Criterion–Referenced Grading vs. Flipped Classroom Format vs. Visual Voice-Leading Cues: Which Undergraduate Music Theory Teaching Technique Actually Worked?

Robert T. Kelley, Lander University

Abstract

Undergraduate music theory coursework is an obstacle to graduation for many music majors. Music theorists have tried a variety of techniques to improve student success, but the relative benefits of different teaching strategies in the music theory classroom have not been rigorously studied. This study evaluates the effectiveness of different pedagogical interventions in sophomore music theory courses. After years of lecture-format teaching with a weighted-average grading system, I attempted to reach failing students by implementing several teaching strategies over several years, including giving separate grades for the course's core competencies (criterion-referenced), teaching students to add tendency arrows to their part writing (visual voice leading cues), and changing from lecture to in-class problem solving (flipped) with pass/fail grading on each course topic and unlimited reattempt possibilities (specifications). I conducted an analysis of final exam data from these four student groups across a ten-year period to see which of these changes had the greatest impact on student success. The results indicate that the flipped classroom format, combined with specifications grading, was the most effective strategy in improving student performance in harmonic analysis and part writing. This approach significantly outperformed traditional lecture-format teaching, with students in the flipped classroom showing higher success rates in meeting standardized performance benchmarks. The study suggests that active learning with immediate feedback is a powerful tool in fostering deeper engagement and improving outcomes for at-risk students in music theory. These findings advocate for further investigation and wider adoption of active learning and flipped classroom methodologies in music theory pedagogy.

Introduction

Decades of pedagogy research have introduced and validated many new methodologies for the college classroom,[1] but the music theory community has been slow to adopt these new approaches in the core undergraduate music curriculum.[2] College music majors who struggle with their traditional theory courses would benefit from teaching techniques that have been shown to reduce achievement gaps, such as culturally relevant pedagogy,[3] sense of belonging initiatives,[4] active learning strategies,[5] and enhanced advising systems.[6] In the context of the undergraduate music theory curriculum, Duker et al. (2015), Johnson (2015), Shafer (2018), and others have explored the use of alternative grading and active

---

[1] Many valuable teaching innovations that have been rigorously studied are summarized in Lang 2016.
[2] Snodgrass 2020.
[3] Byrd 2016, Hernandez 2022.
[4] Gopalan and Brady 2019, Murphy et al. 2020.
[5] Theobald et al. 2020.
[6] Erwin et al. 2021.

learning strategies, and Yu Wang (2022) has offered a framework for culturally relevant music theory pedagogy. Little work has been done, however, to evaluate the effectiveness of active learning and alternative grading in the context of the undergraduate music theory curriculum. In this study, I seek to begin this undertaking by measuring the impact of different pedagogical interventions on sophomore theory final exam scores.

Before fully committing to changing my sophomore theory course in 2016, I attempted to reach failing students by rewriting assignments to be more directly instructive, incorporating team-based learning activities,[7] and providing step-by-step methodologies for solving complex problems like interpreting an unusual figured bass symbol. In 2013, I was nevertheless still largely teaching part writing in the way that I had learned it as an undergraduate using lecture-format teaching and a weighted-average grading system. In 2016, I changed my grading system to separate out behavior-based grades, like attendance, homework completion, and participation, from the core competencies of the class, like music analysis, part writing, and remembering terminology and concepts.[8] I also joined the small number of undergraduate theory teachers to try a scale-degree-function approach to part writing and analysis of tonal harmony based on Quinn 2005.[9] In 2019, I began requiring students to use a symbolic system for the scale-degree functions and to write arrows after notes in their part writing and analysis as reminders of their resolution tendencies. An example of this analytical notation can be seen in Figure 1. In 2021, I implemented a flipped-classroom format in sophomore theory.[10] Instead of lecturing, leading students in group harmonic analysis, and demonstrating part writing on the board, I made every class involve students practicing analysis and part writing in the same format as the assessments, with direct immediate feedback from me as I walked around the classroom and assisted each student in turn. The flipped teaching format requires students to work outside of class on the reading assignments and studying for the terms and concepts quizzes. At the same time, I also implemented a specifications grading system based on the course content learning modules like secondary dominants, modal mixture, or modulation, rather than on overall competencies like part writing, analysis, and theory concepts.[11] The specifications grading system requires students to complete a predetermined number of pass/fail assessments to achieve a certain course grade. Students may reattempt failed assessments multiple times.

After some success using inverted pedagogy and specifications grading, I decided to collect and analyze data from final exams across the four groups of past students:

Traditional: Lecture-format class with traditional part-writing pedagogy and weighted-average grading system.

Criterion-Referenced: Lecture with part writing by scale-degree functions and grades separated into four competencies:

- factual knowledge
- part writing
- analysis
- professional conduct

---

[7] Both of these strategies came from a POGIL workshop that I attended (Simonson 2019).
[8] I based this intervention on Duker et al. 2015.
[9] Kelley 2006 and Kelley 2017 outline my work inspired by Quinn 2005 (and Shaffer 2014) to implement scale-degree functions in my classes.
[10] Talbert 2017.
[11] Nilson 2015.

Visual Voice Leading Cues: Lecture with tendency arrows required in part writing and criterion-referenced grading.

Flipped/Specifications: Activity-based classes with visual voice leading cues and specifications grading by course topic.

Table 1 shows the teaching format, grading system, and part-writing methodology for each group.

Table 1: Pedagogical interventions used in each treatment group.

	2013 Traditional	2016 Criterion-Referenced	2019 Visual Voice Leading Cues	2021 Flipped/Specifications
Classroom Format	Lecture Format Classroom			Flipped Classroom
Grading System	Weighted Average Grading	Criterion-Referenced Grading		Specifications Grading
Analysis System	Traditional Harmonic Analysis	Scale Degree Functions	Visual Voice Leading Cues	

Methods

To see whether any of the classroom interventions that I have listed was effective, I collected student work samples from second-semester sophomore theory final exams given between 2014 and 2024. To make sure that the student activities were comparable, I extracted the part writing and analysis section of students' final exams in the first three groups, and the last part writing and analysis assessment that I had in my files for students in the flipped/specifications group.

Since I did not have part writing and analysis samples from all students who failed or dropped out, I only used work from students who completed the university's four-semester theory sequence. While this makes it impossible to consider the difference between students who would have passed under one system and failed under another, it more accurately represents the impact of my teaching on the abilities of successful graduates.

To obtain consistent performance data from the student part writing exercises, I decided to rely upon the well-established norms and grading procedure of the AP Music Theory Exam, which evaluates part writing based upon Roman numeral analysis, chord spelling, and voice leading, and requires a 60% score to pass. Figure 1 shows one of the anonymized part writing and analysis samples, with the AP-exam-style scoring added in purple. To eliminate any possibility of bias in scoring, I gave the 41 collected work samples to Nancy Rogers at Florida State University, who has served as an AP Music Theory Exam reader for many years. The priorities of the AP exam's grading system are not precisely aligned with my own, but they nevertheless effectively differentiate performance levels among students in their harmonic analysis, chord spelling, and voice leading skills.

With achievement metrics for 41 students in the four treatment groups, I compared the mean scores in the four groups using an ANOVA with a Tukey-Kramer post-hoc test. I then regrouped the data into two treatments based solely upon the classroom format: lecture (2013–2021) and flipped (2021–2024). I performed a Fisher's exact test to determine the significance of differences between the proportion of students in each treatment group meeting and not meeting the AP exam's 60% benchmark score. To account for the exclusion of students who did not pass sophomore theory, I also compared the DFW rate (the percentage of students who received a D or F grade or who withdrew (W) from sophomore theory) between the two groups using a Fisher's exact test.

Figure 1: Student work sample from the visual voice leading cues group, with scoring in purple showing a total of 31/33 and subscores of 9/9 for Roman numerals, 8/8 for chord spelling, and 14/16 for voice leading. Two points were deducted from voice leading for parallel fifths and octaves between the fourth and fifth chords.

Results

Students in the flipped/specifications group had significantly higher mean scores than those in the traditional and criterion-referenced groups (p < 0.05; Figure 2). The visual voice leading cues treatment had a mildly positive effect over the previous two treatments, though not reaching statistical significance, likely because of the small sample sizes in this study. Adding visual voice leading cues nevertheless seems to have increased the performance at the top of the class but not the bottom, while the active learning class format helped all students and was the only treatment to substantially improve scores at the bottom of the class.

Students who passed sophomore music theory with a flipped classroom had a significantly higher likelihood of achieving the 60% AP exam passing score compared to those who had a lecture-based class (p < 0.01; Figure 3). Specifically, 87.5% (14/16) of students passed the AP exam part writing and analysis standard at the end of hands-on activity–based sophomore music theory courses with specifications grading, while only 44% (11/25) passed the standard after the same course in any of the lecture-format treatment groups.

DFW rates in sophomore music theory were slightly higher in the flipped/specifications group than in the previous treatments (Figure 4). These rates, however, vary dramatically from year to year, and the difference between the groups is not statistically significant (p = 0.19).

Discussion

Of the classroom interventions tested, only the flipped classroom with specifications grading significantly improved student exam scores. Students who passed sophomore music theory in a flipped classroom were nearly twice as likely to meet the benchmark score of 60% than in a lecture-format classroom. These results show that a flipped classroom leads to a dramatic improvement in student mastery of part writing and harmonic analysis skills.

Only students who passed sophomore theory were included in this study. Specifications grading could make passing the class more difficult, resulting in fewer students passing the course. Flipping the classroom requires that students review factual content outside of class, which may have the effect of retaining only students who are willing or able to memorize
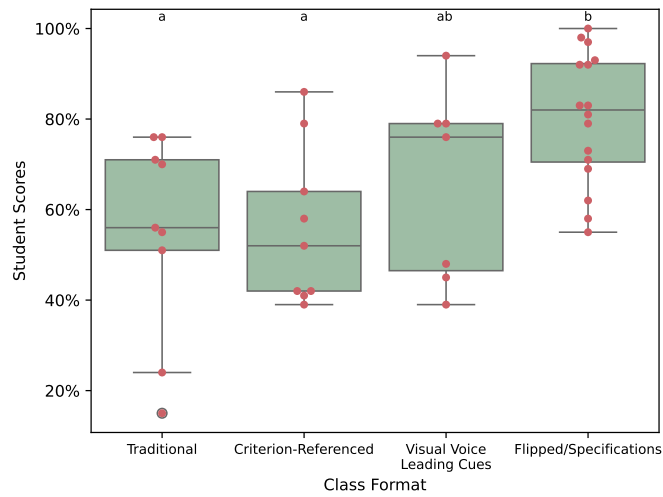
Figure 2: Sophomore music theory completers' work sample scores grouped by teaching technique. Different letters indicate significant differences between groups (p < 0.05). The flipped/specifications group performed significantly better than traditional and criterion–referenced.
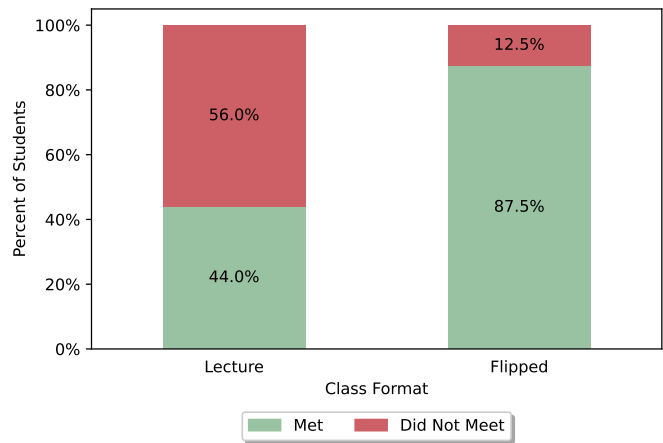

Figure 3: Percent of students meeting 60% benchmark score in lecture and flipped class formats. The flipped format nearly doubled the number of students meeting the AP Music Theory Exam's pass mark (p < 0.01).
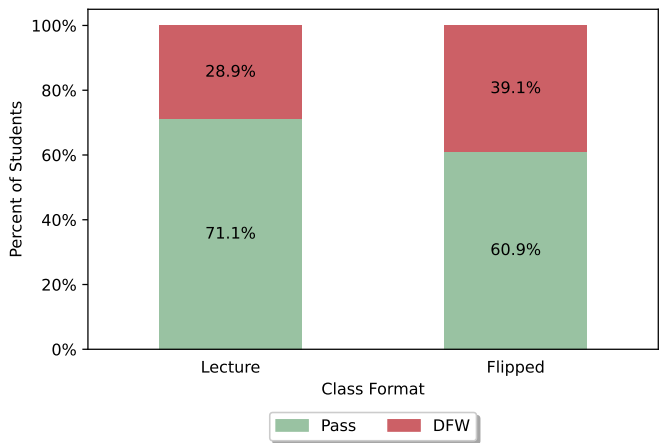

Figure 4: Percent of students receiving a passing final grade in sophomore theory classes in lecture and flipped class formats. The small difference in pass rate between lecture and flipped formats was not significant (p = 0.19).

material before exams. One possible reason that the performance of successful students improved in this study might have been that the weakest students failed or dropped out. There was no difference, however, in DFW rates between the traditional and flipped/specifications course formats, indicating that the improvement in exam scores is a result of increased student learning rather than decreased retention of weak students.

The AP exam's achievement bar of 60% is low compared to the expectations of many college music theory courses. The positive effect of flipping the classroom is still present, however, at higher benchmarks. Increasing the passing score to 70% changes the result to 36% (9/25) of lecture students passing and 75% (12/16) of flipped students, and an 80% pass mark changes the result to 8% (2/25) and 56% (9/16), with significant differences in each case ($p = 0.02$ and $p = 0.001$).

It was impossible to fully control each variable in this study. As shown in Table 1, the treatment with significantly higher results is a combination of inverted pedagogy and specifications grading added onto the existing visual voice leading cues treatment, whereas the criterion-referenced grading group did not build upon an existing teaching intervention, and the visual voice-leading cues group only added one treatment onto the existing criterion-referenced grading intervention. The results in Figure 2 nevertheless show minimal differences when introducing a nontraditional grading system and when adding visual voice leading cues on top of the existing treatment, while adding a flipped classroom format did result in distinguishably improved outcomes. This suggests that, while the efficacy of problem-based active learning may rely upon combination with other pedagogical best practices, it still was the most effective teaching intervention in the undergraduate theory classroom.

Conclusions

All teaching innovations evaluated in this study resulted in small improvements in student performance, but these results indicate that flipping the classroom is the most effective for the heavily skills-oriented undergraduate theory curriculum. The change in score distribution between the visual voice leading cues and flipped/specifications groups further suggests that among the beneficial data-driven pedagogical interventions in this study, in-class problem-based learning with immediate feedback is the most effective for reaching at-risk students. Of all the teaching techniques tested, only flipping the classroom significantly improved student understanding of harmony and counterpoint, likely because it promoted deeper student engagement with the course material.

While the most effective teaching strategy may not be the same for more knowledge-based courses like undergraduate music history, achieving student success will always depend upon how effectively the instructor elicits student engagement with the course material.[12] This analysis of various teaching techniques' efficacy in the undergraduate music theory classroom seeks to ground the discussion about best pedagogical practices with conclusive empirical data. Further research by other instructors will surely uncover additional useful information about effective techniques for student success. By adopting active learning strategies and rigorously testing their effectiveness, we can build a body of evidence for best practices in college-level music pedagogy.

---

[12] Carini, Kuh, and Klein 2006.

Acknowledgments

References

Bakker, Sara. 2020. "Creating Measurable Learning Objectives." *Engaging Students: Essays in Music Pedagogy*, vol. 7.

Byrd, Christy M. 2016. "Does Culturally Relevant Teaching Work? An Examination From Student Perspectives." *Sage Open* 6, no. 3.

Carini, Robert M., George D. Kuh, and Stephen P. Klein. 2006. "Student Engagement and Student Learning: Testing the Linkages." *Research in Higher Education* 47 (1): 1–32.

Chenette, Timothy. 2018. "Part Writing Can Be Student Centered." *Engaging Students: Essays in Music Pedagogy*, vol. 6.

Duker, Philip, Anna Gawboy, Bryn Hughes, and Kris P. Shaffer. 2015. "Hacking the Music Theory Classroom: Standards-Based Grading, Just-in-Time Teaching, and the Inverted Class." *Music Theory Online* 21, no. 1.

Duker, Philip, Kris Shaffer, and Daniel Stevens. 2014. "Problem-Based Learning in Music: A Guide for Instructors." *Engaging Students: Essays in Music Pedagogy*, vol. 2.

Erwin, Christopher, Melissa Binder, Cynthia Miller, and Kate Krause. 2021. "Performance-Based Aid, Enhanced Advising, and the Income Gap in College Graduation: Evidence From a Randomized Controlled Trial." *Educational Evaluation and Policy Analysis* 43 (1): 134–153.

Gopalan, Maithreyi, and Shannon T. Brady. 2019. "College students' sense of belonging: A national perspective." *Educational Researcher* 48 (8): 466–478.

Harrison, Daniel G. 1994. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents*. University of Chicago Press.

Hernandez, Susana. 2022. "Closing the Achievement Gap in the Classroom Through Culturally Relevant Pedagogy." *Journal of Education and Learning* 11 (2): 10–20.

Johnson, Vicky. 2015. "Proficiency-Based Learning with Muscle in a Music Theory Classroom." *Engaging Students: Essays in Music Pedagogy*, vol. 3.

Kelley, Robert T. 2006. "A Mathematical Model of Tonal Function." Paper delivered at the annual meeting of Music Theory Southeast.

———. 2017. "A Corpus-Based Model of Voice Leading in Tonal Music." Paper delivered at the annual meeting of the Society for Music Theory.

Lang, James M. 2016. *Small Teaching: Everyday Lessons from the Science of Learning*. Jossey-Bass.

Murphy, Mary C., Maithreyi Gopalan, Evelyn R. Carter, Katherine T. U. Emerson, Bette L. Bottoms, and Gregory M. Walton. 2020. "A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university." *Science Advances* 6, no. 29.

Nilson, Linda. 2015. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*. Stylus Publishing.

Quinn, Ian. 2005. "Harmonic Function Without Primary Triads." Paper delivered at the annual meeting of the Society for Music Theory.

Shafer, Jennifer. 2018. "'Calling Captain Scaffold!' Interactive Handouts to Stimulate Student Learning." *Engaging Students: Essays in Music Pedagogy*, vol. 6.

Shaffer, Kris. 2014. "Tendency tones and functional harmonic dissonances." Chapter 23 of *Open Music Theory, Version 1*, edited by Kris Shaffer and Robin Wharton. Hybrid Pedagogy Publishing.

Simonson, Shawn R., ed. 2019. *POGIL: An Introduction to Process Oriented Guided Inquiry Learning for Those Who Wish to Empower Learners*. Routledge.

Snodgrass, Jennifer. 2020. "Foreword for *Engaging Students* Vol. 7." *Engaging Students: Essays in Music Pedagogy*, vol. 7.

Talbert, Robert. 2017. *Flipped Learning: A Guide for Higher Education Faculty*. Routledge.

Theobald, Elli J., Mariah J. Hill, Elisa Tran, Sweta Agrawal, E. Nicole Arroyo, Shawn Behling, Nyasha Chambwe, Dianne Laboy Cintrón, Jacob D. Cooper, Gideon Dunster, Jared A. Grummer, Kelly Hennessey, Jennifer Hsiao, Nicole Iranon, Leonard Jones, Hannah Jordt, Marlowe Keller, Melissa E. Lacey, Caitlin E. Littlefield, Alexander Lowe, Shannon Newman, Vera Okolo, Savannah Olroyd, Brandon R. Peecook, Sarah B. Pickett, and David L. 2020. "Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math." *Proceedings of the National Academy of Sciences* 117 (12): 6476–6483.

Yu Wang, Anna. 2022. "Teaching the Cultural Commitments of Musical Structures and their Theories." *Engaging Students: Essays in Music Pedagogy*, vol. 8.