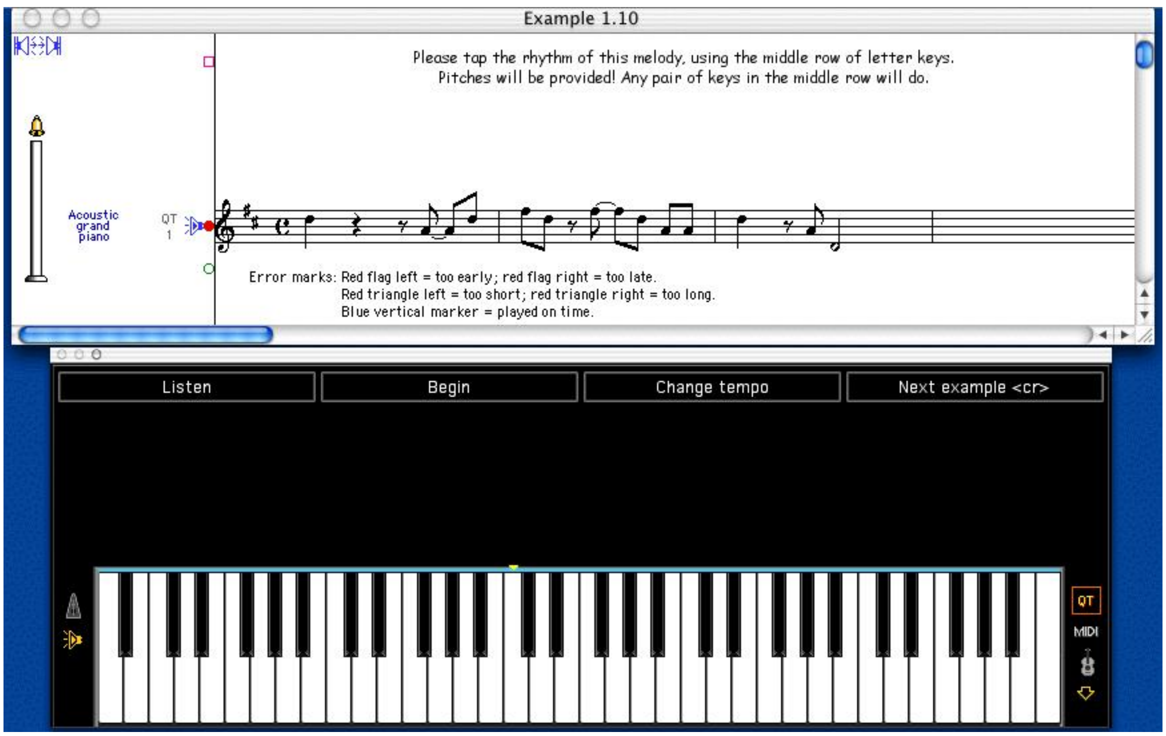
The application will track relatively extensive learner data, and allow that data to be viewed via an administrative interface. A test creation tool allows for instructor-defined assessment to be administered.

This is probably the most extensive package I reviewed in regard to the range of material it covers. The design is appealing and relatively intuitive, and navigation is simple. It's apparent that significant resources were dedicated to the development of the instruction, which includes text, audio, graphics, and animation. However, some students that used our software suggested that some supplemental instruction might be useful. If we were to pursue that suggestion we would likely examine The Essentials of Music Theory as an example.

Practica Musica

Practica Musica from Ars Nova (http://www.ars-nova.com/) is probably the most professional-looking product I reviewed. The application provides a significant amount of control to the learner, as opposed to the system determining what activities are appropriate to do at a given moment.

After selecting an activity the learner is asked to designate the level of difficulty he or she would like to start at. Activities are numerous and varied, and include: pitch matching and reading, interval playing and ear training, scales, pitch dictation, and melody writing. Instructions for a given activity are available when requested.



<u>Figure 9.</u> An activity on rhythm.

In many cases the clef and key can be set by the learner, allowing them to focus on areas they feel the greatest need to practice. Additionally, various keyboards can be chosen, including a plain piano keyboard, a guitar fret board, and a labeled keyboard (showing the names of the notes written on the keys). Other notes displayed on the screen can also be played at any time.

Instructional Materials

Design Process and Summary

In this section I describe the process we went through in determining the content, interface, and instructional design of the learning modules. Rather than build our instruction on a specific instructional design theory, I decided instead to briefly research general instructional design principles related to the types of activities we would be building, and then allow the design to be shaped largely through formative evaluation. This formative evaluation contained two principle facets. The first consisted of students reviewing prototype activities and offering feedback, which would inform our design. The second came as expert reviews by professional instructional designers. I will describe this formative evaluation process only briefly here, and will give more detail later in the evaluation section.

One of the first steps Dr. Jones and I took was to establish a series of general instructional objectives, and then extrapolate them to specific learning outcomes. This would provide the basis for both the instruction and the integrated assessment. We also mapped out the content (see the “Scope and Sequence” section), which reflected the instructional objectives, and delineated the exact topics we would cover.

Because of our short time schedule we decided to cover a few topics that we felt were most essential (e.g., key signatures, intervals, and triads), while leaving others to get to as time allowed (e.g., meter and rhythm). Because certain special features would need to be programmed into the system, we also prioritized certain requirements for the musical component that I would be programming. For example, Dr. Jones suggested that it was critical that the component be able to handle double sharps and double flats, but not quite as important that it allow students to construct key signatures.

As we discussed the types of activities to develop, based on the instructional objectives it became apparent that much of what we were after was factual knowledge recall and recognition. This led us to focus largely on "drill and practice" types of activities. As such, the initial design of most of the activities was relatively straightforward. Students would typically be given the task to either recognize or construct a particular musical element (e.g., a triad or scale), which also included the more complex task of resolving one chord to another. For example, the activity displayed in Figure 10 was intended to drill the student on recognizing various intervals:

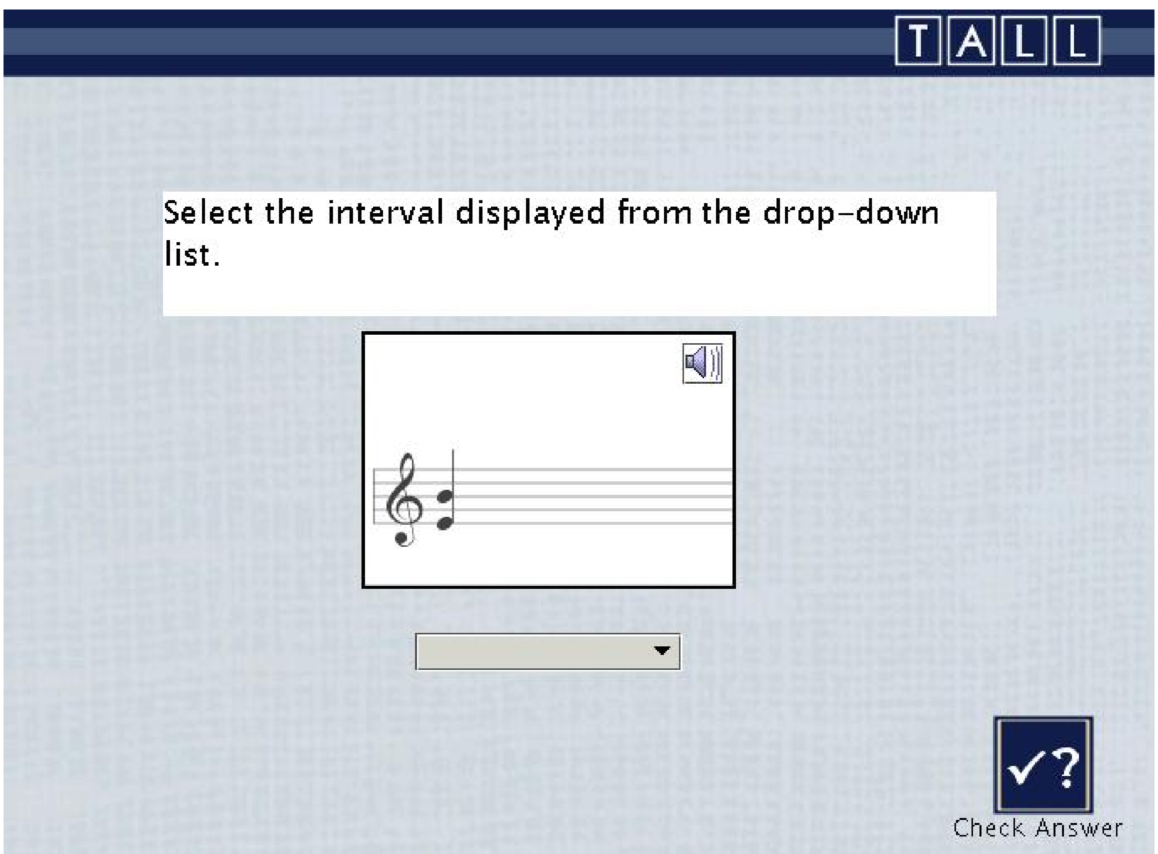


Figure 10. An activity used to practice interval recognition.

Simultaneous to our discussing the types of activities we would build, I performed research into other computer-based systems used to teach music theory. While this research was not extensive, it did help to validate some of the ideas we had regarding our approach. I also took the opportunity to discuss our ideas with Dr. Paul Merrill, who has published works on the drill and practice approach to learning. Dr. Merrill referred me specifically to his *Computers in Education* text, which contains a chapter specifically on drill and practice. Along with researching methods for teaching music theory through computers, I also surveyed several works that covered more traditional approaches to music pedagogy.

Having established basic preliminary designs for the activities, it became my task to construct a few of those activities using the TALL system. The majority of my personal time in developing content was spent programming a musical component that could be embedded into a TALL activity. This component had the following principle features:

1. Display a single staff with any clef (alto, bass, tenor, treble) or a grand staff (bass and treble)
2. Allow notes to be added, moved, and deleted
3. Allow accidentals (sharps and flats added to individual notes) to be added and deleted
4. Play the notes displayed (which was originally considered as only secondary in importance, which we anticipated putting in only if time allowed)

5. Recognize certain musical patterns (such as a correct harmonic minor scale or a successful dominant seventh resolution)

Figure 11 is a screenshot of the final version of the musical component:

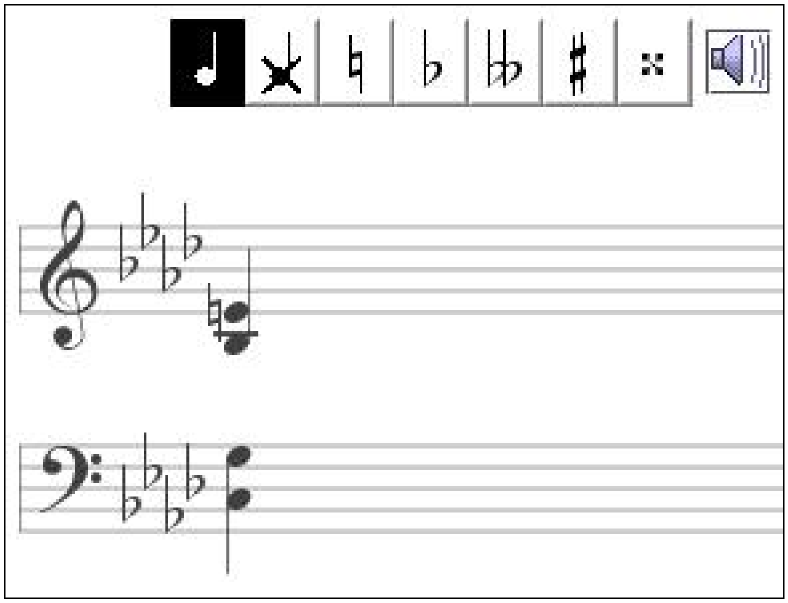


Figure 11. The musical component interface.

Most of the component's features (especially those related to recognition of certain musical patterns) were largely dictated by the activities we chose to develop. For example, the activity displayed in Figure 12 required that students be able to construct four types of scales beginning on a randomly selected pitch. They achieved this by adding accidentals (sharps and flats) to notes, which meant that the component had to be able to provide that pitch and recognize when the student had successfully spelled the scale.

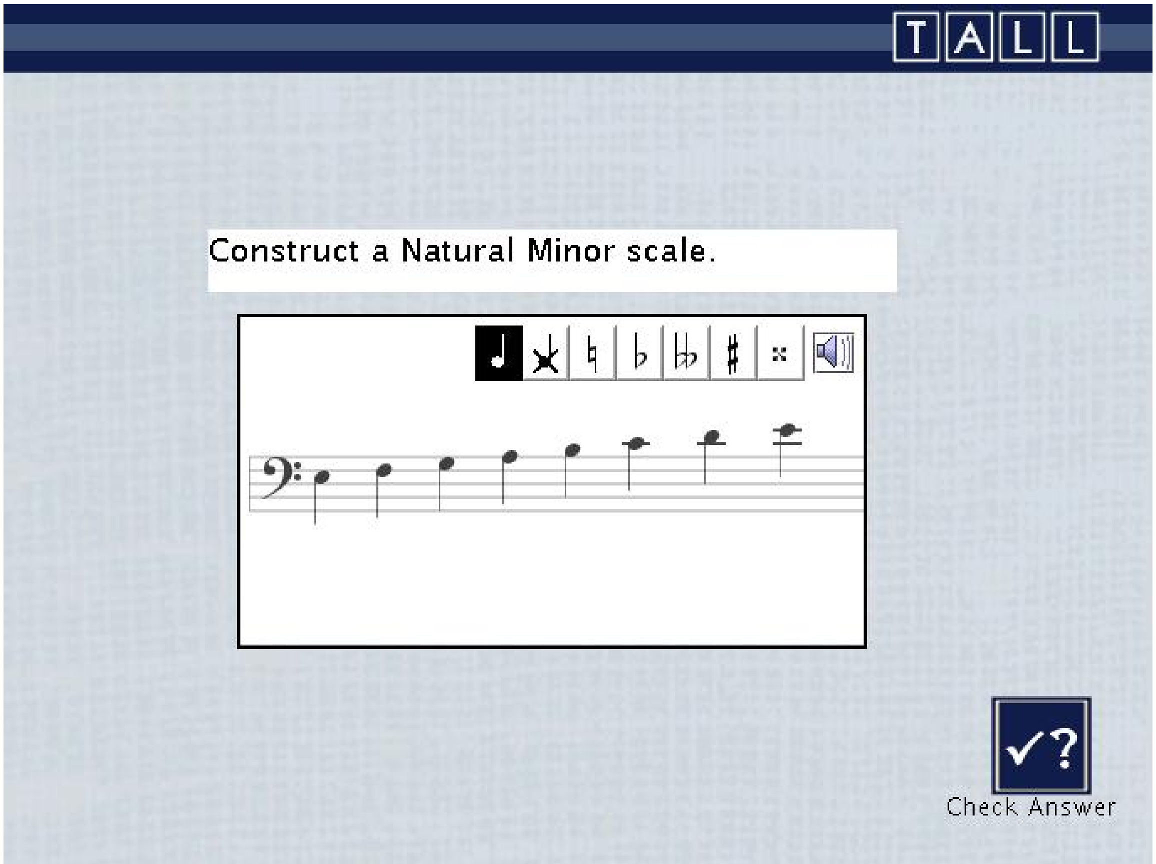


Figure 12. The “Spell Scale” activity.

TALL has the ability to allow for very sophisticated sequencing of activities, but because of time constraints we ended up using only basic sequencing. In most cases the student would be given a few initial activities for a given concept (e.g., a diminished tenth interval or a melodic minor chord), then would be required to review an activity until he or she was able to correctly answer it on the first try. On the activities where we allowed multiple attempts, students were required to review activities until they could correctly answer them within a certain number of attempts.

Figure 13 shows the sequence of activities used for intervals:

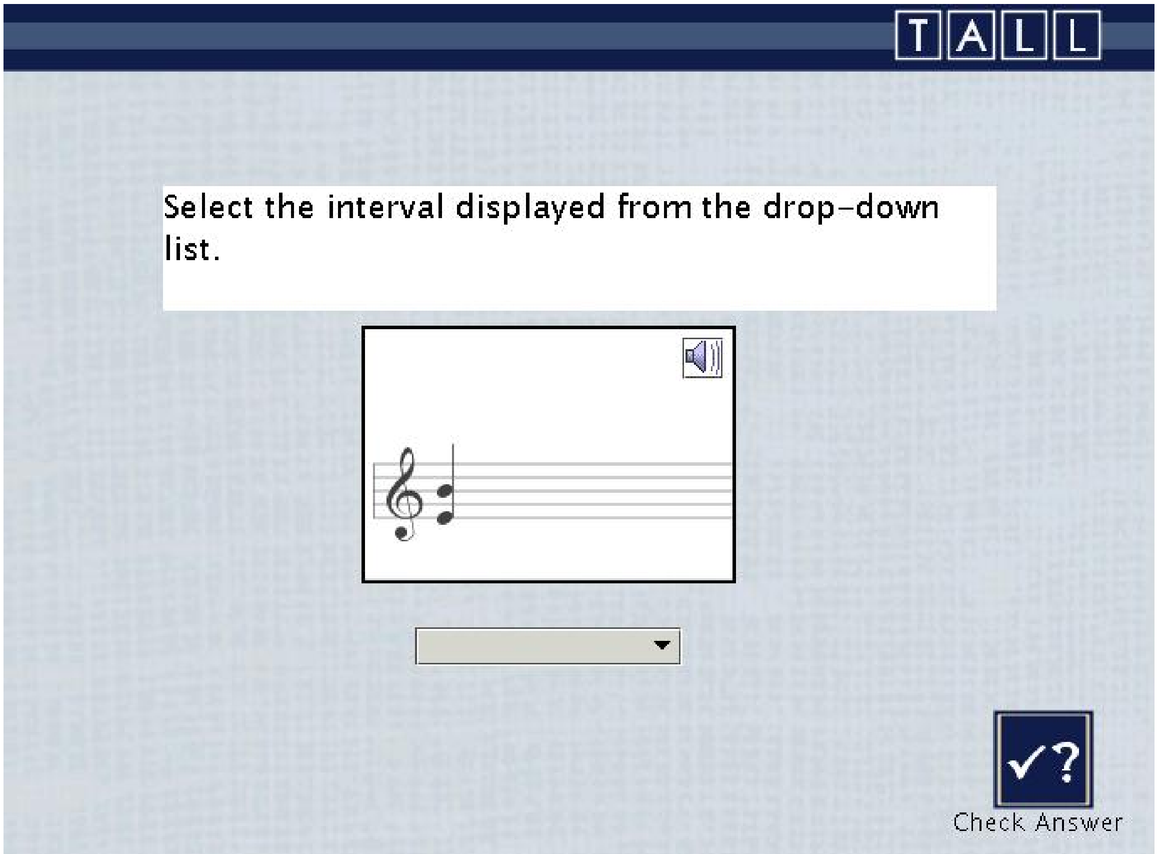


Figure 13. The “Spell Scale” activity.

The learner would first be given the “Recognize Interval” activity (the simplest) two times, followed by the “Spell Inverted Interval”, “Spell Enharmonically Equivalent Interval” and “Spell Interval” activities in a random sequence. Finally the learner would be given the “Spell Interval” activity, which he or she would be required to review until he or she could answer it correctly on the first attempt. Note that this sequence of activities would be followed individually for each of the intervals (e.g., perfect fifth, diminished tenth, major sixth).

After having decided upon the basic features of the musical component, I started into the prototyping phase of the project. This was a cyclical process wherein I designed a few activities, which I then had reviewed by students. The feedback they provided in these reviews helped to shape the design of the next prototype. This cycle was completed three times, and is described in more detail in the formative evaluation section below.

Concurrent with having students review these prototype lessons I also held two separate meetings with expert instructional designers. At these meetings I demonstrated for the designers the prototypes I had created, then allowed for their feedback, which also helped to inform the design and sequence of the activities. Again, this part of the evaluation is described in more detail below.

This "design by evaluation" approach seemed to work very well in improving our design. While theories and principles helped to provide a foundation for our approach, getting direct feedback from actual users and designers moved us toward design changes that were specific to our project.

Detailed descriptions and screenshots of the activities can be found in Appendix F.

Scope and Sequence

There were four main areas of the content, which can be broken down as follows:

1. Intervals. These were given in treble, bass, tenor, and alto clefs. Inversion of intervals was covered, as well as both melodic and harmonic intervals.
    a. Unison
        (1.) Diminished
        (2.) Perfect
        (3.) Augmented
    b. Seconds
        (1.) Diminished

(2.) Major

(3.) Minor

(4.) Augmented

c. Thirds

(1.) Diminished

(2.) Major

(3.) Minor

(4.) Augmented

d. Fourths

(1.) Diminished

(2.) Perfect

(3.) Augmented

e. Fifths

(1.) Diminished

(2.) Perfect

(3.) Augmented

f. Sixths

(1.) Diminished

(2.) Major

(3.) Minor

(4.) Augmented

g. Sevenths

(1.) Diminished

(2.) Major

(3.) Minor

(4.) Augmented

h. Octaves

(1.) Diminished

(2.) Perfect

(3.) Augmented

i. Ninths

(1.) Major

(2.) Minor

j. Tenths

(1.) Major

(2.) Minor

2. Triads. These were given using various key signatures and all four clefs.

   a. Choral spacing

   b. Keyboard spacing

   c. Open spacing

   d. Dominant seventh chord resolution

3. Scales. These were given using various key signatures and all four clefs.

   a. Major

   b. Natural minor

   c. Harmonic minor

   d. Melodic minor

4. Key Signatures

    a. All 14 key signatures were given.

Yet another factor we considered when intervals, triads, and scales were presented was the range of notes (e.g., how far above or below the staff a given interval should be placed).

Objectives

The objectives we based our instruction and assessment on can be broken down as follows:

General instructional objective

Identifies and constructs key signatures, triads, scales, and intervals.

Specific learner outcomes

1. Identifies intervals both in context (in a musical score) and in isolation.
2. Constructs intervals, given interval names and starting pitches.
3. Constructs interval inversions, given existing intervals.
4. Constructs enharmonically equivalent intervals, given existing intervals.
5. Identifies all major and minor scales in all clefs and key signatures.
6. Constructs all major and minor scales in all clefs and key signatures.
7. Identifies all key signatures.
8. Identifies triads in root and close positions in any clef.
9. Constructs all triads on the grand staff in open, choral, or keyboard spacing.
10. Resolves the dominant seventh chord in root position.

A Brief History of TALL

TALL, or Technology-Assisted Language Learning, is a computer software application that's been used for several years at the Missionary Training Center for the Church of Jesus Christ of Latter-Day Saints. Missionaries required to learn a foreign language will study via classroom instruction, personal study, instructor tutoring, and computer-assisted instruction (TALL). The current TALL system was written in the Visual Basic programming language, and has a relatively rigid instructional design approach.

Because of the success that's been achieved in language instruction with the current TALL system, it was decided that the system should be re-written so as to allow it to be more pedagogically flexible and allow it to take advantage of newer technologies. This is the project that I'm personally involved in. The newer version of TALL is designed to be much more flexible in terms of the type of content it can handle, and the instructional approaches it allows for. Part of the purpose of my project was to test how well TALL could handle non-language content.

Implementation Description

The instructional activities we developed were to be used as a supplement to Dr. Jones' introductory music theory course. We installed the software in three separate computer labs in the Harold B. Lee Library at Brigham Young University, which provided access on both weekdays and weekends.

Our resource constraints were very limited in terms of technical support and quality assurance, as it was intended that I would handle most of the work in these areas. I was also acting as the programmer on the project, which unfortunately left little time extensively test the software before students were to use it. As a result, the students were made to suffer through bugs in the system, especially at the outset. Even while the implementation was underway I was left to work "ahead" of the students, completing features and content that would be required for upcoming lessons. As students experienced problems with the software they would call or email me. Unfortunately, when severe bugs came up the best I could do was to ask that they stay out of the system until I could provide a fix. Based on the evaluation findings, this technical support dilemma was a significant source of frustration for the students.

I worked closely with the supervisor of the computer labs, Scott Hunt, to ensure that the software was correctly installed, and that lab attendants were notified of the implementation. I provided Scott with a document that gave a basic description of our software, instructions on launching it and logging in, as well as procedures to follow if something went wrong. The document also included my contact information so that I could be easily reached in the event that lab attendants had questions or major issues arose. Scott posted this document on a web site used by lab attendants. Following the implementation I inquired with Scott as to the frequency that issues related to TALL arose, and he mentioned that there were only a few that he was aware of. Additionally, I was never personally contacted by a lab attendant regarding TALL. Our instructions to the students were that I was to be personally contacted in the event that they needed help, and I would assume this is the reason the lab attendants received so few questions on TALL.

During the course of the implementation, I provided periodic statistical reports to Dr. Jones. In its current state TALL lacks a basic reporting system to be used by instructors, so my reports had to be derived by analyzing raw data from the database. This arrangement turned out to be problematic because Dr. Jones wasn't always able to get the information he needed in a timely fashion. Also, the reports did not provide many details on the performance of individual students.

Because of technical problems with the musical component I created, data was not always accurately captured. TALL does record performance-related data as the learner progresses through activities, but oftentimes the software would freeze and result in incomplete or inaccurate data being recorded. This proved to be especially problematic during the posttest, which ultimately resulted in very few student scores being accurately reported.

Evaluation

There were two main facets to my evaluation. The more formative part was performed before the learning modules were given to the students, and was used to shape the design of our instruction. The second was more summative in nature, and was intended to answer specific questions reached by interviewing stakeholders.

Context and Background

The evaluation for this project actually began even before its inception. Stephen Jones brought up the idea to use TALL in one of his courses, but didn't actually provide many details at the outset. A necessary first step was to flesh out more details related to his ideas and evaluate if what he had in mind was even a viable undertaking. Factors affecting the viability of the project related principally to time constraints and availability of resources. I also performed some of my own evaluation in determining whether or not the project would meet the requirements for a master's development project through the BYU Instructional Psychology and Technology department. This consisted mostly of discussion with the IP&T professors on my committee.

The instructional modules were developed within the context of a larger system (TALL), which is important to the evaluation; however, the evaluation of the TALL system itself is beyond the scope of my project. Background on this project and the TALL system were given earlier in this document, which should also serve as background to the evaluation.

Two primary evaluation activities were performed with the intent that they would guide our instructional design. These were one-on-one student evaluations and the expert reviews. This portion of the evaluation differed from the field test evaluation in that the activities were used to inform design, as well as the fact that we didn't formally consult with stakeholders in structuring them. In the field test section I address working with stakeholders, their questions, and the evaluand.

I should point out that I made every effort to ensure that the values of stakeholders were central to the evaluation. This primarily took place as part of the field test evaluation where stakeholders were formally interviewed regarding questions and issues that they were interested in. Data gathered from these interviews were boiled down to a series of questions and the related criteria used in determining the answers to those questions. The intent in centralizing the evaluation around the values of the stakeholders was to increase the probability that the data and conclusions derived from it would be used by stakeholders.

The evaluation was carried out with attention given to the Program Evaluation Standards (http://www.eval.org/EvaluationDocuments/progeval.html). Please see the "Meta-evaluation" section for a description on how I addressed each of the standards.

One-on-One Student Evaluations

During development of the instructional modules I performed one-on-one evaluations, using the "talk-aloud protocol," with six students as they tried out the learning modules. In these sessions I would sit the students in front of a computer displaying one of the activities we created. I would then ask the students to go through the activity verbalizing what they were thinking while doing so. The intent was to bring to the surface any issues that seemed confusing or ineffective, as well as to inform evaluation questions related to user experience. During these sessions I also took the opportunity to inquire about their general attitudes toward the instruction.

I will now describe in more detail the format of these review sessions. I would begin by sitting the student next to me while I sat at the computer. I then demonstrated how the talk-aloud protocol was done by verbalizing my thoughts as I used Microsoft Excel to sum a sequence of numbers. Before performing these one-on-one sessions it was suggested to me by a student that I might consider using paper prototypes initially, as he had read that oftentimes people are less inclined to suggest changes to an interface because of the perception that a software application is more difficult to change. In an attempt to mitigate this possibility I then explained to the students that the application was very open to change, and that they should feel free to offer any suggestions.

At this point I would describe to the student the actual context in which a student would do the activity they were about to see. I explained that they were seeing the activity disconnected from any others, but in an actual implementation a student would get a series of activities, somewhat like groups of flash cards.

I then brought up the first activity on the screen, and asked the student to sit in front of the computer. Rather than provide instructions on what the student was expected to do, I simply relied on the user interface and brief instructions within the activity to guide the student.

As the learner made their way through the activity I took detailed notes on what they were verbalizing. As much as possible I tried to say nothing at all until they were finished, though there were a few times when I made minor suggestions at points where the students seemed to get stuck.

After a student had completed an activity I would ask a series of questions about what they had just experienced, for example:

1. How much of a hindrance was the software?
2. Did the computer respond as you expected it to? If not, what response would seem more intuitive to you?
3. What aspects of the user interface seemed confusing? Did you understand the purposes of the buttons based on the icons?

While I used these questions to help guide the students, I also tried to allow for a free discussion so that they could offer any suggestions that came to mind. I performed these one-on-one reviews with six students on three separate occasions.

It was astounding to me the value of the students' feedback. Having done all of the programming and much of the instructional design of the activities, it helped me realize how blind I had been to many of the flaws. Here is a sample list of the insights gained by my observations of the students:

1. A drag-and-drop interface was expected to add notes to the staff, as opposed to the existing point and click interface.
2. The button used to play the notes was confusing. Some thought it would advance them to the next activity while others thought it would shift all of the notes to the left.
3. Certain icons were particularly confusing, such as the button used to remove a note and the button used to check the student's answer.
4. The mechanism for adding an accidental was confusing. Students often tried to click to the left of a note head instead of directly on it in order to add an accidental.

5. Oftentimes students had difficulty finding the “tool tips” (boxes containing short descriptive text when the mouse cursor was hovered over an element, such as a button).
6. Drop-down lists were often confusing.

We made specific changes to the design of the musical component based on student feedback. For example, initially we had placed as low on our priority list the ability for the musical component to be able to play the notes on the staff; however, in the first prototype more than one student suggested that this would be very useful, as ear training is often difficult for students when working alone. One student described how she would often record herself playing at the piano, then play the recording back later in order to train her ear. Another described how, in order to practice ear training, she and a friend would alternate playing chords for each other at a piano. It didn’t occur to me at the outset that this feature would be so desirable, so I moved it up in priority. We also made minor changes to the user interface when behavior seemed confusing or non-intuitive. For example, after clicking on one of the accidental buttons (e.g., flat or sharp), and adding an accidental to a note, we had initially designed it so that the selected accidental button would remain selected. A few students mentioned that they had expected the button designed to add and move notes to likewise become selected after adding an accidental, so I changed it to behave that way.

Expert Reviews

Another important aspect to our instructional design approach involved expert review of prototype activities. I met on two separate occasions with instructional designers—once with those at TALL, and the other with two professors on my committee.

One of the aspects of TALL that makes it unique is its ability to handle dynamic instruction, both within activities as well as in sequencing activities. As such, I met with TALL instructional designers in order to get feedback and suggestions on how best to take advantage of these features. Dr. Jones was also able to attend this session.

The format of the meeting with Dr. Jones and the TALL designers involved demonstrating for them the prototype activities I had created, discussing plans for sequencing the activities, then requesting that they give both general feedback, as well as feedback specific to the capabilities of TALL. This session was invaluable not only because of the level of experience each of these designers provided in general principles of instructional design, but they were also the originators and designers of the TALL system, and as such, provided unique suggestions based on the specific features of TALL.

For example, in terms of general feedback, one of the designers suggested that I try to minimize the amount of textual instructions given to the learner. He suggested that instructions be given perhaps only on the first few activities, then removed afterward. Additionally, several suggestions were given specifically relating to TALL. We discussed how best to fit the musical content (triads, scales, etc.) into the existing TALL information structure. The designers also provided specific suggestions on how I might take advantage of TALL's unique ability to adapt to learner performance.

Shortly after meeting with the TALL designers I also took the opportunity to meet with the professors on my committee, whose suggestions were equally valuable. They also helped to validate the approach we were taking and had several suggestions that we decided to incorporate into our design.

The format of the meeting with members of my committee was similar to that of the review with the TALL designers—I demonstrated prototype activities, and then invited suggestions. One of the more valuable items that came from this meeting related to the use of feedback in the system. In the prototypes I demonstrated there was essentially no feedback (really just a “correct” or “incorrect” label appearing after the check answer button was clicked), but subsequent to getting the feedback I tried to incorporate specific feedback into the design of the activities. At a minimum, in each case I displayed to the students the correct answer, and in many cases was able to provide feedback specific to the mistake the student had made in the activity.

Field Test Evaluation

The field test evaluation was designed to answer specific questions regarding the use and outcome of our implementation, which questions I arrived at by interviewing stakeholders. I’m hopeful that the results will help to inform future decisions in using TALL to teach non-language topics.

Stakeholders

There is a broad range of groups and individuals that might be indirectly affected by my evaluation, but in forming evaluation questions and determining the focus of my evaluation I decided to narrow the group of stakeholders to the following:

1. TALL administrators and instructional designers, which include Kent Parry, Joseph South, Bret Elzinga, and Sherrie Spencer.

2. Stephen Jones (the faculty member teaching the course the software was created for)
3. BYU music majors

Individuals that were excluded as stakeholders include other TALL employees and BYU music professors, that may indeed be indirectly affected by the evaluation in the future, but it is difficult to say how much. I decided to narrow the group of stakeholders to those mentioned principally because of time and resource constraints. At the same time, I feel that those I chose were able to adequately represent the interests of the others, given that they belong to essentially the same groups. The primary intended users of my evaluation are TALL administrators and Professor Stephen Jones.

Evaluation Questions

Representatives from each of the stakeholder groups were interviewed in order to form evaluation questions. The interviews centered on these four questions:

1. Which aspects of the basic music theory instructional modules do you think would be most important to evaluate?
2. What are the specific questions you would like answered?
3. What criteria should be used to determine whether your expectations or desires for the system have been met?
4. Could you give an example of an acceptable performance level?

I decided to interview each stakeholder individually and didn't reveal to any what the others had said prior to interviewing him or her. I feel as though this approach allowed each individual to focus on the questions and issues that most interested him or her, without being influenced by the opinions of others. While there was some overlap in comments made by the stakeholders, it was interesting how significantly their questions and issues varied.

I took extensive notes from these interviews, collated them, and boiled them down to the following primary evaluation questions:

1. Given that TALL was originally designed for language learning, how effectively can it be used to teach other topics?
2. As a development platform, how does TALL compare to other tools?
3. How effective is the instruction built for this project?
4. What are student attitudes toward the instruction?
5. Does the software perform such that it supports (or is transparent in) the learning process, or does it prove to be a hindrance?
6. How reusable is the software and instruction?
7. Is this system implemented and utilized as intended?

Generating these evaluation questions was a critical step in ensuring that the evaluation itself would be useful to the stakeholders. I was careful to center all of my data collection and analysis activities on these questions.

Data Collection

In order to more concretely support my evaluation findings I drew from several different sources of information. Using such a “triangulated” approach allows for strengths in one method to compensate for weaknesses in another. For example, I was concerned that in the focus group students would be less inclined to speak freely about the negative aspects of their experience, given that I was directly involved in developing the instructional modules. As such, I decided also to administer a questionnaire, which would allow the students to express their sentiments anonymously. In reverse, the survey provided only a small sample of information that, while very useful, was not nearly as detailed as the information I was able to glean by conducting a focus group. My sources of information follow.

Expert Reviews. To evaluate the validity and soundness of the content I conducted personal interviews with Dr. Stephen Jones and with Dr. Laurence Lowe, both experts in music theory. Evaluating the content involved both verifying that the visual and audio representations of the music were valid and accurate, but also examining the rules governing the activities. For example, there are a series of checks that need to be performed to verify that a student has correctly resolved a dominant seventh chord.

At the end of the project I asked another computer programmer to review the programming code I had written in order to get feedback on its potential reusability.

TALL-Recorded Responses. Another unique aspect of TALL is its ability to collect data regarding learner performance. Examples include the number of attempts a learner makes on a given activity, the time taken to answer a question, and the number of questions the learner has answered correctly. Because of time constraints we were only able to make use of a few of these data points, which came in the form of tracking how many activities a student had encountered as well as how many he or she was able to answer correctly. Additionally, on the dominant seventh resolution activities we tracked the number of attempts a student had made at a given activity.

Two identical tests were administered before and after students were given access to the main body of instruction. The tests covered a sampling of the five main content areas (intervals, scales, key signatures, triads, and dominant seventh resolution), and were administered via the TALL system. The intent of the tests was to give evidence of student learning as a result of the instruction.

It was my hope that I could use the pre and posttests as one of the principle sources for informing questions regarding the effectiveness of TALL in helping students to master the material; however, this source ended up not being as effective for three reasons. First, only nine of the twenty-three students in the course took the pretest, with sixteen taking the posttest. Second, two of the students who took the pretest did not take the posttest. Third, because of some of the technical problems we had with TALL, not all of the data should be considered accurate. For example, because of issues that arose while the posttest was being administered, five of the student scores had to be thrown out completely. Having mentioned that, I will still present the information as one of several evidences, though the other sources I draw from will also be emphasized in supporting my conclusions.

Resource Tracking. As some of the evaluation questions dealt with the relative cost and difficulty of developing instruction using TALL, another source of information I drew from was my own personal tracking of the estimated and actual hours required for discrete pieces of the project.

Interview with Professor. A few of the evaluation questions were best informed by gleaning feedback from Stephen Jones on his experience with the project. Because I was working closely with him throughout, I was able to gather information along these lines during the course of the project. However, after the students had finished the instruction by taking the posttest, I sat down with Dr. Jones for a formal interview, using the evaluation questions to guide our discussion.

Questionnaire. The students were administered a questionnaire that was principally designed to measure attitudes toward their experience. The statements and questions contained within the questionnaire were generated based on the evaluation questions I had derived earlier. Before I administered the questionnaire to the students I gave TALL instructional designers and Stephen Jones a chance to review it, and made a few modifications based on their feedback. The survey made use of a six-point Likert scale, consisting of fifteen separate statements intended to measure student attitudes toward their experience with TALL. I also included four questions that addressed the amount of time they typically spent in a single sitting with TALL, and also gave them a chance to provide more general suggestions and feedback. Twelve of the twenty-three students enrolled in the course responded to the survey. Appendix A contains the full text of the questionnaire.

Focus Group. A focus group session was held after the students had completed all of the instruction (including the posttest). Eight of the twenty-three students that were enrolled in the course attended the focus group. I came to the session prepared with a series of specific questions to address (please see Appendix D for a listing of these questions), which were derived from the evaluation questions. While I was able to get answers to all of the questions I came prepared with, the open nature of the focus group allowed the students to also give useful suggestions and comments that fell outside of the evaluation questions.

Student Interviews. In the preparation stages for this experience I had planned on also performing one-on-one interviews with students after they had completed the instruction. I later decided not to go through with this for several reasons. First, the focus group turned out to be a very open discussion in which all of the students who took an interest were able to voice their opinions. After having had that experience it seemed unnecessary to rehash the same questions with students on a one-on-one basis, especially since the questions wouldn't have changed. Second, the focus group was valuable, but much of the information I gleaned from it was redundant with some of the other sources I was utilizing. Third, I felt like one-on-one interviews would have placed an unnecessary burden on the students, who had already freely given of much of their time to assist with the evaluation. Finally, I was able to get feedback from students by visiting them in class, as well as email correspondence and phone conversations that I had with them, which I hadn't anticipated at the beginning of the project. I feel like these sources of information also helped to supplant what I would have gained through individual interviews.

Evaluation Findings

Primary Question: Given that TALL was originally designed for language learning, how effectively can it be used to teach other topics?

Related Questions and Issues.

1. How effective are TALL content development tools at representing and manipulating information unrelated to human language (e.g., music)?
2. How well does TALL handle non-text-based content (e.g., musical notation)?

Criteria.

1. An expert in the field would determine that the content is instructionally valid.
2. Music learning activities would not be "forced" to be like existing language-learning activities.

3. Content is organized in such a way that it makes sense to an expert in music theory.
4. Content development tools require little or no alteration in order to handle non-language-learning content.
5. Music-based content “fits” into the existing TALL data structure, without alteration of the latter.
6. TALL is able to accurately display musical notation.
7. TALL provides for a flexible display of the musical notation. For example, graphics will scale properly, colors can be modified, etc.
8. Generating musical notation dynamically was not significantly harder than generating it statically.

Data Analysis. There were three areas of data collection that informed these questions: first, my own experience in designing and developing the instruction with Stephen Jones; second, Stephen Jones’ assessment of the content as an expert in music theory; and third, the assessment of Laurence Lowe, another expert in music theory who gave a third-party assessment of the content.

There were two principle aspects to representing the musical content. The first were the audio and visual aspects—displaying musical notes, staves, and ledger lines, as well as playing the audio representation of the musical elements. The second dealt with defining rules that govern where notes are to be positioned in order to fulfill a presented problem. As an example of one of these rules, in the dominant seventh resolution activity a student must resolve the seventh note of the dominant seventh chord to the third note of the resulting tonic chord.

According to Stephen Jones, the TALL tools were generally able to effectively represent musical content. Visual depictions of musical elements (notes, staves, ledger lines) were accurate, as well as the audio depictions of the music (when the music was played by the application). Dr. Jones also commented that concepts and activities used to teach the content fit well into the music theory curriculum, and that they were consistent with other methods used to represent similar content.

Visual depiction of the musical elements fell short in a few areas, but none of them were considered by Dr. Jones and me to be so severe that they would significantly inhibit learning. In some cases notes were positioned too close to the key signature, in which cases sharps and flats on the key signature could have been mistaken for accidentals. When displaying a grand staff (treble and bass clefs), at times notes on the top staff didn't align vertically with notes on the bottom staff. Also, TALL was unable to display a single note intended to represent two choral voices (e.g., both the soprano and alto voices using the same note in the treble clef).

There were also software bugs that affected the accuracy of the representation of the music. For example, I programmed the musical component to be able to play the notes that were represented; however, this didn't always seem to work for students, and on occasion would even crash the software application. There were also a few quirks in the musical component regarding where it would allow notes to be added. For example, often if a student would attempt to add a note head above the staff to an existing note it would actually position the new note to the right of the existing note instead of above it. While none of the software bugs were "show-stoppers," they were at best a minor annoyance to students and likely the source of some frustration.

Comprehensively defining the many rules related to music theory was particularly challenging. As I provided early prototypes of particular exercises to Dr. Jones he was reminded of exceptions to the rules, which I then had to account for in my code. For example, in the dominant seventh resolution activity there were ten different checks that the computer would make to determine if the student had correctly resolved the given chord. Two shortcomings in the system seemed to stem from the complexity in defining these rules. First, on more than one occasion the computer would give inaccurate feedback to the student on a rule they should be following. This problem wasn't too widespread, but it did seem to create a sense of doubt in the students that the software was always providing accurate information. Second, even after a few iterations of Dr. Jones defining rules, my programming them, and him reviewing them, we still seemed to discover rules that we either weren't covering or that we weren't treating correctly. It's very possible that there are still inaccuracies in this regard. Because of the difficulties we had in working with the rules, in my final interview with him, Dr. Jones made the comment that the system seemed to be able to handle more fundamental concepts (such as key signature and triad identification) more adeptly, and that it may be more effective to focus on the more basic concepts.

Discounting bugs that occurred from time to time, the system was able to give accurate feedback to students. However, Dr. Jones made the comment that the feedback could have been better targeted to the specific mistake that the student was making.

The use of "concepts" was a shortcoming of the TALL system in representing our content. TALL centers on the notion of "concepts," which can be generically described as a learnable unit of knowledge. A concept could be as small as a single note on a staff, or as large as an entire music theory course. Concepts can be related to each other in various ways, including hierarchically (called a “whole:part” relationship in TALL). This provides for a flexible representation of learning content; however, a given activity can only center on one concept, which turned out to be somewhat of a shortcoming for us. For example, when testing interval recognition we used the interval itself as the “central” concept, but other important concepts also included the clef (alto, bass, tenor, or treble) as well as the position on the staff where the interval was placed (e.g., whether it was high or low on the staff). In designing the adaptivity of the instruction, it was relatively easy to adapt based on student performance on central concepts—it was less so (though possible) to adapt based on the peripheral concepts.

Laurence Lowe validated that the music displayed by the component I had programmed was correct, and that the content was logically organized. I demonstrated several of the activities for him and allowed him to try a few of them as well. Aside from validating the content, Dr. Lowe also made several helpful observations. He commented that the software might be more suitable for a lower-level music theory course, as those entering the music major would likely already be versed in much of the material we covered. This validated one of the premises underlying our project—that the instructional material was most suitable for entry-level students. Dr. Lowe made the observation that, because the textbook for the course provided very few details and examples of concepts, this type of software could be valuable to many students. He also suggested that we include links to instructional content, so that students could have something to refer to as they were performing the activities. One of the main disadvantages to the software Dr. Lowe noted was that it disallowed the students from writing music notation software by hand, as it traditionally has been done. He conceded that it could be simply that he's from an "older" generation, but he saw the act of clicking and dragging to edit music as taking away from students an important traditional aspect of learning music theory.

Primary Question: As a development platform, how does TALL compare to other tools?

Related Questions and Issues.

1. How do development costs using TALL tools differ from costs using other tools?
2. What kinds of expertise are required to use TALL tools (e.g., programming, content development, knowledge of tools)?
3. How much programming is required for TALL to be effectively used in teaching content outside of human language?
4. If the same instruction were developed using a tool like Flash, how much more difficult would it be? Does TALL provide any inherent advantages over a tool like Flash?

Criteria.

1. Other groups on the BYU campus, such as the Center for Instructional Design and Independent Study, could viably use TALL to develop instruction at a realistic cost. For example, Stephen Jones and a graduate student could develop content beyond this project without any further programming work done and at a reasonable cost.

Data Analysis. My approach to addressing this question is to attempt to answer it myself based on my personal experience in programming and developing the instruction; however, ideally, others, such as those at the Center for Instructional Design, would be able to try their hand at developing similar instruction with the tools. Unfortunately, because the TALL system is still in a formative stage, and because of intellectual property issues, this wasn't a possibility. I realize that this is a less-than-ideal approach, but I'm hopeful that, having worked for both the Center for Instructional Design and Independent Study, I've been able to inform this question based on past experience (and hopefully with a minimum of bias).

On the whole it's my estimation that TALL is a very effective tool for both developing and delivering content. Much of the legwork in designing and creating instruction can be handle by existing TALL functionality. In this particular case the only programming I had to do dealt with representing music in visual and audio forms, and defining rules in the system directly related to music theory. Other features related to developing and delivering instruction, such as constructing activities, sequencing activities, and collecting user data were already found within the system.

In terms of relative costs of using TALL to create and deliver the instruction, the only way to get an accurate measurement of this would be to attempt to generate identical instructional modules using an alternative tool (such as Macromedia Flash or Director). However, given my experience in using both TALL as well as many other similar tools, I would guess that the costs of using TALL would be significantly less than if I had elected to use an alternative tool. TALL was already built to handle many of the underlying features needed in developing and delivering instruction, which I've described previously. Additionally, TALL is written in Java, which is a high-level programming language that is frequently used in software development projects. Because I was able to utilize a high-level programming language many of the features (such as enforcing many of the rules related to music theory) could be programmed relatively easy, which likely would not have been the case had I elected to use a tool like Flash.

Regarding expertise required for this project, familiarity with the TALL tools was obviously a requirement. Having taken part in the programming of much of TALL this was a relatively small obstacle for me, but for someone unfamiliar with the system this could be a bit of a learning curve. In addition to creating the instruction, I had to program a small widget to be embedded in an activity that would allow for the musical content to be displayed. Because TALL is written in Java, proficiency in this language was another requirement.

In order for TALL to be used in teaching content outside of human language, the content must be represented in the TALL system. In this particular case this meant programming a widget that would display and play musical content. At the same time, TALL is designed to handle a broad range of multimedia formats, so it seems that it could handle other subject areas relatively easily (e.g., history or physics) without requiring extra programming.

In regard to developing the same content using another tool, such as Macromedia Flash, it's very possible that much of what I had to program would be prohibitively difficult to program. As I mentioned previously, I was able to do all of the programming in Java, which is able to handle just about any programming task. Tools like Macromedia Flash, though useful for developing things that use simple logic, would likely fall short in being able to enforce the musical theory rules. For example, because Java is object-oriented I was able to create underlying programming structures in code that represented musical content in a form that the computer could understand. As such, the computer could tell, for example, when a student hadn't resolved the seventh note of a dominant seventh chord to the root of the tonic chord. Because Flash is not a high-level programming language, and doesn't allow for the same type of representation of data structures, enforcement of the same types of rules would be extremely difficult. Having said that, if the content were relatively simple, for instance, just representing single notes and key signatures, it's likely that a tool like Flash would be adequate.

Primary Question: How effective is the instruction built for this project?

Criteria.

1. The students feel as though the instruction has assisted them in mastering the concepts.
2. The professor feels as though the instruction has assisted the students in mastering the concepts.
3. The professor accepts this instructional approach as a valid method of teaching.
4. An expert review of the software, done by TALL instructional designers, will verify that the capabilities of the TALL system are being used appropriately.
5. Students are able to retain the information at least throughout the duration of the course.
6. Students feel like they were able to master the material more quickly than they would have been able to without using the software.
7. Students felt that TALL's systematic spaced review method was helpful for long-term retention.
8. Students achieve mastery over concepts handled by the instructional modules, as determined by the professor.

Data Analysis. I am addressing these questions based on data drawn from the pre and posttests, a survey, conversations with the professor, and the focus group.

In order to whittle down the results of the pre and posttests to something useful, I've decided to utilize the scores from only four of the students. I realize that this is a very small number, but these are the only scores that I can say with some confidence accurately reflect the performance of the students. In all cases student scores from the pretest to the posttest went up. The average increase was 23.27%, with the lowest increase being 1.21% and the highest being 32.57% (see Appendix C for the raw data related to the pre and posttests). While there was a rise in the test scores, this data should be taken with a grain of salt given its lack of reliability. Notwithstanding, it does serve as yet another source of information to justify conclusions.

On the whole, students seemed to feel that TALL helped them in mastering the material for the course. After adjusting the questionnaire responses to account for the positive and negative nature of each of the fifteen statements, the average student response was 3.6 (on a 6-point Likert scale).

The first three statements on the survey specifically addressed whether or not TALL helped the students master and remember the material in a short amount of time. On the 6-point scale, the average response to these first three statements (which were similar in what they were intended to measure) was 4.8, which seems to support the conclusion that TALL can be an effective tool for learning.

Dr. Jones commented that it was difficult for him to know how much TALL was helping students given the lack of a convenient reporting system. That is, student performance on activities and assessment instruments wasn't readily accessible to him. I would provide reports to him periodically, but they were insufficient for him to make reasonable judgments about the effectiveness of the software. Also, he made mention of the fact that, because of technical glitches, some of the information may not have always been accurate, which would have tainted student attitudes toward it. In spite of these problems, Dr. Jones did validate the content and overall approach that we took, and mentioned that he would seriously consider using TALL again to teach.

Primary Question: What are student attitudes toward the instruction?

Related Questions and Issues.

1. Is it too easy? Too hard? Too repetitive? Too overwhelming?
2. Do students enjoy learning this way?
3. How would students rank learning using TALL relative to other methods (e.g., from a book, in a classroom setting)?
4. How long do students prefer to work on the computer in one sitting?

Criteria.

1. Students generally found the instruction appealing.
2. Students felt that the software was easy to use and inviting.
3. Students would use the software again.
4. Students would recommend the software to a friend who wanted to learn the same thing.

Data Analysis. Based on the survey results it's apparent that the activities were not too difficult, but may have been somewhat repetitive; though student responses were not strong enough in either of these areas to merit significant revamping of the activities.

Four statements addressed the technical problems that students may have encountered while using the software. The average response on these four was 3.25, which was somewhat surprising to me. I had anticipated that students would have considered this to be one of the principle negative aspects to their experience, but apparently students didn't consider the problems to be that severe.

The three statements intended to address student preferences toward various methods of instruction relative to TALL were fairly mixed, though it's apparent that students generally prefer classroom or private instruction over working with TALL. When asked to rate learning from a textbook over using TALL, though, the average student responded with a 3, and gave all 6 points of the Likert scale at least one response.

Three statements were designed to measure general attitudes of students toward their experience. They dealt specifically with: whether or not students would continue to use TALL, whether or not they would recommend it to a friend, and if they enjoyed using it. The average response to these statements was 4, which supports the conclusion that, on the whole, students seemed to have a positive experience.

Student responses to another survey question indicated that they would typically spend an average of 54 minutes in a single sitting using TALL.

In terms of study methods students used aside from TALL, nearly all of the students cited their textbook. Students also used notes from class, personal sessions with Dr. Jones, fellow students, flashcards, and homework.

In general, the focus group I held supported the same conclusions drawn from the survey. Students expressed that there was a good balance between easy and hard activities. They felt that it was convenient to be able to use TALL on their own time.

Technical problems caused stumbling blocks, but during times when the software was working as it was designed to, students felt like it was instructionally effective. Several students highlighted the usefulness of getting immediate feedback on activities. As in the survey, students mentioned that their textbook was one of the principle methods of study they used. They also indicated that classroom instruction and feedback from the professor on homework was useful.

While student feedback in the survey seemed to support the general evaluation conclusions, one of the most useful aspects of the session was the specific feedback students provided. For example, they would have liked to have some kind of an indicator as to their progress in the system (e.g., some kind of a gauge to show them how many activities are left). They also suggested that a timed drill might be effective.

Another source of information, that I hadn't necessarily planned on at the beginning, was the times I was able to visit the class throughout the term. My intent for these sessions was principally to allow students to report problems with the software, but they also gave the students a chance to provide general feedback on the system. Again, the feedback that students expressed during these times supports the overall evaluation conclusions—there were technical problems with the software, but, on the whole, students felt like it was an effective method of instruction.

Primary Question: Does the software perform such that it supports (or is transparent in) the learning process, or does it prove to be a hindrance?

Criteria.

1. The software runs at an acceptable speed.
2. The application doesn't crash or freeze.

3. It is generally found to be usable by students.
4. The software is able to run within reasonable memory constraints.
5. The software doesn't take too long to load.

Data Analysis. This question is partly addressed under the previous question: "What are student attitudes toward the instruction?" On a number of occasions, technical issues caused problems resulting in sluggish performance, abnormally high repetition of activities, malfunctioning audio, failure to recognize when the learner had obtained the correct answer, and even the crashing of the program. Though it was not voiced directly to me by students, Dr. Jones reported that on numerous occasions students reported a high degree of frustration with these problems that would have been a hindrance to their learning.

The machines we ended up running the software on were robust enough that the application was able to load quickly and run within reasonable memory constraints.

Primary Question: How reusable is the software and instruction?

Criteria.

1. The instruction could be used in other music courses at BYU (e.g., Music 190).
2. The musical notation widget developed for this project could be effectively used in developing other instructional modules.
3. A somewhat tech-savvy content expert could use the musical widget to develop instruction.
4. The professor in this project would want to use it for future projects.
5. The professor in this project comes up with other areas where he would be interested in using TALL to teach.

Data Analysis. Dr. Jones expressed interest in using TALL for future projects, provided we were able to clean up the technical problems we encountered while using the software. He also mentioned that it may be more effective to use TALL in a more fundamental music theory course, given the difficulty we had in programming some of the more complex rules.

I asked another programmer on the TALL project, Robert Sweeney, to review the software code I had written in order to give an assessment of reusability. Robert mentioned that the source code files were large, which is generally an indicator that it's not very modular (which is generally required for reusability). He suggested breaking them up into smaller pieces so that they would be more easily understandable and accessible from code that might link to it. Notwithstanding, Robert did comment that the code could have a high degree of reusability because it "has so much functionality crammed into it." That is, if it were broken up into smaller pieces, the various algorithms could be useful to other programmers.

Primary Question: Is this system implemented and utilized as intended?

Related Questions and Issues.

1. What percentage of the instructional material are students actually viewing?
2. Are there any technical issues regarding the installation and execution of the software that deviate from the original plan?

Criteria.

1. Students view as much of the instructional material as they feel is helpful to them.
2. The software is installed and executed according to the deployment plan.

Data Analysis. Data collected by the TALL system provides a good measure as to how much students used the system. This was one of the more interesting aspects of the experience, as some students decided not to use the system at all, while others made extensive use of it. Five of the twenty-three students made almost no use of the software. Other students took the time to review certain concepts over a hundred times. One girl, for example, went through 232 separate activities relating to spelling and recognizing intervals. Seven other students reviewed those same activities over a hundred times. It should be noted, however, that the higher frequency of these activities was most likely a result of a bug in the software that forced the student to abnormally repeat them. I was directly involved in deploying the software, and took the time to review it on several machines in order to ensure that it had been installed and was functioning correctly.

Technology Evaluation

A second significant evaluative aspect of my project dealt with technology selection. Some programming was required in order to represent musical notation in TALL (which it doesn't inherently do). As such, I had to find or build an existing software package to render the notation. Because the content was being developed in the context of the larger TALL system, there were already some criteria that it had to meet, which had been determined prior to my project. Using these criteria, and after discussing the matter with Bret Elzinga, who is in charge of the technical aspects of TALL, we arrived at the following criteria for selecting a software package:

1. The package is written in the Java programming language (the language TALL was developed in.
2. The license the software is released under is not too restrictive.

a. It doesn't require that the source code to TALL be publicly disclosed.

b. It preferably does not carry a cost.

c. It allows for the manipulation and extension of the source code.

3. The package seems to be credible.

a. It is used in other applications.

b. A review of the source code deems it to be well designed.

4. It must not require a persistent connection to the Internet.

5. It has no unreasonable dependencies on other third-party software packages.

6. Its feature set is adequate, either out of the box or with a reasonable amount of extra programming, to support the instructional requirements.

a. Must be able to handle the following musical elements:

(1.) Treble, bass, alto, and tenor clefs

(2.) Double sharps and flats

(3.) Triads

(4.) All key signatures

7. The package performs well (in terms of speed).

8. Its usability level is determined to be adequate for the audience.

9. It has a reasonably low memory footprint.

After scouring the Internet for possible packages to use, I arrived at four. Based on the criteria outlined above, I ranked them and listed out the advantages and disadvantages to each:

1. XEMO (http://www.xemo.org/)

a. Advantages:

(1.) Produces immaculate notation

(2.) Released under an open source license that fits our criteria

b. Disadvantages:

(1.) Would require a fair amount of programming on our behalf

(2.) User interface may be somewhat clumsy

2. JMSL (http://www.algomusic.com/jmsl/)

a. Advantages:

(1.) Sports a relatively complete feature set

(2.) Would likely require the least amount of programming

b. Disadvantages:

(1.) Does not handle double sharps or double flats

(2.) Does not easily allow changing of the key signature

(3.) Would require payment of a licensing fee for long-term use

(4.) Source code is unavailable in order to make fixes and facilitate integration

(5.) User interface is clumsy and unattractive

3. GUIDO (http://www.informatik.tu-darmstadt.de/AFS/GUIDO/)

a. Advantages:

(1.) Produces immaculate notation

(2.) Released under an open source license that fits our criteria

b. Disadvantages:

(1.) Would require a fair amount of programming on our behalf

(2.) User interface would likely be somewhat clumsy

    (3.) Requires a persistent network connection

4. jMusic (http://jmusic.ci.qut.edu.au/)
    a. Advantages:
        (1.) Relatively easy to integrate into TALL
        (2.) User interface is more or less intuitive
        (3.) Released under an open source license
    b. Disadvantages:
        (1.) Doesn't support chords, sharps and flats combined in a given chord, double-sharps, double-flats, or tenor and alto clefs
        (2.) Released under the GPL, which may make it unusable in the long term

After about 16 hours of experimenting with the various packages I settled on XEMO. The biggest disadvantage to this package was that it likely required the most amount of programming (relative to the other libraries) in order for it to meet our instructional requirements; however, it was the only one that met enough of the other criteria that it was truly viable. Ultimately the package turned out to be adequate, but not until I had invested significant time into evolving it to the point that it supported our instructional design. In the future I think it would be a good package to continue to use, assuming sufficient resources can be devoted to it to ensure that it functions as it is designed to.

Meta-evaluation

In an ideal situation a third-party would perform a meta-evaluation, but lacking the resources I've made the determination to provide this evaluation myself.

Additionally, however, members of my committee (one of which is an expert in evaluation) will be reviewing my work, and, while it may not be considered a formal meta-evaluation, this does allow a third-party to make a judgment on the merit and quality of my work.

When evaluating one's own work, subjectivity and bias will obviously play into the final result. In an attempt to minimize my own personal bias as I judge my work, I've decided to utilize the Program Evaluation Standards (http://www.eval.org/EvaluationDocuments/progeval.html) as a framework, with the hope that they will provide a more formal guide then I may have followed otherwise. What follows is a brief description of each of the standards, along with my assessment on how I addressed each.

Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1 Stakeholder Identification--Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed. At the outset of the evaluation I made a conscious choice about whom I was going to include and exclude as stakeholders in my evaluation. Having done that, I consulted with each of these individuals in order to formulate evaluation questions. Taking this approach helped to ensure that their needs, issues, and questions would be addressed.

U2 Evaluator Credibility--The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance. My background in evaluation isn't extensive, but I believe it was adequate for this particular task. I've formally studied evaluation in a university setting, and completed an honor's thesis (somewhat like a master's thesis for undergraduates) that was an evaluation. Dr. David Williams and I later published an article that originated from this honor's thesis in a peer-reviewed journal on evaluation. Additionally, prior to carrying out the evaluation I drafted up a formal plan for it, which was reviewed and approved by an expert in evaluation.

U3 Information Scope and Selection--Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders. As much as possible, I tried to draw from a broad range of sources in order to answer the evaluation questions. As described earlier in this document, taking such a "triangulated" approach allowed me to compensate for weaknesses in some sources, and capitalize on strengths of others. Throughout my data collection and analysis I centered my activities on the evaluation questions that were generated by the stakeholders, allowing me to respond to their individual needs.

U4 Values Identification--The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear. In this document I have tried to describe in detail the approach I took to my evaluation, along with my justification for doing so. Being the meta-evaluator of my own work, this particular standard is a little difficult to address, simply because it's more difficult for me to realize when my own subjectivity may be influencing my inferences. Notwithstanding, with each conclusion I've drawn I've also tried to provide a clear justification based on the data I collected. In addition, as I've noted previously, stakeholder values and questions guided the evaluation and provided the basis for data collection and interpretation.

U5 Report Clarity--Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood. Again, I've made an attempt to provide as much detail as seemed appropriate on my methods, activities, and findings. In terms of the context of the evaluation, I described, for example, how the instructional content we were developing was being done within a larger system, but that the evaluation was not intended to address that larger system.

U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion. In the case of my evaluation there are no fixed deadlines in terms of when the findings need to be utilized. As such, I haven't felt the need to provide interim findings to the stakeholders. One area where this might come into play would be the statistical reports I provided to Stephen Jones that were generated by the TALL system regarding the performance of the learners. Also, one of the instructional designers requested to know the results of the questionnaire, so I allowed him (as well as the other instructional designers) access to the raw results.

U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased. Probably the principle way I feel I complied with this standard was to center all of my evaluation activities, findings, and analysis on the questions and criteria that were generated by the stakeholders. In doing so I will hopefully be addressing specific needs and interests that they have, which should allow them to put the evaluation results to use. After delivering to the stakeholders the results of my evaluation, I plan on having at least one formal meeting in which we can discuss them, and talk about ways that they might be utilized.

Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures--The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained. It seems that this standard would come into play most in data collection procedures. In my particular case I tried to be sensitive of the time constraints of those who were providing data to me. Meetings and interviews were kept as short as possible, while still allowing time for me to gather adequate data to address the evaluation questions. In the case of working with students, the times when I was invited to attend class in order to get feedback I tried to keep disruption to a minimum by taking only a small portion of time at the beginning of class. Toward the end of the evaluation I actually made a conscious decision to forgo one of the data collection procedures I had initially planned on: one-on-one student interviews. One of the reasons I decided not to do this was that I felt the students had already contributed sufficient information to answer the evaluation questions, and that interviewing them further would be an unnecessary waste of their time, as I was confident that it wouldn't provide me significant information that I didn't already have.

F2 Political Viability--The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation *may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.* I believe that the political impact of the evaluation should be relatively low. The principle stakeholders I interviewed (Stephen Jones and TALL instructional designers) work together on the TALL system, and should have aligned interests in terms of it.

F3 Cost Effectiveness--The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified. Financial resources for the evaluation were low. I did request $100 in funding from the Instructional Psychology and Technology department to fund gift certificates, which I provided to the students who took part in the initial one-on-one interviews. Outside of that, the only cost was time required by myself and those I collected data from.

Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation--Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants. The entire structure of my evaluation has centered on identifying and seeking to cater to the needs of those that I initially identified as stakeholders. The intent of this approach has been to increase the chances that the evaluation findings will be used.

P2 Formal Agreements--Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

Prior to beginning the main portion of my project I drafted up and had approved a prospectus, which contained an outline of the evaluation I intended to perform.

P3 Rights of Human Subjects--Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects. Because of time constraints prior to working with students, I was unable to receive formal IRB approval for my project; however, I did complete a small tutorial on IRB compliance standards, which were all adhered to throughout my project.

P4 Human Interactions--Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed. See P3.

P5 Complete and Fair Assessment--The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed. Given my closeness to the project this standard is difficult to assess. I have strived to be as impartial as possible, but objectivity is difficult when assessing one's own work. Because my intent has been to maximize the usability of my findings, I have tried to point out both strengths and weaknesses that should be addressed.

P6 Disclosure of Findings--The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results. My plan is to make available all of my reports and raw data to the stakeholders. I will also hold a formal meeting with the instructional designers I worked with in order to discuss the findings and how they might be used.

P7 Conflict of Interest--Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results. This is also a bit difficult to address, given my level of involvement in the project; however, I have no vested financial interest and little political interest in the use of the evaluation. As such, it seems unlikely that a conflict of interest would negatively affect the evaluation.

P8 Fiscal Responsibility--The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate. The small amount of money that was spent has been reported in this document.

Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1 Program Documentation--The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified. In this document I've strived to provide a detailed description of the instructional product, as well as its implementation. Additionally, I've included screenshots of both the activities themselves, as well as a visual of a sample activity sequence.

A2 Context Analysis--The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified. I mentioned in the beginning that the evaluation was happening within the larger context of TALL, and provided a short history on it. Additionally, I tried to provide a thorough description of the setting of the implementation, including notes on problems that occurred in the course of it.

A3 Described Purposes and Procedures--The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed. I have included an entire section detailing my procedures and data collection methods. I believe it contains sufficient detail, but this will also be checked by my committee.

A4 Defensible Information Sources--The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed. I've tried to describe in detail each of the sources of information I made use of in my evaluation. Because of the "triangulated" approach I took in addressing questions utilizing more than one information source, I feel as though weaknesses in certain information sources were generally compensated for through the strengths of others.

For example, in addressing the evaluation question, "How effective is the instruction built for this project?" I made use of four information sources: pre and posttests, a survey, conversations with the professor, and the focus group. The pre and posttests had the advantage of providing quantitative data to back up claims (which are especially useful in giving evidence of learning having taken place), but they were also quite weak because of the reliability problems caused by TALL's data collection mechanism. The other three sources hopefully helped to compensate for this weakness, even though they were qualitative sources of information.

A5 Valid Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use. I believe the "triangulated" approach I used addresses this standard. It seems very unlikely that a single data collection procedure could be adequate simply because all will have pros and cons. Using multiple collection methods allows for strengths and weaknesses of the various procedures to balance each other out.

One weakness that would likely fall into this area would be the fact that I didn't pilot my questionnaire with real students before administering it to the students in the study. I did allow stakeholders to review it, but, ideally, I would have tried it out on actual students as well. If I had taken the time to pilot the questionnaire with students it's more likely that their feedback (probably gathered via formal interviews) would have instigated changes in the questions I was asking such that they would better get at the information I was trying to draw out.

A6 Reliable Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use. On obvious weakness in regard to reliability that I've mentioned previously was the data collection mechanism within TALL that was to gather student performance information. Because of software bugs this feature of TALL never worked consistently.

In regard to expert reviews, it would have been ideal to consult regularly with experts throughout the design process; however, because of time constraints this simply wasn't realistic. It should be noted, though, that this lack of regular consultations could have a direct impact on the reliability of the information source.

I feel as though my interviews with the professor should be regarded as reliable for the fact that they happened on a regular basis. Feedback I received from Dr. Jones was generally consistent.

In regard to data collection of student feedback (e.g., the questionnaire and focus group) its difficult to judge the reliability of these sources, simply because they were only performed once. In an ideal situation we would have held a few focus groups throughout the term in order to collect data on multiple occasions, thereby increasing the reliability of the information source.

A7 Systematic Information--The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected. This is one area that could have been addressed more thoroughly. While the project was underway I should have spent more time analyzing the data that was being collected through the TALL system so as to make adjustments when it became apparent that much of the data was inaccurate. I'm also anticipating that as Stephen Jones reviews the evaluation he will be able to point out inaccuracies in the areas that he was directly involved with.

A8 Analysis of Quantitative Information--Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered. See A7 for a partial answer to this question. The quantitative data in my study were the data collected by the TALL system and most of the survey results (the non-open-ended questions). These data were analyzed to the level of detail that seemed appropriate, but my success in selecting a good level of detail will also need to be determined by those that use the evaluation findings.

A9 Analysis of Qualitative Information--Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered. See A7 for a partial answer to this question. Qualitative information was gathered through personal interviews, the focus group, and the open-ended questions in the survey. Again, I've done my best to analyze the data as thoroughly as seemed appropriate, but that is obviously a subjective determination on my part.

A10 Justified Conclusions--The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them. I have attempted to justify all of my findings and conclusions by citing specific evidence from my data sources.

A11 Impartial Reporting--Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings. This is another difficult standard to address, given my closeness to the project. I've tried to protect against personal bias by justifying any conclusions via specific evidence drawn from my data sources.

A12 Metaevaluation--The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses. This has been done by myself, but will be addressed, to a degree, by others. This evaluation will be formally reviewed by an expert in evaluation who is obviously very familiar with industry standards. It will also be reviewed by other instructional designers who have formal training in evaluation.

Budget and Schedule

Schedule

Implementation of the instruction had to be completed before Spring Term of 2003 (which began April 29), as that is when the course was offered. Here is the basic schedule I followed:

1. Feb. 1 to Mar. 21. Perform research, create preliminary design (including such products as a scope and sequence, a task analysis, and a detailed evaluation plan), and perform initial programming in preparation for a prototype.
2. Mar. 22 to Mar. April 3. Create a prototype sequence of instructional modules.
3. Apr. 4. Perform the first set of one-on-one interviews.
4. Apr. 5 to Apr. 10. Adjust existing modules based on feedback, and expand the path to 75% of its ultimate length.
5. Apr. 11. Perform the second set of one-on-one interviews.
6. Apr. 12 to Apr. 24. Complete first full version of instructional modules. Perform the third set of one-on-one interviews.
7. Apr. 25 to Apr. 28. Finalize instruction. Handle final arrangements for deployment of the software.
8. Apr. 29 (first day of Spring term). Introduce the software to students. Administer the pretest to those interested in using the software.
9. May 21. Hold a focus group meeting to get feedback from students on the system.
10. May 23. Perform interviews with a sampling of the students.

11. June 16 through Aug. 22. Collate and organize feedback from the implementation. Write report on the development project.

Budget

All of the development and programming of the instruction was done by me. I consulted briefly with a graphic designer once the functionality of the activities had been established, which took less than an hour. Table 1 is a breakdown of the tasks that were preformed, along with the project and actual hours.

Table 1

Projected vs. the actual hours spent by task

Task	Projected	Actual
Programming		
Note adding, editing, and deleting (foundational programming)	24	52
Interval recognition	4	6
Interval spelling	6	12
Interval inversion construction	3	5
Spelling enharmonically equivalent intervals	4	6
Scale identification	5	5
Scale spelling	6	8
Key signature identification	3	3
Triad identification	8	6
Triad spelling	8	16
Dominant seventh resolution	6	18
Designing and building activities	10	40
Designing and building concepts and concept types	2	2
Designing and building adaptive paths	6	8
Evaluation		
Perform one-on-one student interviews prior to launch	6	5
Perform focus group meeting	1	4
Prepare and administer questionnaire	2	2
Total	111	194

On the whole, the largest discrepancy between the projected and actual hours falls in the area of content development and programming. At the time I created estimates to program each of the features I hadn't looked extensively through the source code of the music library we decided to use. On the surface the library seemed to do most of what we needed; however, as I delved more deeply into the source code I realized that the library handled the visual aspects of rendering music, but little else. An analogy might be looking at a sports car that lacks most of the parts needed to make it run. On the surface it looks relatively complete—you can sit inside and examine the interior, but until you take it out for a spin it's hard to know exactly what it can do. Such was the case with the XEMO library in that it visually rendered musical elements beautifully, but lacked the underlying data structures that allowed a computer to understand and manipulate those elements. As such, I ended up performing much of the programming to allow for this to happen.

In regard to building activities, I attribute much of the time discrepancy to my lack of prior experience in building them. Not only was it a bit of a learning curve for me, but I also ended up having to modify the code of my musical component so that it worked correctly with other components that comprised the activity. TALL allows for disparate components to communicate with each other within an activity, but connecting them up correctly took a bit more work than I had anticipated.

The remainder of the estimates was more or less on target, with the focus group being the only item that took significantly longer than I had anticipated. In this case I had failed to take into account the amount of time it would require to organize the meeting, make adequate preparations, and collating some of the data immediately following the meeting.

Critique

In retrospect, I'm satisfied with the outcomes of our project. In the heat of it, it often seemed as though there were so many problems surfacing that student learning would be severely negatively affected. In looking back over it, though, there were a number of positive aspects to the project that are worth highlighting, along with those that didn't turn out so well.

On the whole, I think that many of the weaknesses of our project stemmed from the lack of resources (both time and manpower) we had at our disposal. Our time prior to the implementation consisted of about three months, which turned out to be just sufficient to get the project off the ground. In the process there were many supporting activities that we were forced to set aside. Given unlimited (or at least more) time and resources it could have made a significant difference to hire at least two part-time assistants—one to help with the programming, and another to perform software testing. This would have freed me up to focus more on the instructional design and evaluative aspects of the project.

In my estimation, two of the principle weaknesses of our project were a lack of quality assurance activities and a lack of technical support for students. It is often said that computer programmers make very poor software testers, and I think this was especially true in our case given that I was the one that had programmed a significant portion of the software. I think it would have changed things dramatically to have someone with a background in music theory spend time prior to the implementation reviewing the activities for content problems and software bugs. Taking this step would likely have reduced the number of technical problems the students encountered during the implementation. On a related note, I think students would have felt much more confident in using the software had we been able to provide better technical support. Given more time, I likely would have set up “office hours” when I could have been in the labs to assist students when they encountered problems.

Another significant weakness of our system was a lack of reporting tools for the professor. Dr. Jones had a very difficult time knowing how well students were doing in the system, or even if they were using it, because of a lack of reporting features. In any type of classroom setting it’s important for a professor to know whether or not students are understanding the material, so as to provide help when needed. The TALL system is designed to gather extensive data, but we just happened to be using it at a stage in its development where that data was not readily accessible. Unfortunately, even given more time this is probably not a portion of our project that we could have had much control over. TALL is being developed on a schedule that is subject to the needs of a variety of customers, of which we were only one.

Given more time, I think it also would have improved the instruction if I could have spent more personal time reviewing the basic concepts of music theory we covered. It was fortunate that Dr. Jones was able to spend so much time in collaborating with me in designing activities, but I think I could have been more useful had I known the subject better. Related to this, it also would have been helpful if we could have spent more time reviewing together some of the “rules” of music theory. For example, on the triad and dominant seventh resolution activities, we went through multiple iterations of design and revision as we discussed and discovered rules that they system needed to cover.

To some extent, I think we simply tried to cover too much material. This turned out to be somewhat of a catch-22—Dr. Jones expressed that, for the project to be genuinely helpful to students taking his course, we would need to cover a minimum number of topics. Given the time constraints we were under, though, it became difficult to cover all of those topics at a high level of quality. As such, we probably erred a little too much on the side of covering the material at the expense of the quality of the software itself.

In spite of all of the weaknesses, I feel that there were several important strengths to our project. First, I'm satisfied with the fact that we focused on evaluation throughout the life of the project. From the very beginning I was consulting with stakeholders and making efforts to gather data that would inform their questions and interests. Evaluation also played a significant role as we iteratively designed activities and programmed the software. Also, I feel that taking a triangulated approach by gathering evaluative data from multiple sources allowed me to support well inferences I was able to draw.

The one-on-one student reviews that took place prior to the implementation strike me as an especially significant portion of the project. Based on my experience, I'm convinced that one of the best ways to improve instructional design is by having it reviewed by actual end-users. As designers we may be blind to aspects of our instruction that act as stumbling blocks to student learning. Features as simple as a confusing icon on a button can be a source of unnecessary student frustration. As the one-on-one interviews I performed were so helpful, this is definitely an area that, given more time, I would have focused more on. It would have been especially helpful to have students spend more time reviewing activities once they had been sequenced into adaptive paths.

We would have been able to focus more on issues such as determining the ideal length of time that should pass before students are required to review material previously covered.

In spite of some of the shortcomings, I feel a strength of our project was that we were able to build our instruction on an existing system that provided a feature-rich framework for delivering the material. The amount of programming I had to do was a drop in the bucket when compared with the amount of time that had already gone in to designing and building the TALL system. TALL made it simple to provide features such as targeted remedial feedback, which I feel was another strength of the system.

In spite of the fact that we may have extended ourselves a bit too much by attempting to cover too much material, I feel like we were able to be very specific about the requirements of the system. This would apply both to the topics and sub-topics we had outlined, as well as the feature requirements we established for the musical software component. Having clearly defined requirements made the instructional design and programming significantly smoother than it would have been otherwise.

Conclusion

“The software should take away from the teacher, at least to some extent, the least exciting part of classroom teaching, namely the necessary drill and practice, and subsequent evaluation” (Decoo, 1994, p. 154). The statement sums up well the intent of our project. In spite of the criticism drill and practice often receives, it can provide an invaluable supplement to learning, when used appropriately.

One of the more significant conclusions I would draw from my experience is that evaluation should play a central role in any instructional project. It was apparent to me that evaluation played a significant part not only in increasing the effectiveness of our design, but also in giving strength to the inferences I was able to draw from the experience. Evaluation plays a key role in ensuring and providing evidence of the successfulness of any instructional project.

It has also become more apparent to me just how personal an experience learning is. Students receiving identical instruction can respond in significantly different ways. This fact must be taken into consideration when designing computer-assisted instruction such that varying learning styles and preferred paces can be supported.

The experience our students had in interacting with the software also underscores the significance of quality assurance in implementation. The instructional material and content may be exceptional, but if the software doesn’t perform as it was designed to, students simply won’t learn as effectively as they could. This is a difficult balance is it is often difficult to anticipate how long quality assurance activities will require. Designers need to be flexible in allowing for some aspects of the project to be scaled back so as to allow for sufficient focus on assuring a high level of quality.

Notwithstanding the technical problems that students encountered while using the system, I was pleased to find out that they felt like it genuinely helped them in their learning. While it seemed like I was constantly trying to field and resolve bug reports with the software, I also received positive feedback from students during the implementation, and especially afterward in the focus group and questionnaire. It's rewarding to know that our efforts apparently had a positive impact, however small.

References

Berz, W. L., & Bowman, J. (1995). An historical perspective on research cycles in music computer-based technology. *Bulletin of the Council for Research in Music Education.* 126, 15-28.

Decoo, W. (1994). In defence of drill and practice in CALL: a re-evaluation of fundamental strategies. *Computers and Education.* 23(1/2), 151-158.

Duckworth, W. & Brown, E. (1978). *Theoretical Foundations of Music.* Belmont, California: Wardsworth Publishing Company, Inc.

Kuhn, W. E. (1974). Computer-assisted instruction in music: Drill-and-practice in dictation. *College Music Symposium.*12, 89-101.

Kuhn, W. E. (1967). Computer-assisted teaching: a new approach to research in music. *Computer-Assisted Teaching*. 11, 1-13.

Merrill, P. F. (1986). *Computers in Education*. Englewood Cliffs, NJ: Prentice-Hall.

Merrill, P. F., & Salisbury, D. (1984). Research on drill and practice strategies. *Journal of Computer Based Instruction*. 11(1), 19-21.

Parrish, R. T. (1997). Development and testing of a computer-assisted instructional program to teach music to adult nonmusicians. *Journal of Research in Music Education*. 45(1), 90-102.

Rogers, M. R. (1984). *Teaching approaches in music theory : an overview of pedagogical philosophies*. Illinois: Southern Illinois University Press.

Salisbury, D. F. (1990). Cognitive psychology and its implications for designing drill and practice programs for computers. *Journal of Computer Based Instruction.* 17(1), 23-30.

Salisbury, D. (1988). Effective drill and practice strategies. In D. Jonassen (Ed.), *Instructional designs for microcomputer courseware* (pp. 103-124). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vazquez-Abad, J., & LaFleur, M. (1990). Design of a performance-responsive drill and practice algorithm for computer based training. *Computers and Education.* 14(1), 43-52.

Webster, P. R. (1992). Computer-based technology and music teaching and learning. In Higgins, W. (Ed.). *Handbook of research on music teaching and learning* (pp. 416-439). New York: Schirmer Books.

Appendix A: Student Questionnaire

Directions: Indicate your level of agreement to the statements below.

1. Using TALL helped me to master the concepts of Music 195.

   a. Strongly disagree

   b. Disagree

   c. Somewhat disagree

   d. Somewhat agree

   e. Agree

   f. Strongly agree

2. Using TALL helped me to better remember the concepts I studied in class.

   a. Strongly disagree

   b. Disagree

   c. Somewhat disagree

   d. Somewhat agree

   e. Agree

   f. Strongly agree

3. I was able to master the concepts I studied in class more quickly because of TALL than I would have without it.

    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree

4. The activities I worked on in TALL were too difficult.

    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree

5. The activities I worked on in TALL were too repetitive.

    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree

6. I would prefer to use TALL over a textbook to study music theory.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

7. I would prefer to use TALL over classroom instruction to study music theory.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

8. I would prefer to use TALL over private instruction to study music theory.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

9. Given the opportunity, I would continue to use TALL for long-term retention.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

10. I would recommend TALL to a friend taking Music 195.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

11. TALL performed at an acceptable speed.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

12. TALL crashed or froze frequently.
    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree
13. TALL was easy to use.
    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree
14. TALL took too long to load activities.
    a. Strongly disagree
    b. Disagree
    c. Somewhat disagree
    d. Somewhat agree
    e. Agree
    f. Strongly agree

15. I enjoyed using TALL to study music theory.

    a. Strongly disagree

    b. Disagree

    c. Somewhat disagree

    d. Somewhat agree

    e. Agree

    f. Strongly agree

Answer the questions below:

16. How long would you typically spend working on TALL in a single sitting?

17. Aside from TALL, what methods did you use to study for Music 195?

18. What did you like about using TALL?

19. What suggestions for improvement do you have?

Appendix B: Raw Questionnaire Results

The first fifteen questions of the questionnaire were statements that students were asked to rank on a 6-point Likert scale. Table 2 contains the raw results.

Table 2

Coded responses by question

	Question														
Student	1	2	3	4[a]	5[a]	6	7	8	9	10	11	12[a]	13	14[a]	15
1	6	6	6	1	3	3	1	2	5	6	3	1	6	2	6
2	5	5	5	2	4	5	1	4	4	5	4	4	5	2	5
3	4	5	4	2	4	6	3	4	4	5	4	2	4	3	4
4	5	5	5	2	4	3	2	1	4	5	4	5	3	4	5
5	4	4	4	2	5	2	2	2	4	4	4	4	4	3	4
6	5	5	4	1	3	5	4		4	5	3	2	4	3	5
7	5	5	5	2	2	2	1	4	3	4	5	3	5	2	5
8	4	4	4	3	5	1	1	1	3	4	5	5	4	2	3
9	4	4	3	3	6	1	1	1	3	4	4	6	2	4	2
10	5	5	5	4	4	2	1	1	5	5	3	5	4	4	5
11	5	5	6	1	3	4	2	2	5	5	3	5	5	5	5
12	6	6	5	1	3	1	1	1	6	6	4	4	4	4	6
Avg:	4.83	4.92	4.67	2.00	3.83	2.92	1.67	2.09	4.17	4.83	3.83	3.83	4.17	3.17	4.58
Std. Dev.:	.72	.67	.89	.95	1.11	1.73	.98	1.30	.94	.72	.72	1.53	1.03	1.03	1.16

[a] These statements were written with a negative orientation. For positive items, 1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=somewhat agree, 5=agree, 6=strongly agree. For the negative items, 6=strongly disagree, 5=disagree, 4=somewhat disagree, 3=somewhat agree, 2=agree, 1=strongly agree. The coding of the negative items were reversed prior to scoring.

Question 16 of the questionnaire asked how long students would typically spend for a single session working with TALL. Table 3 contains the raw answers.

Table 3

Responses to question 16 of the survey

Student	Time Spent
1	30 minutes
2	30-45 minutes
3	30 min - 1 hour
4	1 hr.
5	1hour
6	30 min
7	1 hour
8	>1 hr.
9	75 minutes
10	1 hr
11	1 hour
12	1. hour

The free-form answers for questions 17-19 are given below (the grammar has been slightly edited in some cases):

Question 17

1. Classroom study, textbook, individual review of class lectures, group study
2. Textbook and classmate.
3. Homework
4. Textbook
5. Class discussion and homework assignments out of our theory book.
6. Composition exercises were useful. Textbook lessons constituted the majority of my study.
7. Just doing the homework exercises

8. Textbook
9. Flashcards and the theory book, doing extra homework problems (from the book)
10. Textbook, notes from class
11. Homework review, textbook review
12. Textbook for concept, professor for personal help, fellow students, practice on the piano

Question 18

1. TALL didn't waste paper, I could use anytime (well, whenever the lab was open), it was quick, easy and straightforward.
2. Good practice and reinforcement.
3. Immediate feedback
4. Good way to practice skills
5. I guess it reinforced in my mind what we had learned in class. I feel like when I have to write out all the notes on paper I tend to learn and associate different concepts better than when I just have to move notes around on the computer.
6. I thought it was great to have the drills to help reinforce the concepts. It was also nice that I could go at my own speed, and on my own schedule.
7. The ability to hear what was going on
8. Quite honestly, I didn't like using TALL. It was awkward and froze when you tried to listen to the examples, the notes were hard to move, or impossible, and it just wasn't very user friendly. I got too frustrated to enjoy using TALL or to think that it helped in any way.
9. I liked being able to hear the examples with TALL, and it was excellent practice.

10. That it gave instant feedback after each problem.
11. It tracked my understanding. Repetition engrained the learning. I liked the immediate feedback

Question 19

Is there anyway we could get it on blackboard? Before we got mid-way through the semester, we 'outgrew' TALL. More skills would be nice. I would have liked to use it for chords V, IV, ii, cadences, etc. etc.

1. More and comprehensive feedback. Ability to practice. If I make a mistake I would like to back up and try it again. If I repeatedly get something wrong I would like some helps as to why.
2. Prizes!
3. Clean up the program; tie up all the loose ends so that it ran smoother.
4. Lessons on the concepts would be great, then we could have something to refer to if we aren't sure on a particular concept. Also, some form of instruction would enable the program to be used completely independently from any other teaching tool.
5. Just continue to iron out the bugs and it should be a really solid program
6. If you offer the choice of listening to the examples, make the sound work. I could listen to it a couple of times, then it quit and every time I tried after that the program froze. AGH!!! It was a very frustrating program. Make the notes move easier and when you click on an accidental (sharp or flat) or a note, make it highlight on the first click. I sometimes had to click on what I wanted ten times for it to highlight so I could continue with the program.

7. It was very frustrating whenever it would freeze and some of the written instructions to the exercises confused me.
8. One thing that bothered me about tall, was that the note button would automatically be highlighted after placing a sharp or a flat. It felt unnatural to me to have to keep re-pushing the sharp button over and over again. Other than that, I really liked tall. Great job.
9. Immediate feedback on the posttest. Easier to make note entries above or below the staves. A note when TALL saw a trend in errors made. Playback of entry without earphones

Appendix C: Pretest and Posttest Raw Results

As described earlier, because of technical problems with the system the data collection performed by TALL was often very inaccurate. Notwithstanding, I have decided to include the raw results of the pretest and posttest in Tables 4 and 5, respectively.

Table 4

Pretest results by student

Sutdent ID	Number Times Viewed	Number Times Correct	Pretest Percentage Correct
541	36	1	2.78%
543	38	32	84.21%
545	103	79	76.70%
546	92	72	78.26%
552	46	13	28.26%
556	35	22	62.86%
561	76	37	48.68%
566	70	52	74.29%
567	59	42	71.19%

Note. Only nine students completed the pretest.

Table 5

Posttest results by student

Sutdent ID	Posttest Times Viewed	Posttest Times Correct	Posttest % Correct
542	48	39	81.25%
543	48	41	85.42%
544	49	43	87.76%
545	46	42	91.30%
546	96	69	71.88%
548	37	30	81.08%
549	48	37	77.08%
551	60	46	76.67%
552	50	1	2.00%
553	48	40	83.33%
554	48	24	50.00%
561	48	39	81.25%
565	48	44	91.67%
566	48	13	27.08%
567	48	40	83.33%
568	48	39	81.25%

Note. Only sixteen students completed the posttest.

Appendix D: Questions for Focus Group and Professor Interview

Prior to holding the meeting with the focus group and my interview with Dr. Jones I prepared a series of specific questions that I drew from my evaluation questions. While I ensured that all of these questions were answered during the course of my discussions, I also tried to maintain an open format so as to encourage any other suggestions or comments that may also arise.

Questions for focus group

1. What did you dislike about using TALL? What was frustrating?
2. What would you change?
3. What did you like?
4. In what ways did TALL help your learning?
5. In what ways did it seem to be a hindrance?
6. How much of a role did TALL play in your learning relative to the other methods of study you used?
7. Was it too easy? Too hard? Too repetitive? Too overwhelming?
8. Did you enjoy learning on the computer? Why or why not?
9. How did TALL perform? Did it crash a lot? Was it fast enough?

Questions for focus interview with Dr. Jones

1. What went well?
2. What didn’t go well?
3. Was the content produced using TALL instructionally valid? Was it accurate in terms of the notation and theory? Did it make sense from the perspective of an expert in music theory?

4. Was this a valid method of teaching?
5. From your perspective, did TALL help students in achieving mastery of the concepts? Why or why not? In what ways?
6. In your opinion, could TALL be used to supplement other music courses (Music 190)? Why or why not?
7. Would you use this in future projects?
8. If we were to do this again, what would you change?

Appendix E: Description of Activities

What follows are screen shots and descriptions of each of the activities. I present how each activity appears when it initially loads, and what it looks like after the learner has made an attempt at it.

The content for our instruction was entered into the system using distinct groupings (called concept types in TALL) as individual concepts. For example, intervals were given the concept type “Musical Interval”, and would include concepts such as “Minor Second”, “Perfect Fifth” or “Augmented Tenth”. The student would then be required to complete at least one activity corresponding to each of these concepts. Also, within an activity certain elements were generally randomized. For example, in the activity requiring a student to spell an enharmonically equivalent interval, the various concepts would be cycled through the activity, while the clef and position of the interval on the staff were simply randomized.

Identify Key Signature

The student is shown a key signature in the treble clef and asked to select from drop-down lists the corresponding major and relative minor keys (see Figure 14). If the student answers incorrectly he or she will be given the correct answer (see Figure 15). Only one attempt at this activity is allowed.

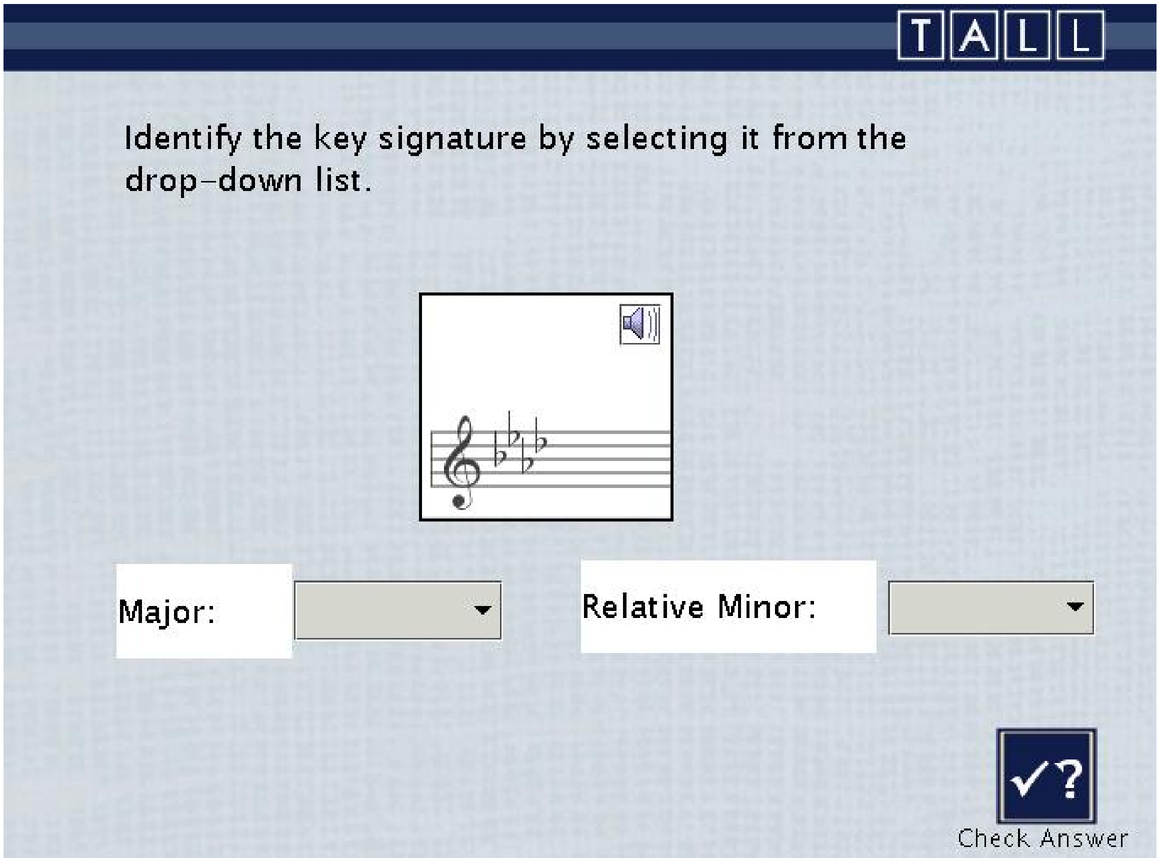


Figure 14. The "Identify Key Signature" activity as it would initially appear to a learner.

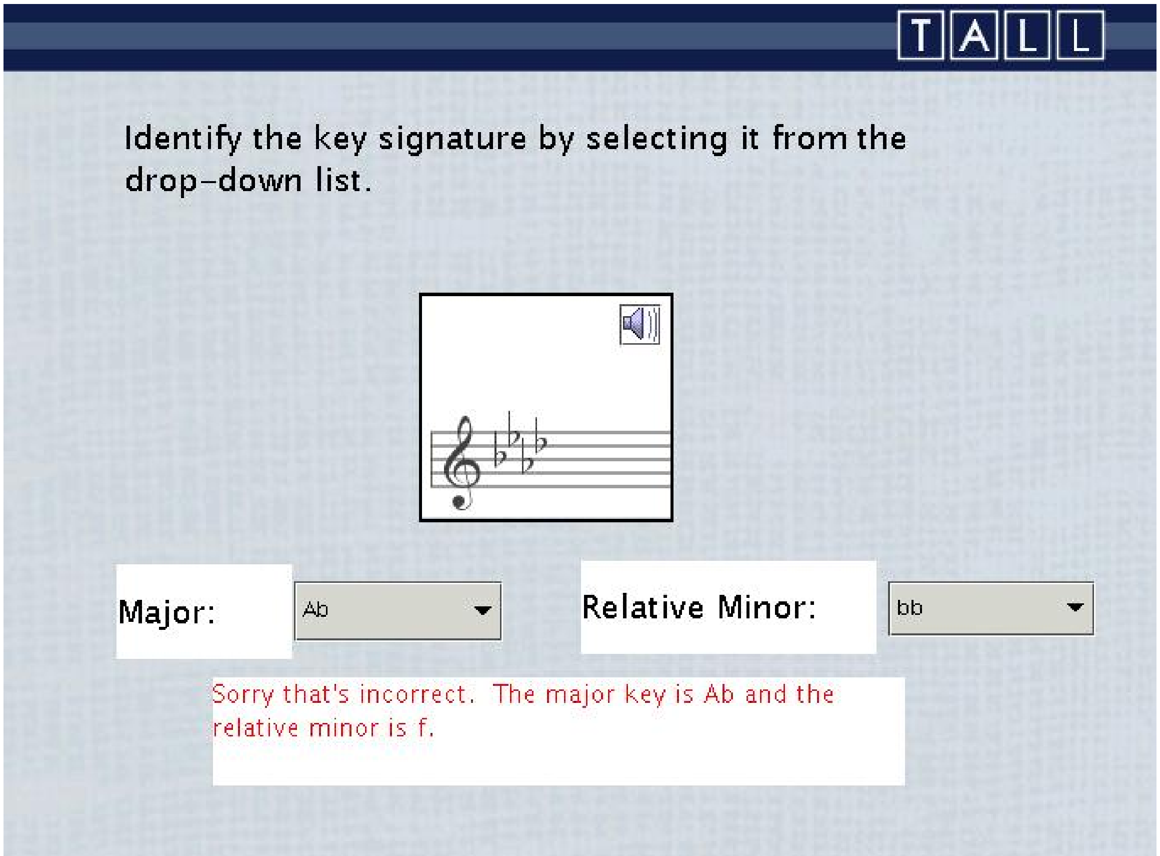


Figure 15. The "Identify Key Signature" activity as it would appear after the learner had answered it.

Recognize Interval

The student is shown an interval on the staff and asked to identify it by selecting it from the drop-down list (see Figure 16). Both the clef and the position of the interval are random. If the student answers incorrectly he or she will be given the correct answer (see Figure 17). Only one attempt at this activity is allowed.

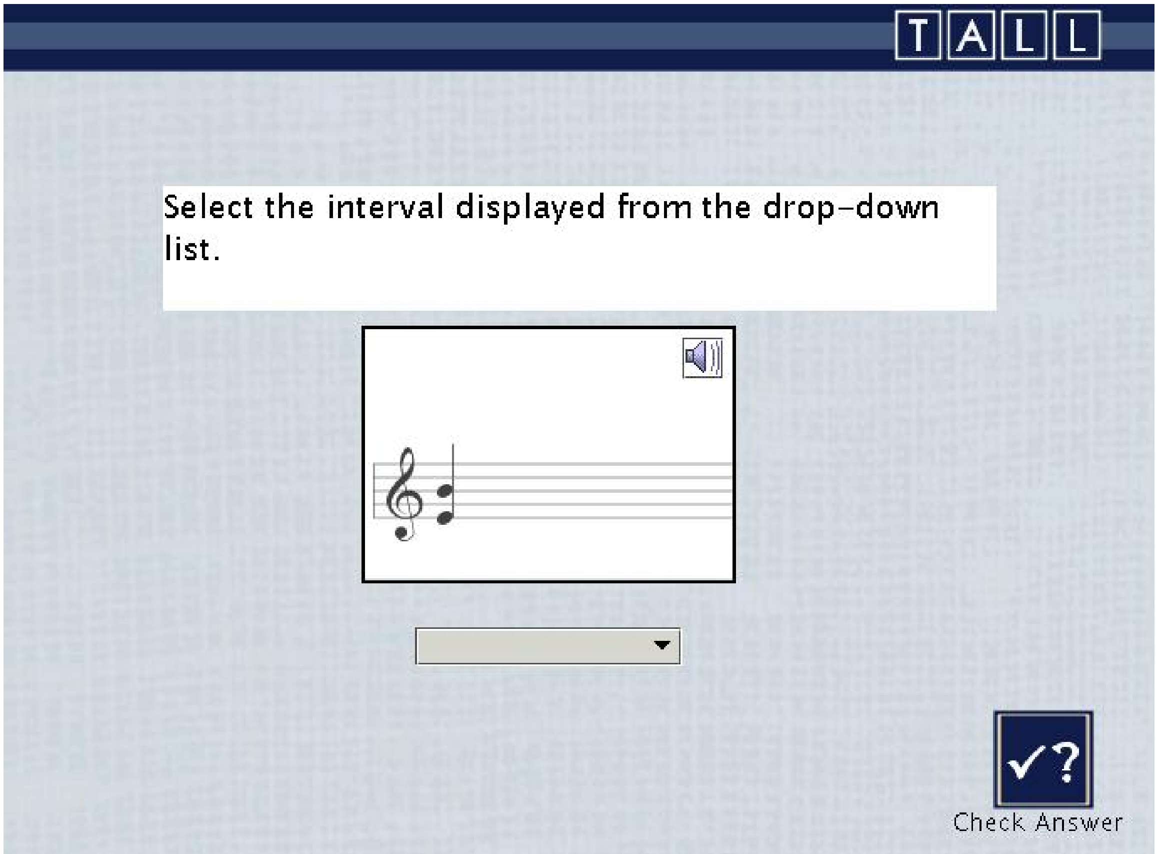


Figure 16. The “Recognize Interval” activity as it would initially appear to a learner.

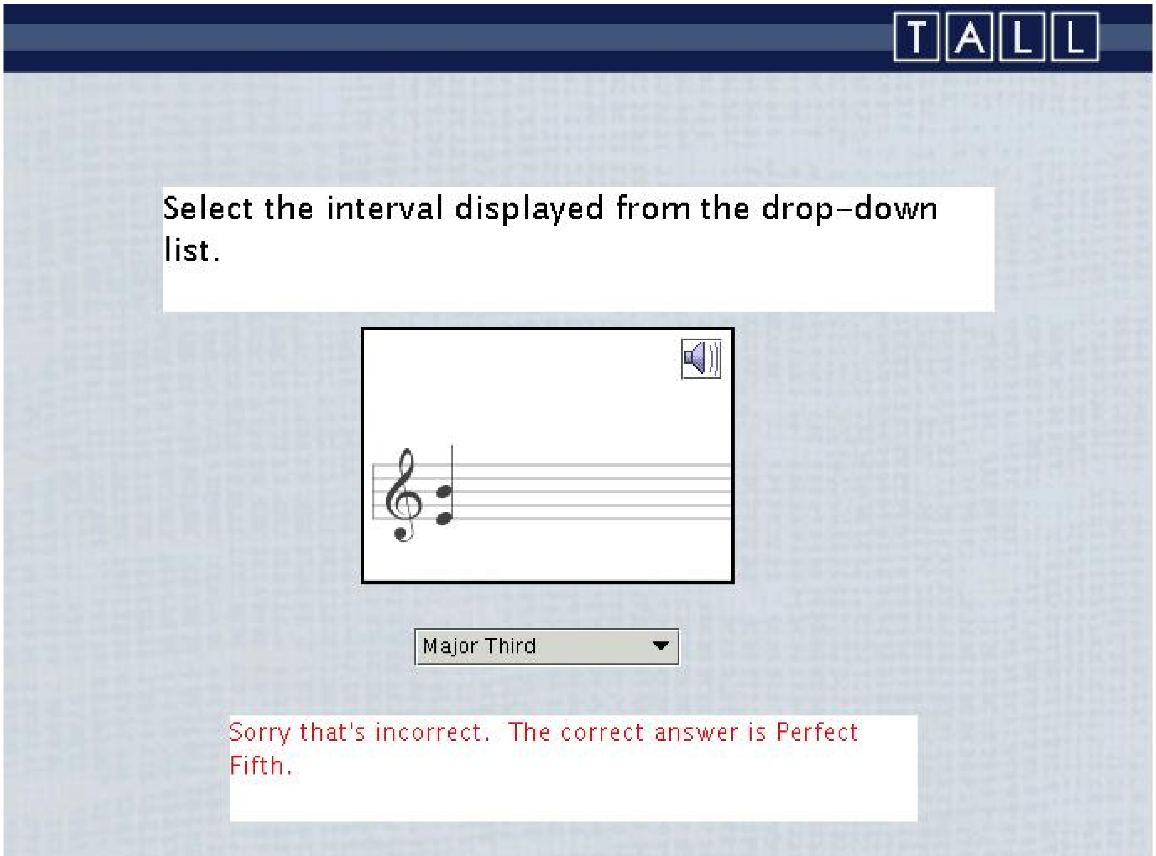


Figure 17. The “Recognize Interval” activity as it would appear after the learner had answered it.

Spell Interval

The learner is presented with a note on the staff, and asked to construct the given harmonic interval by positioning a note above the given note (Figure 18). The clef and the position of the beginning note are randomized. If the student answers incorrectly the musical component will indicate the interval that was actually spelled, but does not display the correct answer (Figure 19). The learner is only given one attempt at this activity.

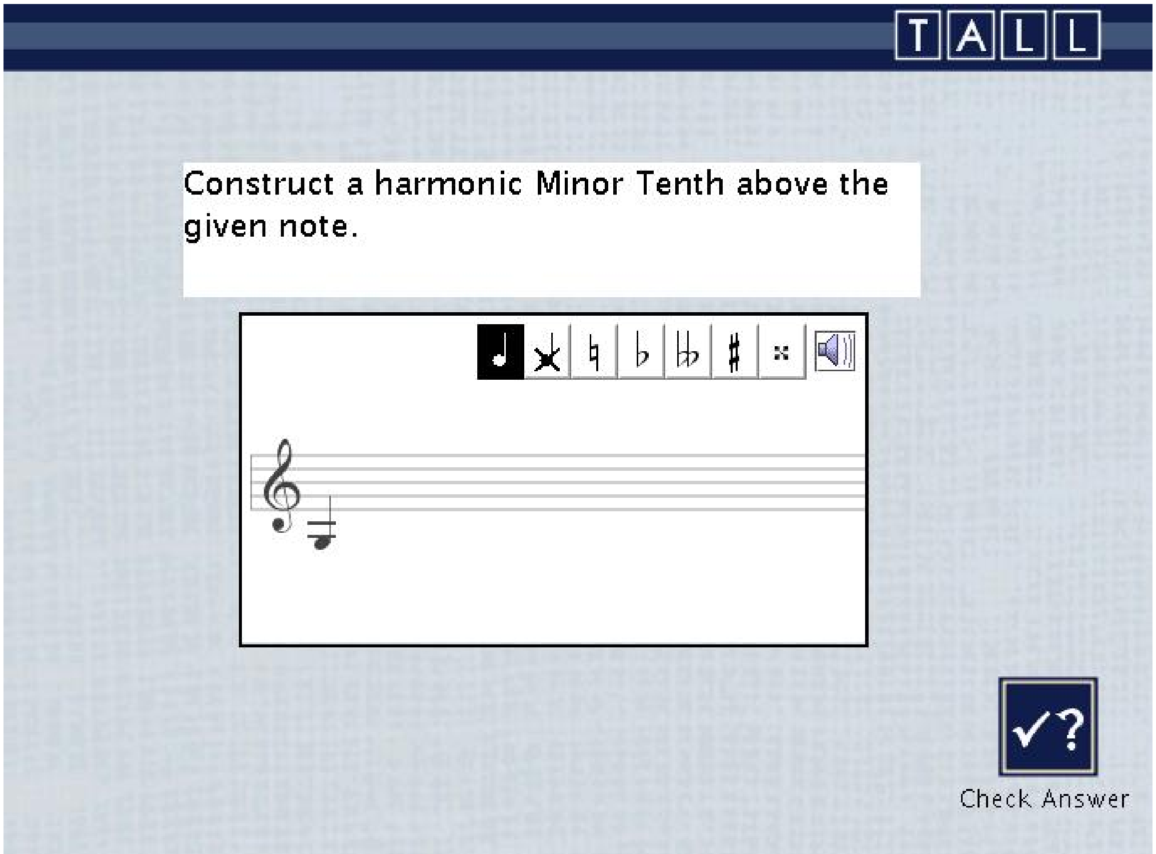


Figure 18. The “Spell Interval” activity as it would initially appear to a learner.

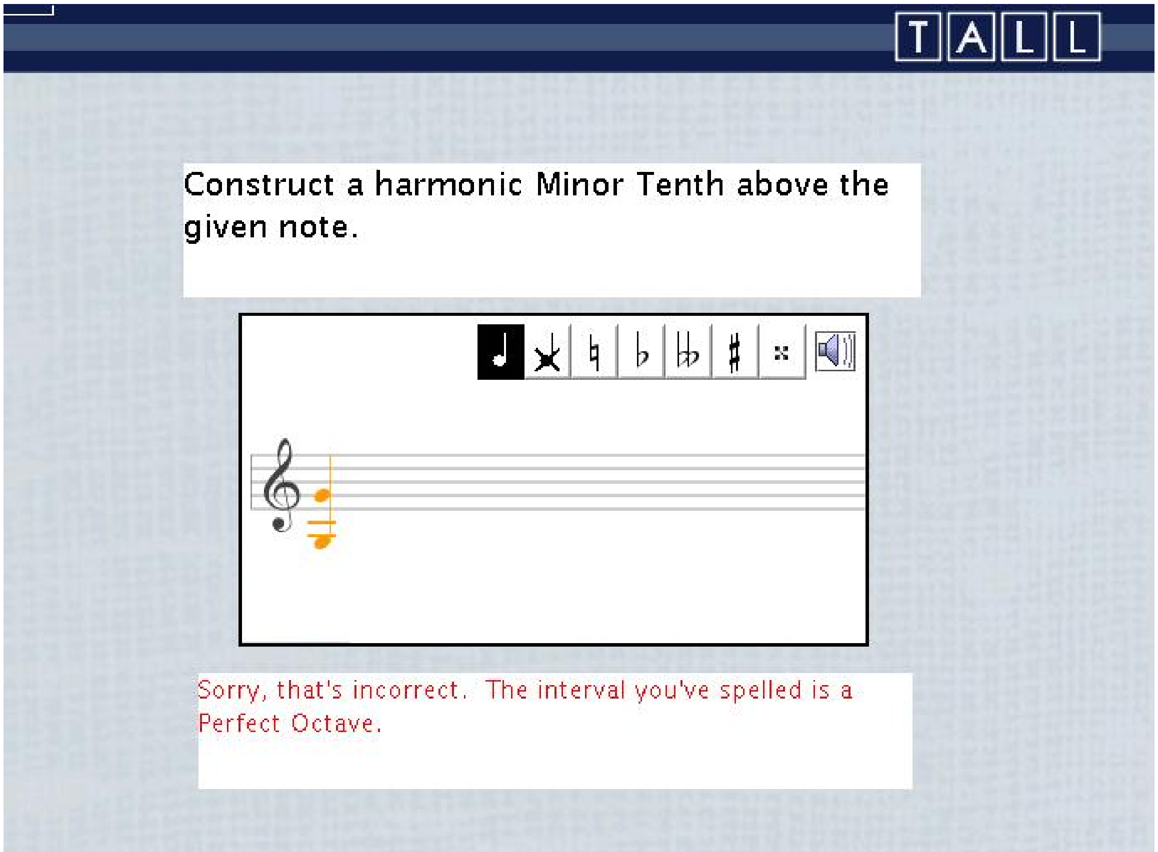


Figure 19. The "Spell Interval" activity as it would appear after the learner had answered it.

Spell Enharmonically Equivalent Interval

Here the student is given an interval and asked to construct an interval that is enharmonically equivalent to it (see Figure 20). The enharmonically equivalent interval can be done a number of ways, requiring that the system have the ability to accept multiple correct answers. The clef and position of the interval are randomized. The student is given one opportunity to answer the activity, after which a "correct" or "incorrect" label is displayed (see Figure 21).

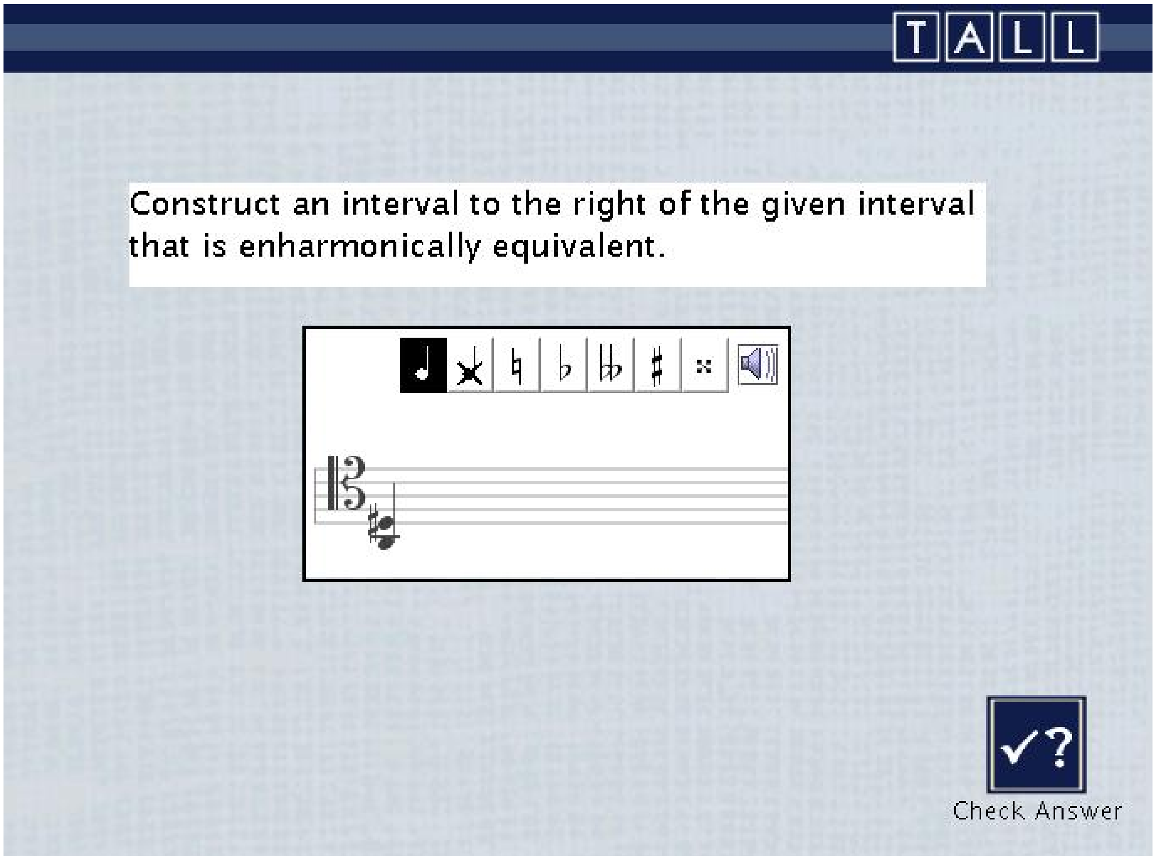


Figure 20. The “Spell Enharmonically Equivalent Interval” activity as it would initially appear to a learner.

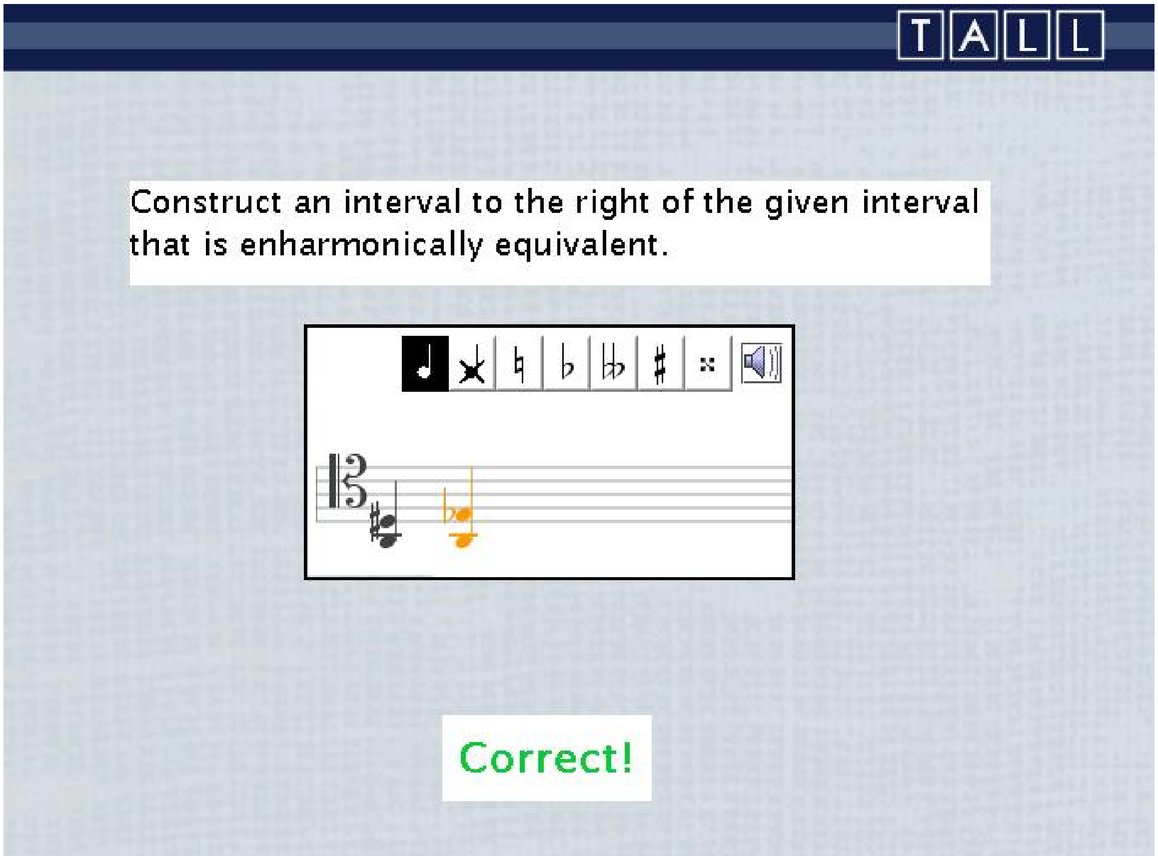


Figure 21. The “Spell Enharmonically Equivalent Interval” activity as it would appear after the learner had answered it.

Spell Interval Inversion

In this activity the student is given an interval and required to generate the inversion of that interval (see Figure 22). The clef and position of the initial interval are randomized. The student is given one opportunity to answer the activity, after which a “correct” or “incorrect” label is displayed (see Figure 23).

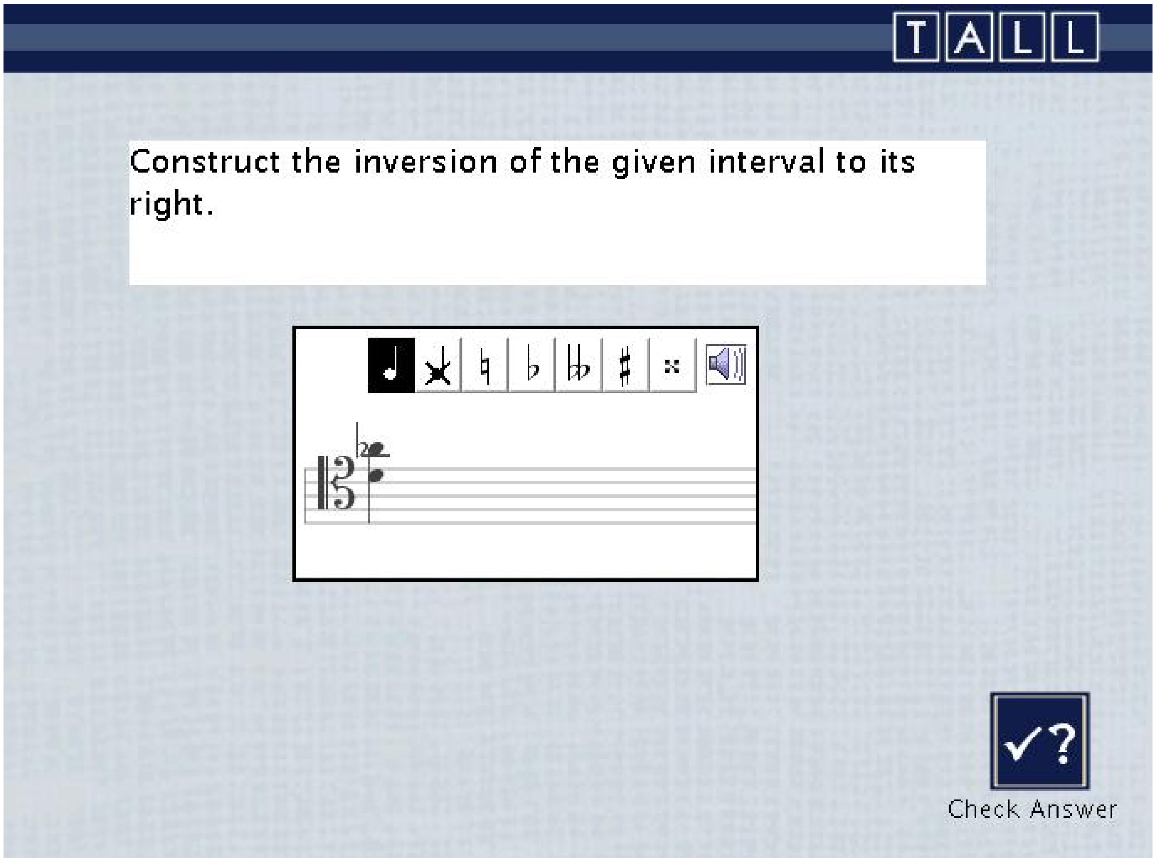


Figure 22. The “Spell Interval Inversion” activity as it would initially appear to a learner.

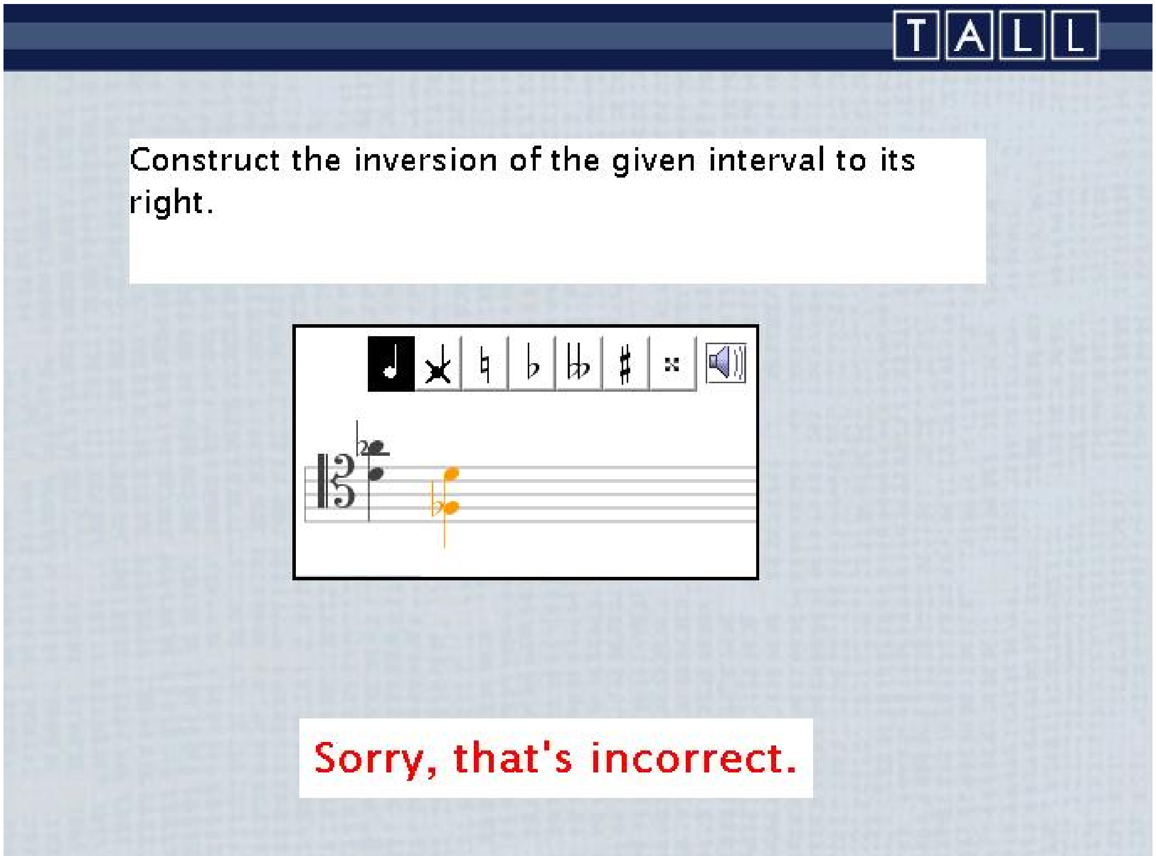


Figure 23. The “Spell Interval Inversion” activity as it would appear after the learner had answered it.

Identify Scale

Given an eight note scale the student is required to identify its type by selecting it from a drop-down list (see Figure 24). If the student answers incorrectly the correct type of scale is displayed (see Figure 25). The clef and starting pitch of the scale are randomized. The student is only given one opportunity to answer this activity.

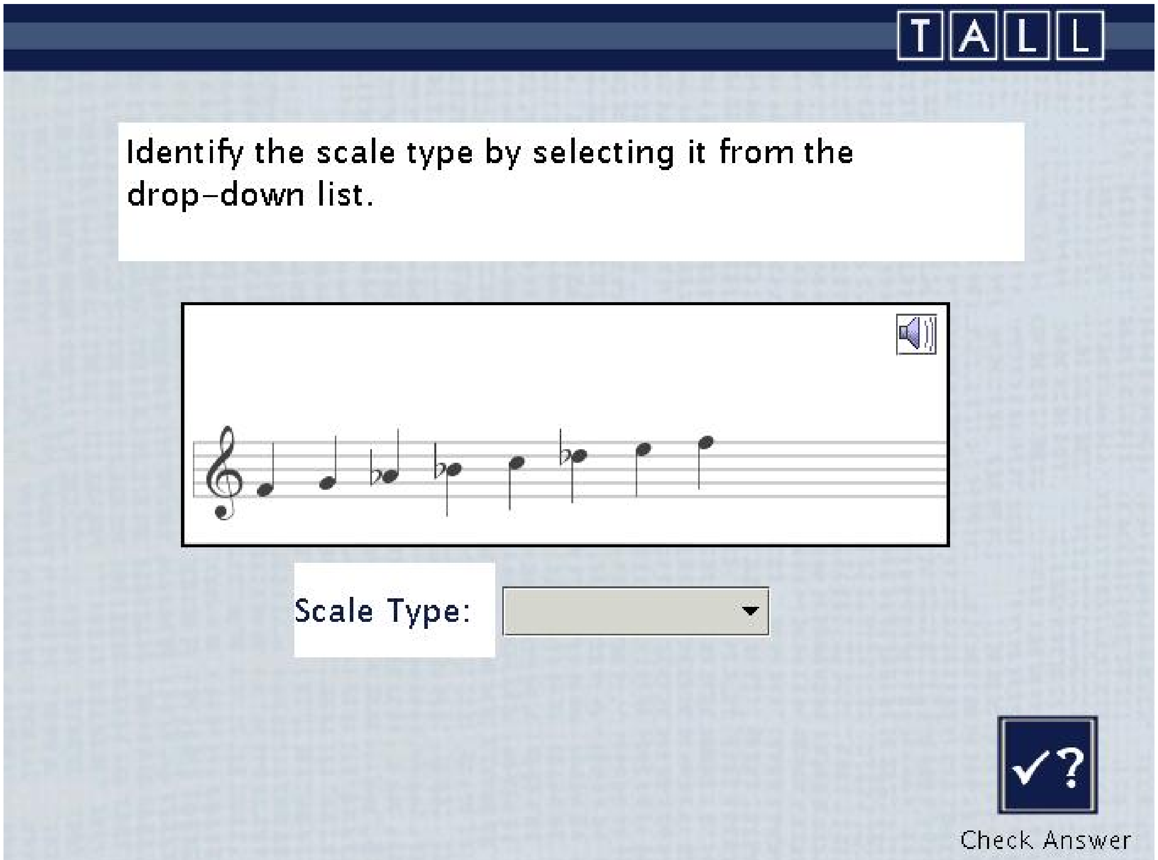


Figure 24. The “Identify Scale” activity as it would initially appear to a learner.

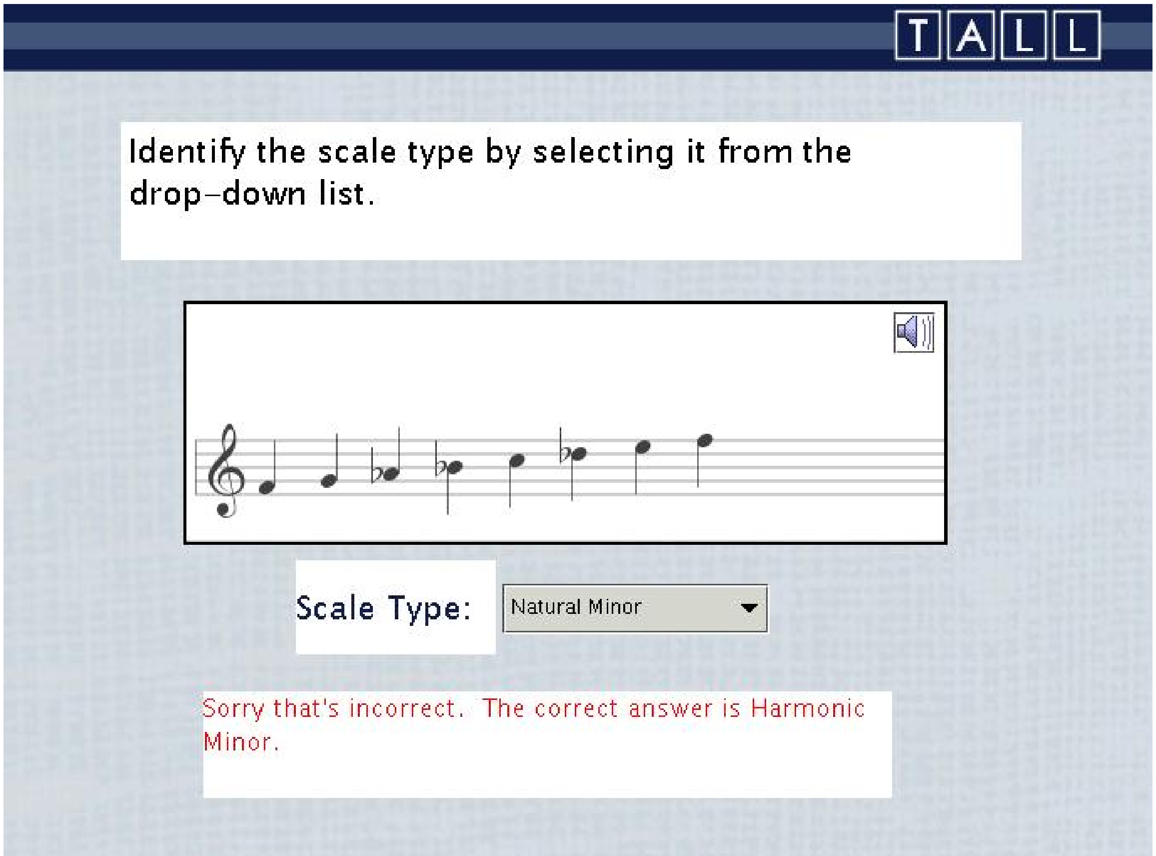


Figure 25. The "Identify Scale" activity as it would appear after the learner had answered it.

Spell Scale

Here the student is provided eight notes in sequence and asked to add accidentals to the scale such that it results in the requested scale type (see Figure 26). The clef and starting pitch of the scale are randomized. If the student spells the scale incorrectly he or she is shown the correct answer (see Figure 27). The student is only given one opportunity to answer the activity.

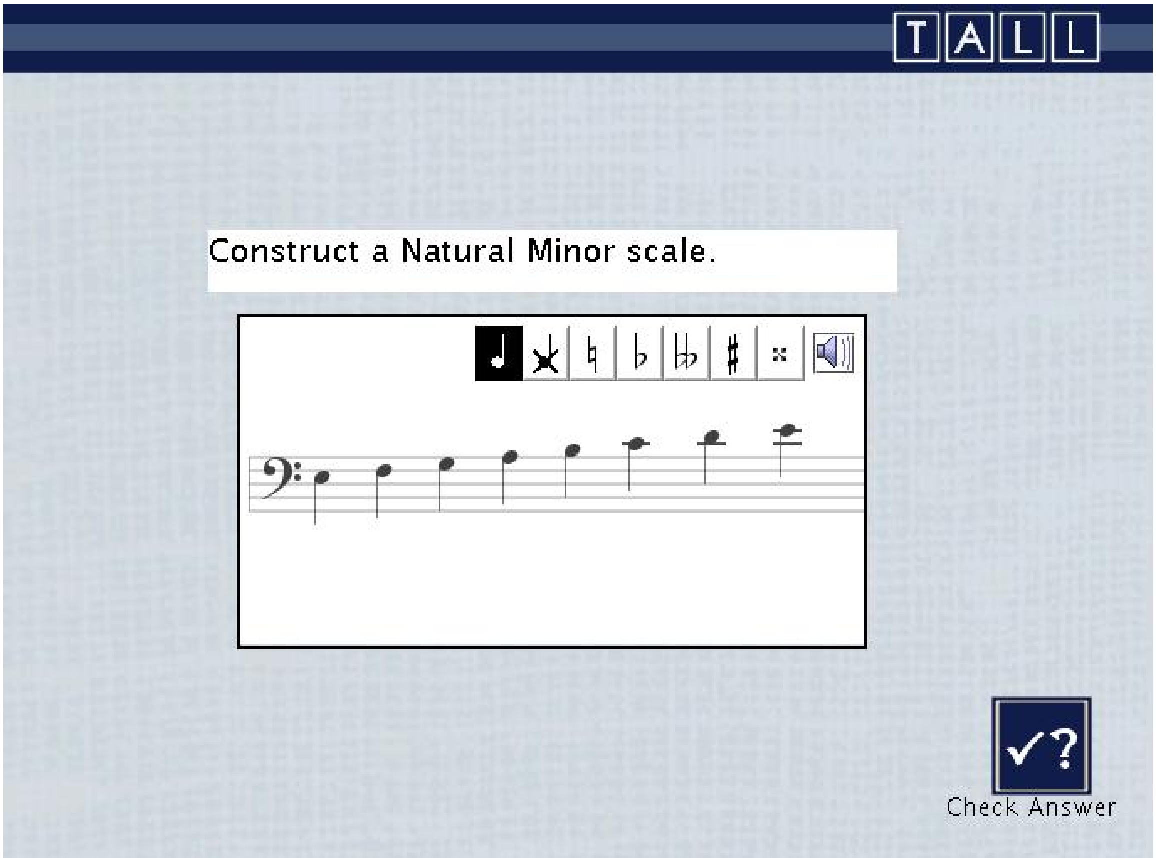


Figure 26. The “Spell Scale” activity as it would initially appear to a learner.

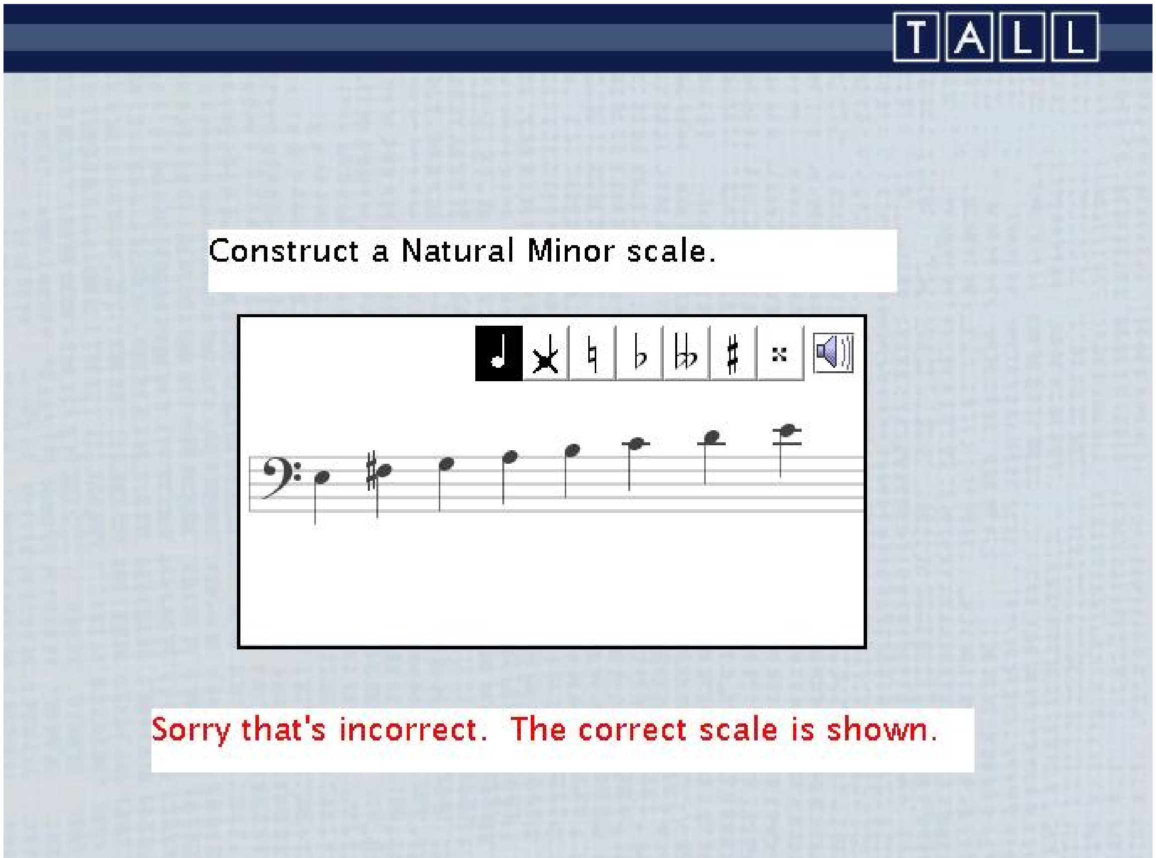


Figure 27. The “Spell Scale” activity as it would appear after the learner had answered it.

Identify Triad

The student is shown a triad and asked to identify its type by selecting it from a drop-down list (see Figure 28). The clef and position of the triad on the staff are randomized. The student is only given one opportunity to answer the activity, and, if answered incorrectly, is shown the correct type (see Figure 29).

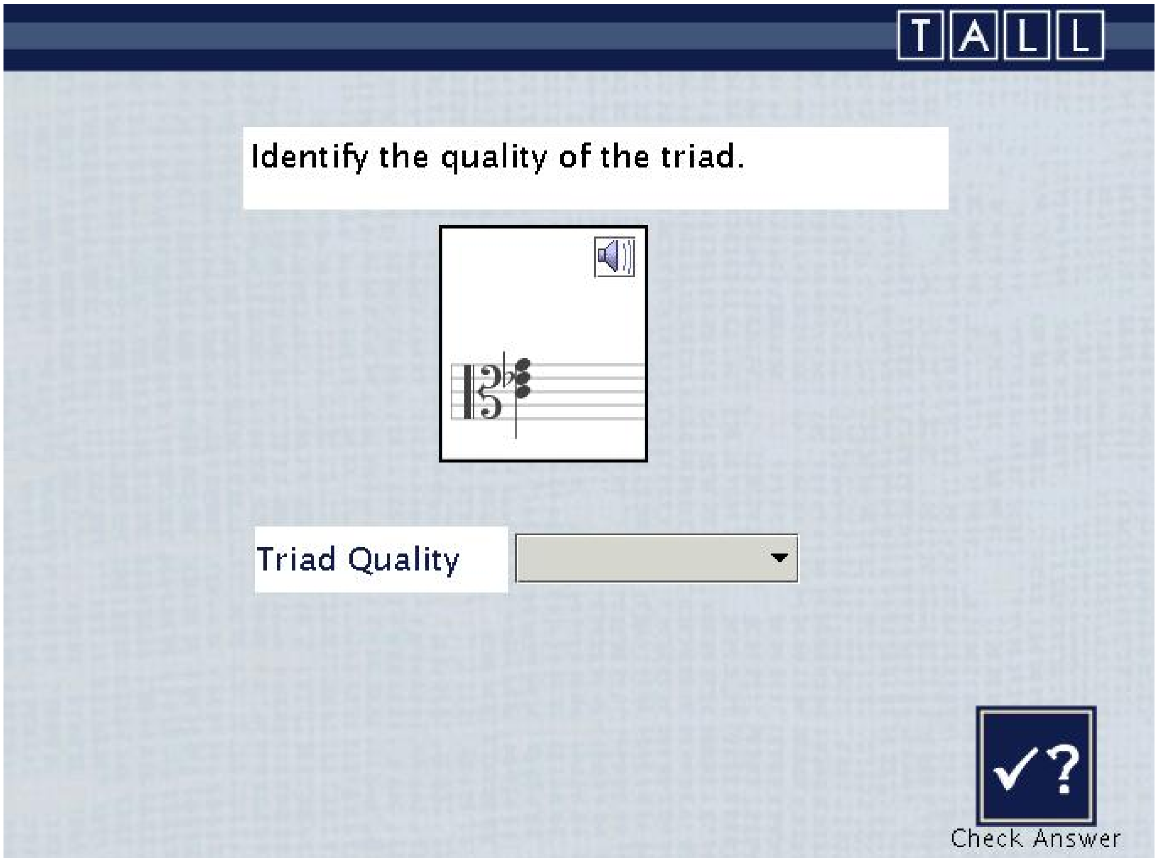


Figure 28. The “Identify Triad” activity as it would initially appear to a learner.

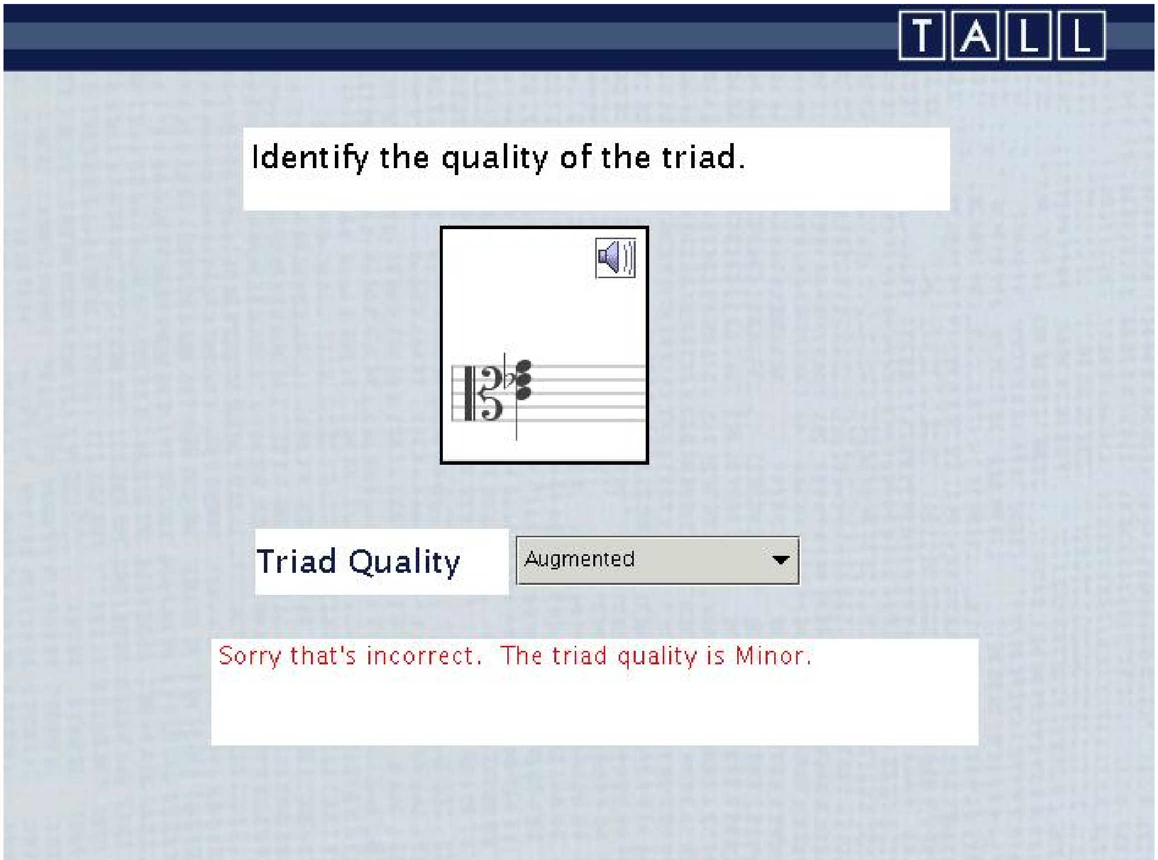


Figure 29. The "Identify Triad" activity as it would appear after the learner had answered it.

Spell Triad

Given a single note on a grand staff, the student is required to construct a triad in four-part harmony using the given type of spacing, which includes choral, keyboard, or open (see Figure 30). This is one of the more complex activities, which requires that the system check that certain rules are followed in the student's answer. Following is a list of these checks:

1. There are not enough notes on the staves.
2. There must be three or four notes for the triad to be in correct spacing.
3. One of the notes on the upper staff is incorrectly spelled.

4. One of the notes on the lower staff is incorrectly spelled.
5. The root of the triad is not doubled.
6. The number of notes on each of the staves is incorrect, according to the spacing requirements.
7. The soprano voice doesn't correspond to the tonic of the triad.
8. More than an octave separates two of the three upper voices.
9. The notes in the treble staff must be as close together as possible.
10. The three upper voices must be as close together as possible.
11. The soprano voice doesn't correspond to the tonic of the triad.
12. More than an octave separates two of the three upper notes.
13. The alto and soprano voices are in close spacing. Please rearrange them so that they are in open spacing.

The beginning pitch and spacing requirement are randomized. The student is given as many attempts at this activity as needed, but must be able to answer the activity correctly in an average of no more than 1.4 attempts in order to avoid continuous review of the activity. Each time the student clicks the “check answer” button he or she will be shown a remedial feedback statement if the answer is incorrect (see Figure 31).

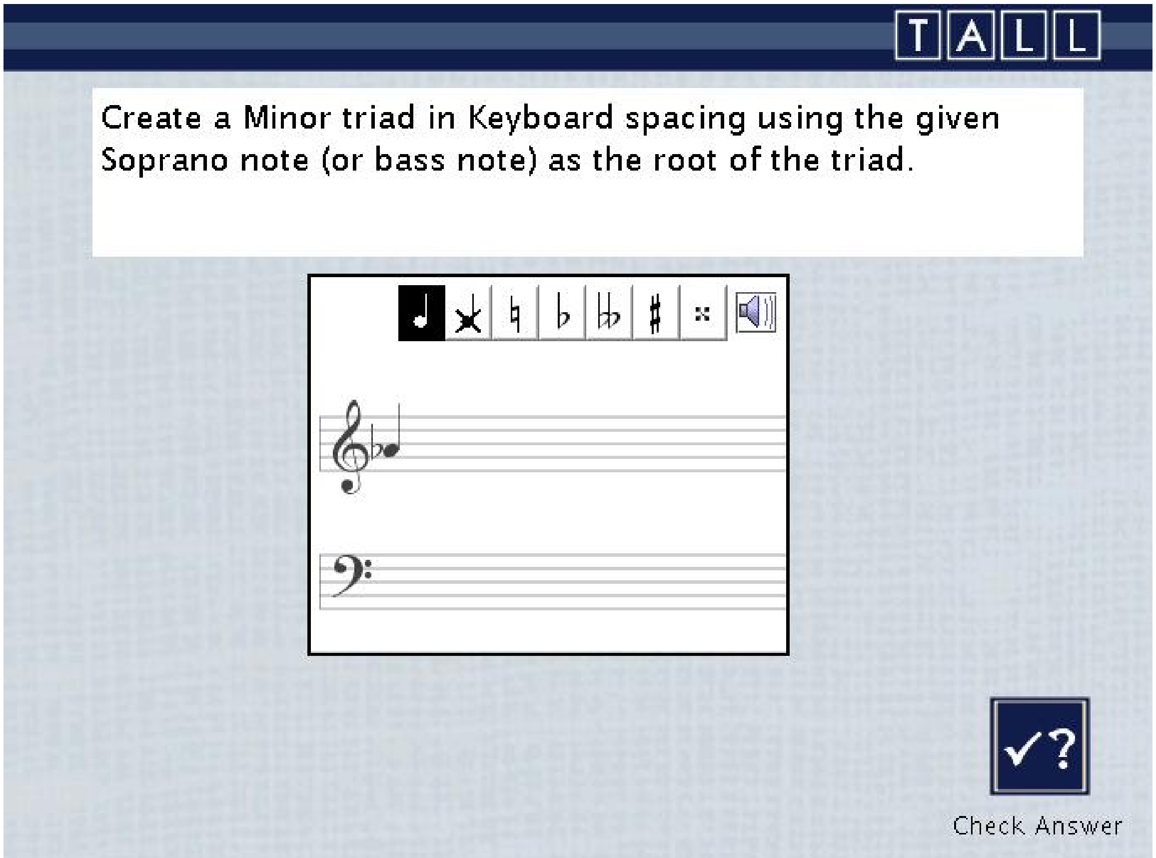


Figure 30. The “Spell Interval” activity as it would initially appear to a learner.

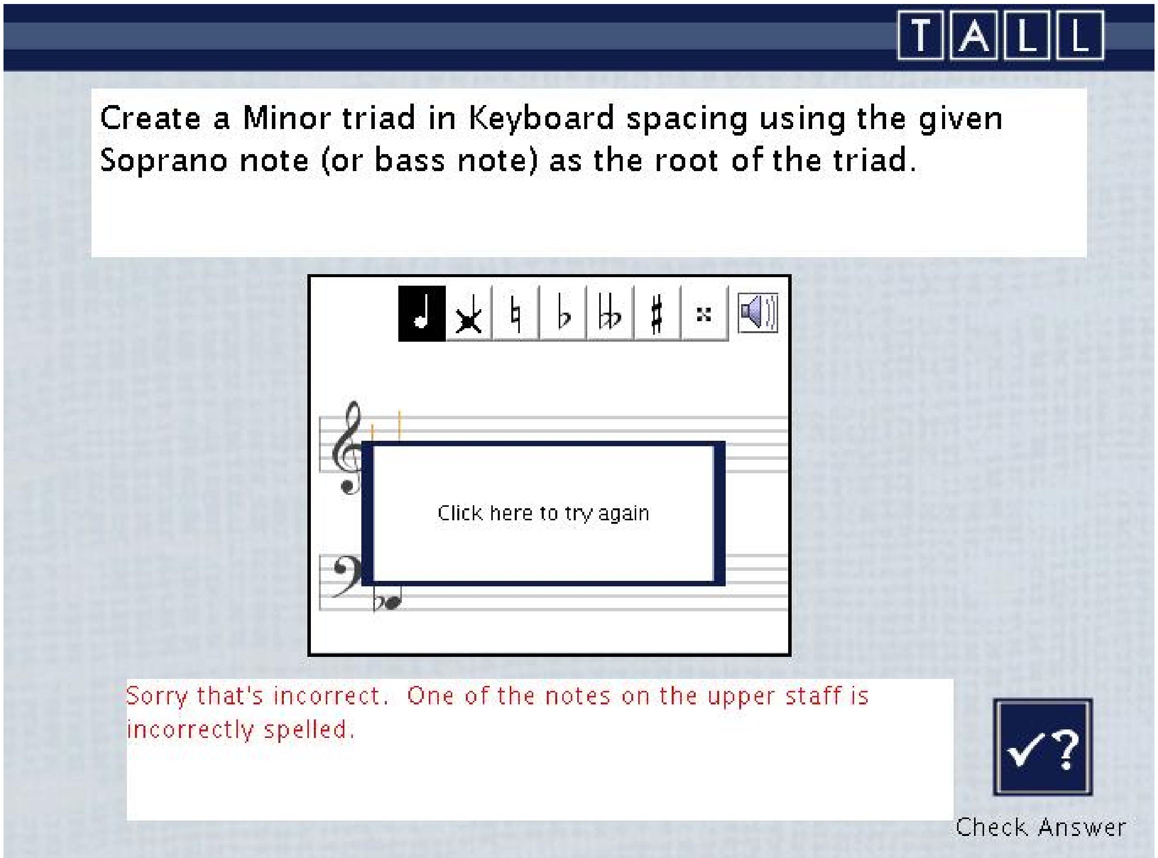


Figure 31. The “Spell Interval” activity as it would appear after the learner had answered it.

Dominant Seventh Resolution

Here the student is given a triad presented in four-part harmony on a grand staff using either a major or minor key. The student is required to correctly resolve the dominant seventh chord to the tonic chord of the key (see Figure 32). Again, there are a series of checks the system makes to ensure that the student has correctly resolved the chord:

1. There must be three or four notes immediately following the dominant seventh.
2. The root of the dominant seventh must resolve to the root of the tonic triad.

3. The seventh note of the dominant seventh chord doesn't resolve to the third note of the tonic.
4. The seventh note of the dominant seventh chord doesn't resolve down.
5. The seventh note of the dominant seventh chord doesn't resolve down.
6. No more than an octave can separate any of the upper three voices.
7. The third note of the dominant seventh chord doesn't resolve to the root or fifth note of the tonic. Remember that if the third note of the dominant seventh is in the outer voice it must resolve to the root of the tonic; however, if it is in the inner voice it can resolve either to the root or fifth of the tonic.
8. The root note of the dominant seventh chord doesn't resolve to the root or fifth note of the tonic.
9. The fifth note of the dominant seventh chord doesn't resolve to the root note of the tonic.
10. Parallel fifths are not allowed.

The key signature and major or minor status of the key are randomized. As with the spell triad activity, the student is given as many attempts at this activity as needed, but must be able to answer the activity correctly in an average of no more than 1.4 attempts in order to avoid continuous review of the activity. Each time the student clicks the "check answer" button he or she will be shown a remedial feedback statement if the answer is incorrect (see Figure 33).

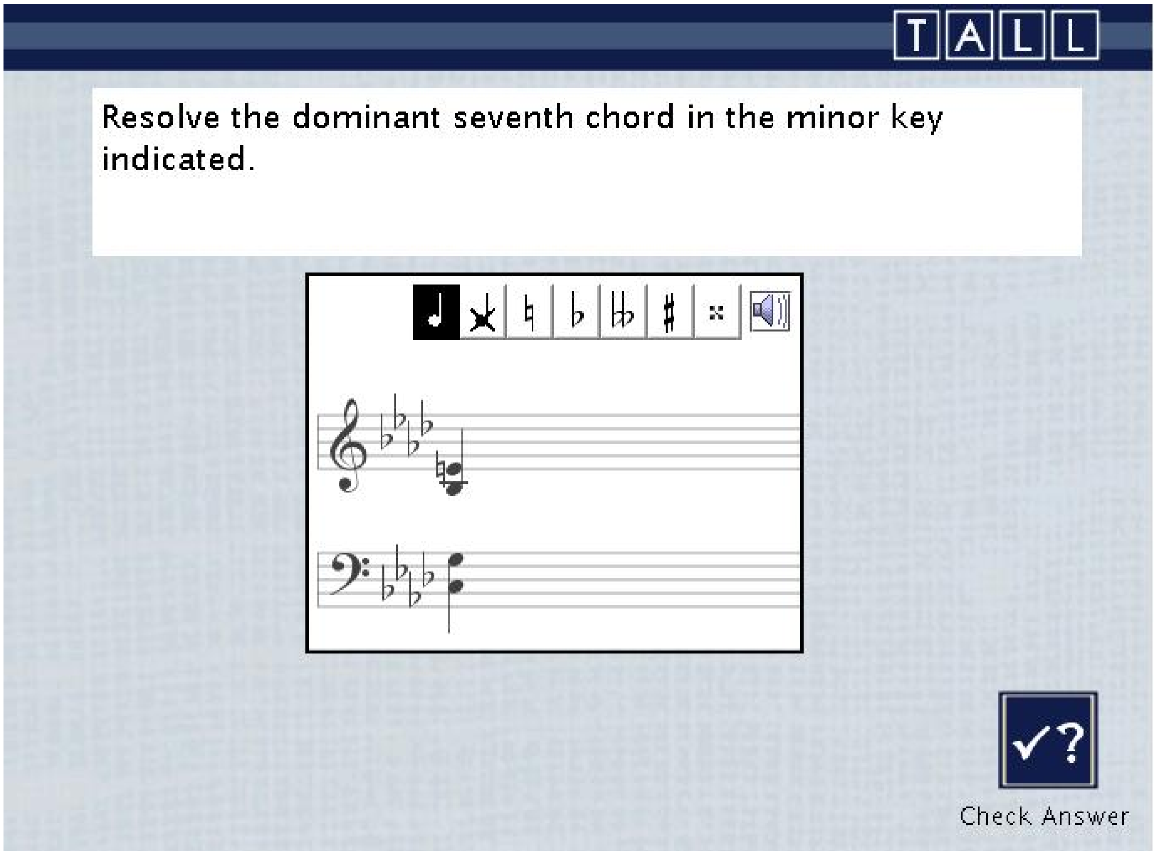


Figure 32. The “Dominant Seventh Resolution” activity as it would initially appear to a learner.

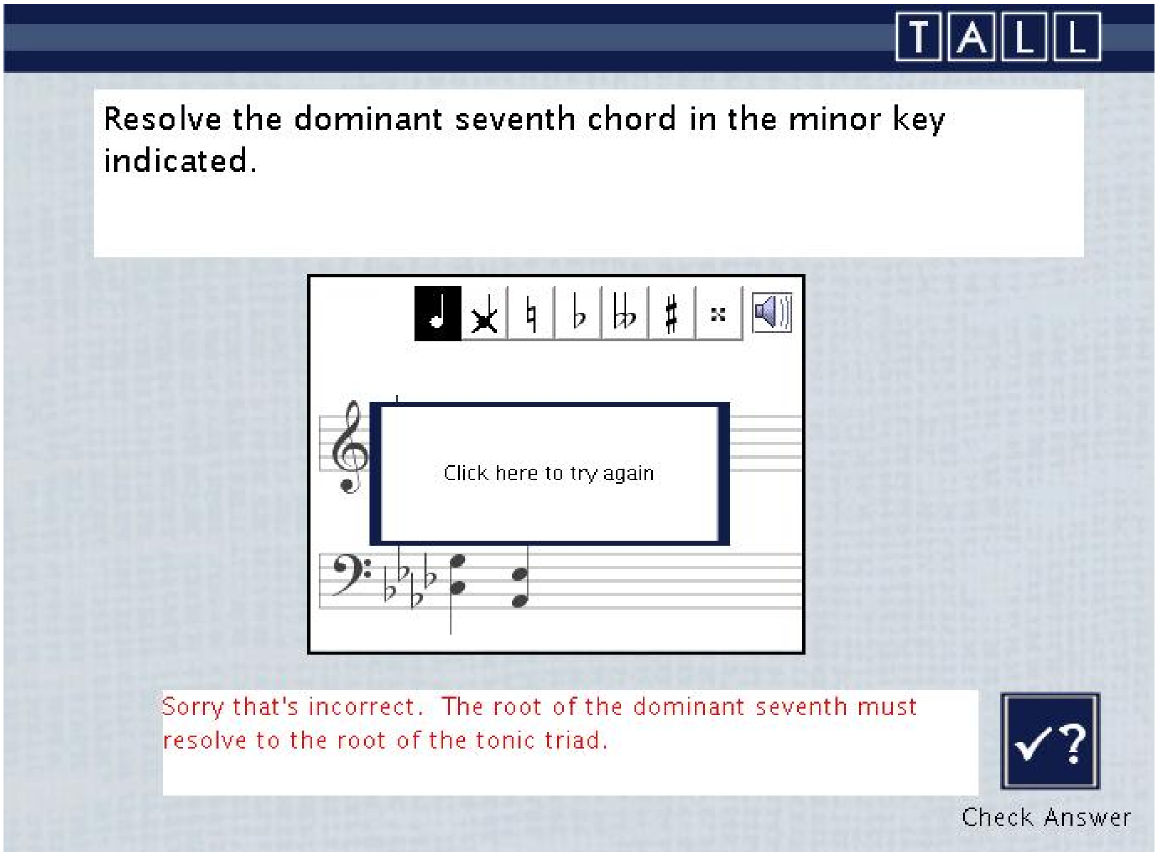


Figure 33. The “Dominant Seventh Resolution” activity as it would appear after the learner had answered it.