Journal of Music Theory Pedagogy

Volume Thirteen 1999

SCHOOL OF MUSIC
THE UNIVERSITY OF OKLAHOMA

JOURNAL OF MUSIC THEORY PEDAGOGY

EDITORIAL & BUSINESS OFFICE

Journal of Music Theory Pedagogy
School of Music
University of Oklahoma
Norman, OK 73019

SUBSCRIPTION RATE (1 issue/yr)

Individual	$25.00
Student/Emeritus	$15.00
Library/Institution	$35.00

(International: $3.00 additional)

Information about back issues and special rates available through the business office

The Gail Boyd de Stwolinski Center
for Music Theory Pedagogy
James H. Faulconer, Director

ISSN 0891-7639

CONTENTS

Volume 13

EDITORIAL

In assuming the editorship of this journal, I am mindful of the contributions that my predecessors, Michael Rogers and Mary Wennerstrom, have made toward its current stature. Prof. Rogers set a high standard with his eloquent, inaugural essay and his subsequent editorial work. Prof. Wennerstrom maintained that standard and expanded the scope of the journal's offerings. We all owe Michael and Mary a huge round of applause for their dedicated leadership and service to our profession.

With the help of the Editorial Review Board, I shall strive to uphold the same standards and expand the journal's scope as appropriate. We hope to identify and publish the very best of current thinking about the teaching of music theory and aural skills. While these ideas will typically be couched in the traditional form of scholarly articles and reviews, authors are encouraged explore the potential of interactive electronic multimedia and the World Wide Web for depicting salient aspects of their work. I and other staff members shall endeavor to provide consultation, within the limits of our expertise, regarding the preparation and presentation of such documents. The journal's web site (http://music.ou.edu/jmtp/) already provides a listing of the contents for all volumes. Links to web documents that supplement journal articles can easily be added at an author's request.

This volume features two articles that address the complex issue of testing and placing entering students in the sequence of required courses in music theory and aural skills. Developers of placement tests are obliged to document the validity and reliability of such tests. Barbara Murphy and Timothy Smith describe two contrasting approaches to testing and then explain how they assessed content validity, the ability of their tests to predict student performance in entry-level courses. In developing their articles, both authors profited from careful reading and constructive criticism provided by my colleague Patricia Sink, Associate Professor of Music Education, School of Music, University of North Carolina-Greensboro. We are grateful to Prof. Sink for providing prompt and thorough consultation in her area of expertise.

Two additional articles explore novel approaches to teaching basic theory and analysis. Bruce Kelley documents the philosophical background of two movements that have influenced college-level teaching in recent decades. He then explains and evaluates his attempt at implementing the writing-to-learn approach in a harmony course. Joseph Swain continues to explore his long-standing interest in harmonic rhythm by describing a method he developed to get students involved with and thinking about this important aspect of music.

Rounding out this volume are critical reviews of two recent conferences on music cognition and perception. While research findings in this discipline could have profound implications for the way we teach music theory and aural skills, the pedagogical relevance of current projects, as described by Edward Klonoski and James P. Mathes, varies considerably. In spite of these uneven results, we shall continue to identify, summarize, and interpret music cognition research projects. Hopefully, such efforts will stimulate a dialogue between researchers and theory pedagogues about the nature of musical knowledge and the process of imparting such knowledge to our students.

J. Kent Williams
University of North Carolina-Greensboro

Predictive Validity of the "Ready or Not" System for the Assessment of Students Needing Remediation in Music Theory

Timothy A. Smith

A perennial problem in the teaching of music theory is how to identify students who are insufficiently prepared for the harmony/ear-training sequence. If it is a logistical problem—how to convene, grade, and advise such a throng before registration, it is also a problem of assessment—what skills and knowledge should one test, how exactly should a student proficient with the major key signatures (but not minor) be advised, and how poorly must one perform to be steered onto the remedial route?

To solve the first problem, some instructors have devised computerized versions of paper-and-pencil tests. While such assessments have the advantage of walk-in administration, online grading and advising, they do not ordinarily employ the more powerful computations of so-called "expert" systems: non-linear and branching pathways, algorithmic computation of probabilities, content shaped by in-progress assessment, random generation of problems, databases, empirically informed cutoffs or norm-referenced advisement.

Given the hierarchical organization of basic music theory, the computer is the perfect tool for tailoring test content, in real time, to the perceived strengths and weaknesses of individual students. Having interpolated reliable estimations from failed attempts at easier problems, for example, intelligent systems are easily engineered to skip more difficult problems where students have little or no probability of success.[1]

[1]The objective is a faster, but no less reliable, assessment. In research leading to the creation of "Ready or Not," it was shown that a high percentage of students who missed an item missed all harder items (and students who solved an item solved all easier items). That measure, as quantified by Cronbach's Alpha, a statistical indication of internal consistency more powerful than a split-half reliability, was .9727.

In 1994 the *Journal of Music Theory Pedagogy* included a report of a project undertaken by this author to design and implement such a system.[2] That article described research preceding the creation of the instrument as well as ExSPRT control structures which enabled it, normally, to identify target populations in under half an hour.[3] The system has since come to be known by the acronym RON (Ready or Not). The RON system was programmed to select skills to be tested from the objectives listed in Table 1; examples of stimuli for presentation and measurement of selected objectives are presented in Appendix II.

RON begins by selecting at random an objective with high discrimination; can the student, to cite one example, notate pitches that are enharmonically equivalent. RON first provides a brief explanation of "enharmonic," then writes a random pitch (in the student's favorite clef) and instructs the student to write its equivalent. The visual stimulus for this type of problem is represented in Figure 1.

RON was designed to present items in response to the test taker's performance on preceding items. Following each attempt, RON would employ algorithms to calculate the probability that the student has mastered this particular objective.[4] When RON is able to

---

[2]"Timothy A. Smith, "An ExSPRT Systems Approach to the Assessment of Students Needing Remediation in Music Theory," *Journal of Music Theory Pedagogy* 8 (1994): 179-200.

[3]The "SPRT" of "ExSPRT" stands for "Sequential Probabilities Ratio Test." The test uses Bayesian mathematics continuously to revise estimates until a reliable prediction can be made, at which time the test is concluded. The mathematical equation at the heart of this system may be found in the next footnote.

[4]The following equation computes a Probabilities Ratio (PR). Because this is done after each problem, the algorithm may be understood to compute Sequential Probabilities Ratios (or SPR, see fn. 3 *supra*). The equation follows in which: Pm = the percentage of students passing the course (in the 1990 study) and who could do this type of problem before having had instruction, Pn = the percentage of students NOT passing the course who could to this type of problem before having had instruction, S = successful answers by the current testee, F = unsuccessful answers by the current testee. (Pm and Pn for "writes enharmonic equivalents" were 85% and 53% respectively.)

$$PR = \frac{Pm^s\,(1\text{-}Pm)^f}{Pn^s\,(1\text{-}Pn)^f}$$

Table 1: Possible Test Items in RON
completes a measure with missing durations
identifies last pitch of a phrase in C Major
identifies correctly paired notes/rests
identifies longest/shortest of four easy rhythms
identifies non-equivalent rhythms
identifies starting pitch of a song
identifies the root of a triad
knows number of sharps and flats in key signatures
matches sounding triad (M m + d) with written
names basic symbols of notation
names pitches in treble alto and bass clefs
plays pitches of treble and bass clefs on keyboard
understands the function of sharps and flats
writes enharmonic equivalents
hears and writes M/m melodic seconds
writes pitches in treble alto and bass clefs
writes sharps/flats of key sig. (correct order and position)
writes short diatonic melody (dictated)

make a reliable prediction, it leaves the objective for another.[5] If a student masters problems of high discrimination, RON switches to a more difficult cluster, where the prognosis of mastery is more quickly confirmed. Conversely, if RON's preliminary assessment indicates that a student may be poorly prepared, it moves to easier

---

Following each problem, the value of PR is compared with three evaluative statements in which: α = error tolerance for misidentifying masters as non-masters, and β = error tolerance for misidentifying non-masters as masters.

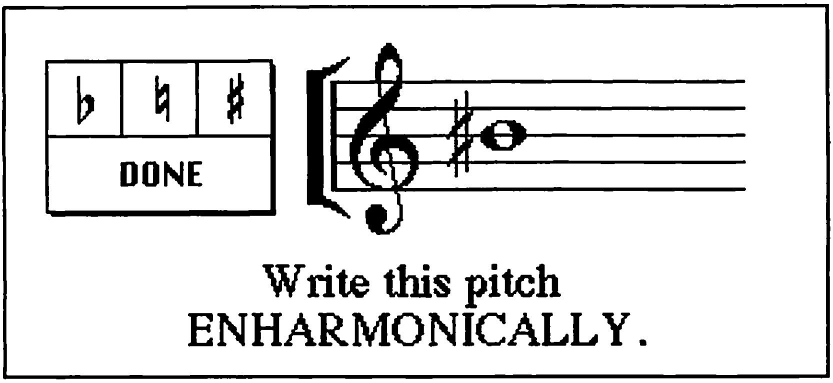
If PR $\geq (1-\beta)/\alpha$ then the student has demonstrated mastery of the objective.

If PR $\leq \beta/(1-\alpha)$ then the student has demonstrated non-mastery of the objective.

If PR $\beta/(1-\alpha) <$ PR $< (1-\beta)/\alpha$ it is not possible to make a reliable prediction, then give the student another problem of the same type.

[5]In instances where failed attempts to answer the question balance successful attempts, RON terminates the line of questioning and ranks the student as "inconclusive" on that particular objective. Inconclusive assessments on individual items do not count for, or against, the student in the overall assessment.

Figure 1: Stimulus for High Discrimination Item (Writes Enharmonic Equivalents)



items where a lack of preparedness can be more quickly established.[6] As one can see from the foregoing description, RON is non-linear and unrepeatable in precisely the same format.

While the 1994 article did establish the internal consistency of the test in a measure known as Cronbach's Alpha (fn. 1 *supra*), it did not, other than to present a handful of case studies, make any assumptions about predictive validity, sufficient data being at that time unavailable. The current article represents analyses of data accumulated since then with a view to establishing the predictive validity of RON for identifying students at risk.

While this article reports on important information for teachers of music theory and individuals involved in the creation and administration of placement tests, it is possible, without sufficient context, that the current research may be difficult to interpret without reference to the companion article (*JMTP* 8, 1994). The introduction has, hopefully, redacted sufficient information to render the conclusions of this investigation understandable. However, readers with a particular interest in the mechanics and pedagogical issues of non-linear tests may wish to study the earlier report (see fn. 2).

The target population for this investigation was comprised of 400 Ball State University students commencing a major in music

[6]Demonstrated mastery of difficult items (or non-mastery of easier items) allows the test to be concluded without moving through irrelevant items in the test inventory. Please refer to fn. 1, *supra*.

from the fall of 1992 through 1995. Data were at hand for 281 students assessed by RON as high school seniors and who subsequently matriculated as music majors, completing a music theory course for grade. No data existed for the more than one hundred students who enrolled without advisement.[7]

Four out of five individuals in the data group completed the first course of the music theory sequence while the other fifth completed the remedial course.[8] Inasmuch as the purpose of this study was to assess the predictive validity of RON as it pertained to performance in the music theory sequence, with the exception of those whose remedial attempts resulted in failure, students enrolled in the remedial course were excluded from consideration.[9]

Because half of the students advised into remediation effectively self removed from the sample by having heeded RON's advice the data were inherently biased.[10] Inasmuch as these individuals chose

---

[7]Because the data pool was not selected at random, it was not possible to infer properties of the larger population from known values. The sample does, however, properly adduce attributes to a sub-group of high school seniors with the means and inclination to participate in on-site assessments prior to matriculation. Conclusions of this report with respect to predictive validity may therefore be considered appropriate in instances where RON has been utilized prior to matriculation, preferably as high school seniors.

[8]Although two fifths of sampled students were advised into the remedial course, half of them opted to disregard RON's notice, attempting the first course in the sequence outright. More than two thirds of this latter group failed.

[9]Throughout this report failure of the course is defined as having earned a grade of D or F. Although D was a passing grade at Ball State, students who earned D's in the theory sequence were required to repeat the course. It seemed reasonable to assume that students who failed remediation would have also failed the first semester of the sequence. Accordingly, the sample retains data on students who engaged themselves in unsuccessful attempts at remediation.

[10]Whereas the statistical ideal would have required denial of student access to the results of their RON assessments, enrollment of students in the sequence regardless of preparedness, would have been unconscionable in view of what was, from case studies, already known about the utility of the instrument. The inclusion, in the data pool, of failed attempts at remediation represented an effort to compensate for bias by reconstituting a reasonable approximation of the ideal population.

not to leave themselves in a position where lack of preparation might have been established empirically, inaccurate estimations of this study would be expected to have erred on the conservative side.

As it was first constructed, RON assessed students either as prepared, unprepared, or inconclusive. For the purposes of this study, the data justified a six-point ranking (Table 2 below) comprising the predictive variable. The outcome variable was represented by the student's grade at the end of the semester.

The bivariate data of this study were readily adapted to two-way contingencies like that of Table 3. While these data indicate

Table 2: Predictive Variable (quantified by RON rankings of 1-6)

RON score Predictive Variable	Qualification for Rank	Prognosis
1	identified by RON as prepared, having mastered at least 85% of objectives	exceptionally well prepared for Music Theory
2	identified by RON as prepared, having mastered 70 - 84% of objectives	prepared for Music Theory
3	identified by RON as prepared, having mastered at least 70% of objectives	may not need remediation, may be prepared
4	ran out of time (or objectives) without RON having made a prognosis	may need remediation, may not be prepared
5	identified by RON as unprepared, having mastered at least 15% of objectives	probably needs remediation
6	identified by RON as unprepared, having mastered less than 15% of objectives	definitely needs remediation

that RON assessments correlate with term grades, the direction of correlation cannot be taken for granted.[11]

The selection of an appropriate measure to interpret these data required a determination of the purpose of the measure—in this case, to establish the validity of RON's predictions. Available measures depended, in part, upon whether the data were ranked in a continuum of infinite range (scalar) or partitioned into discrete units (categorical). Both the semester grade and RON's prognosis are, of course, categorical rankings. As one can see from Table 2, RON

Table 3: RON Scores by Grade

Recommendation	Frequency Total Percent Row Percent Column Percent A	B	Grade C	D	F	Total
Advised into Mus Th Sequence (RON score 1,2,3)	28 11.24 20.00 93.33	34 13.65 24.29 73.91	42 16.87 30.00 67.74	22 8.84 15.71 39.29	14 5.62 10.00 25.45	140 56.22
Advised into Remediation (RON score 4,5,6)	2 0.80 1.83 6.67	12 4.82 11.01 26.09	20 8.03 18.35 32.26	34 13.65 31.19 60.71	41 16.47 37.61 74.55	109 43.78
Total	30 12.05	46 18.47	62 24.90	56 22.49	55 22.09	249 100.00

[11]While any symmetrical value could have been used to establish correlation, the direction of the correlation could not, in such instances, be assumed. Most comparisons, for example, produced Chi-Square values of less than .0005 indicating statistically significant correlation between variables. Such correlation could not indicate, however, which one, or both, of the variables was dependent. Thus, Chi-Square could not properly be used to distinguish between RON assessments as estimators of term grades or the reverse. Table 4 presents data that does indeed establish the direction of correlation.

scores are the product of asymmetrical cutoffs applied *a priori* to prognoses of mastery or non-mastery.[12]

Three measures commonly used to establish statistical relationships are Goodman-Kruskal Gamma (γ), Kendall's Tau-b (τ_b), and Somers' D. The problem driving each of these measures is: for any randomly-selected pair of students, what is the probability that one student's RON score and subsequent term grade will be concordant, discordant, or tied with the other's.[13] This analysis employed Somers' D as the appropriate measure because it was capable of making finer distinctions between variables even in instances where one or both were tied.[14]

---

[12]Such rankings do not allow one to quantify how much more of a master a student ranked 1 is over a student ranked 2. Similarly, whereas a semester grade of F could have represented a range of from 0-69%, all other grade units were comprised of 15% increments. While it is possible to assert that an F is lower than a B, it is impossible to determine how much lower. While the six categories of the predictive variable were not scalar, RON calculated them by means of probability ratios of from zero to one hundred, with an infinite range of decimals between. Product of a Bayesian equation at the heart of the ExSPRT system (see fn. 4 *supra*), a continuous scale was reconstructable, given extant data, and permitted a comparative analysis predicated upon scalar, rather than ordinal, formations. It was hypothesized that validity might be improved by means of ratio variables admitting quantitative comparison. Such a comparison showed, unexpectedly, that the categorical format was, in this instance, more reliable.

[13]Goodman-Kruskal Gamma represents a conditional probability, the condition being a lack of ties. Kendall's Tau-b and Somers' D differ from Goodman-Kruskal Gamma, primarily, in how they recognize ties. Kendall's Tau-b accounts for ties as a value equal to the geometric mean of all types of tied observations. Kendall's Tau-b is, therefore, an unconditional probability lending itself specifically to the analysis of ordinal bivariate data where ties are likely to occur. The disadvantage of Kendall's Tau-b is that it is incapable of making distinctions between the different types of ties. Of the three measures under consideration, only Somers' D distinguishes between ties having ramifications as to the independence of variables with ties accruing to either variable, or both. It does this by expressing itself in two directions. In terms of asymmetric, or two-way, contingencies, these values are commonly phrased as row-predicts-column, versus column-predicts-row.

[14]For example, Joe's RON score may be the same as Jane's but their term grades may be different. In this instance the two observations are

Somers' D for the data represented in Table 3 is .519 (see Table 4 below). This value is indicative of a moderate relationship between RON's categorical rankings and subsequent grades earned in the first semester of harmony/ear-training. Perhaps of more significance, this figure shows that the RON score is in proper orientation to the term grade. In other words, RON functioned as it was designed: to rank order students as per a value congruent with eventual performance in the music theory course.[15] But how reliable was this rank order, and can it be used to predict performance?

Measurements of Somers' D may range from -1 to 1. These values indicate, respectively, that every pair of observations is either discordant or concordant. A measure of zero indicates lack of correlation, with neither variable manifesting statistical evidence of contingency upon the other. Positive values are indicative, therefore,

Table 4: Predictive Validity of RON's Asymmetric Categories

	cut-off point used to predict ...adequate preparation	...need for remediation	Somer's D	Asymptotic Standard Error
Asymmetric Categories	1-3	4-6	.519	.059
(1-6 RON rankings)	1-4	5-6	.517	.072

concordant if the RON score is viewed as the independent variable but discordant if the term grade is viewed as independent. By contrast, Joe's and Jane's term grades may be the same, but their RON scores may be different. Such a tie would indicate concordance with respect to the outcome, but discordance with respect to the prediction. Finally, Joe's and Jane's RON scores and term grades may be identical, indicating a degree of concordance without implying the independence of either variable.

[15]This can be illustrated by reversing the direction and using the final term grade to "predict" the RON score. With such an orientation, Somers' D drops to .323. Please note that what is being described here is not a cause-and-effect relationship. The RON score did not cause the term grade. What is being described, instead, is a statistical measure showing a moderate relationship between variables allowing one to conclude that a student's performance in the sequence may mirror the RON score.

of correlation in which the predictive measure is in proper orientation to the outcome. Such a value is more reliable the further removed it is from zero.[16] Somers' D of .519 would indicate, therefore, a borderline acceptable degree of predictive validity.[17]

How much confidence should one place in such a figure? Some discussion is in order. As a predictive measure, RON scores accounted for 26.9 percent ($.519^2$) of the variance in the music theory grades with 5.9 percent prediction error. It is important to emphasize that the positive linear correlation (Somers' D .519) implies simply that low RON scores can be associated, to a degree, with poor performance in a music theory class. In other words, the "prediction" is not that a student with a low RON score will fail music theory. It is, rather, that in any population of students who have taken RON, the rank order that it produces will correspond, to a certain extent, with the eventual term grade distributions in the music theory class.

But how well does RON identify students at risk? If RON ranks at-risk students at the top, then students at the lower ranks would also be at risk. Conversely, if RON places well-prepared students at the bottom, then students in the higher ranks should outperform them. So, while a correlation coefficient of .519 may warrant some degree of confidence in RON's rankings, it does not indicate where the cutoff should be and whether RON merely happened upon that value. At what point should a student be required to take the remedial course?

---

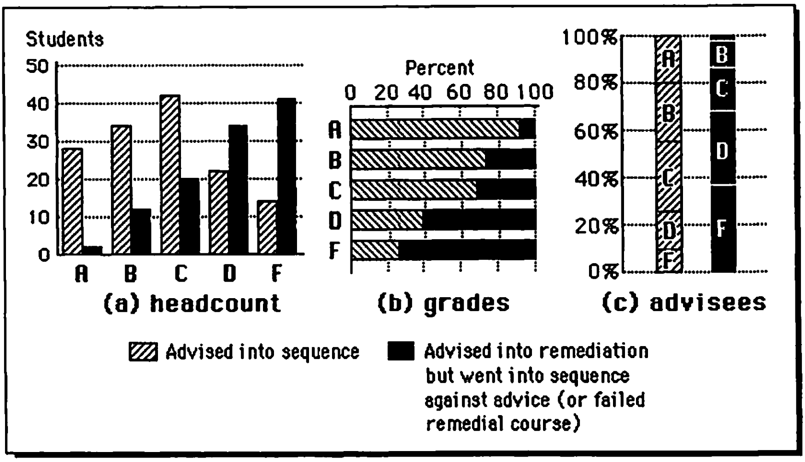
[16]So, if Somers' D for RON had been +1, RON assessments could have been used to predict that every student with a higher score than his fellows would have earned a higher term grade in the sequence then his fellows. Conversely, if Somers' D for RON had been -1, RON assessments could have been used to predict that every student with a LOWER score than his fellows would have earned a HIGHER term grade in the sequence then his fellows (i.e. reliable as a predictor, but in the wrong direction). If Somers' D for RON had been zero, it could not have been used to predict anything.

[17]A correlation coefficient of less than .50 is generally considered to be invalid for predictive purposes. The Somers' D value of .519 falls within the range of acceptable. In statistical analyses of this type, a prediction error of 5% or less would have been considered acceptable. While RON's prediction error of 5.9% is slightly high, it does approach the accepted level and should decline as the test continues to be developed.

In answer to the question of cutoffs, the reader is referred to Table 5. The first view (5a) represents a comparative headcount of students whom RON advised into the remedial course versus those advised into the music theory sequence. The second view (5b) represents percentiles of remedial vs. non-remedial students within each grade bracket. The third view (5c) compares grade distributions of those advised into remediation vs. those advised into the music theory sequence. Note that more than two thirds of the failing grades were earned by students whom RON assessed as needing remediation (Table 5b). Similarly, more than two thirds of those advised into remediation either failed the first semester of the sequence or the remedial course itself (Table 5c).[18]

As noted earlier, RON's prognosis was represented in categorical rankings based upon asymmetrical cutoffs (see fn. 12 *supra*). To determine if a continuous scale might establish a more reliable cutoff, probability ratios for each student were converted to a 1-10 scale,

Table 5: Three Comparisons of Outcomes Between Students Advised into the Sequence Versus Students Advised into Remediation



[18]These percentages are, of course, exclusive of individuals who followed RON's advice to remediate and who subsequently passed the preparatory course.

contributing not only four more categories, but also a truly symmetrical scale capable of rendering quantitative distinctions.[19] Using this new estimator, Somers' D values were calculated and compared with the old, yielding the symmetric categories of Table 6.

Table 6: Comparison of the Validity of Asymmetric vs. Symmetric Categories

	cut-off point used to predict ... adequate preparation	... need for remediation	Somer's D	Asymptotic Standard Error
Asymmetric Categories	1-3	4-6	.519	.059
(1-6 RON rankings)	1-4	5-6	.517	.072
Symmetric Categories	1-9	10	.496	.063
(1-10 Continuous Scale)	1-8	9-10	.482	.061
	1-7	8-10	.459	.062
	1-6	7-10	.434	.062

Table 6 shows, surprisingly, that while the decision to delineate the predictive variable (RON score) in asymmetric categories was intuitive, it happened upon a linear function that better predicts outcomes both in terms of Somers' D and minimization of Asymptotic Standard Error (hereafter referred to as ASE). Whether the remedial line is drawn at 4 or 5, assessments are more reliable using RON's asymmetrical categories than the reconstructed symmetrical scale.

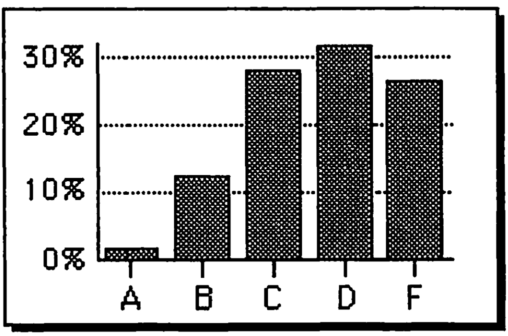
In addition to showing that asymmetrical categories are more reliable, Table 6 suggests that the predictive validity of RON would not be unduly compromised if students with RON rankings of 4 are advised either direction: to remediate or proceed with the first se-

[19]While RON's asymmetrical assessment was not expressed in continuous variables, it was possible to reverse-engineer Sequential Probabilities Ratios (see fn. 4 *supra*) to reconstruct the values RON used to determine if students were prepared. These values comprised a continuous spectrum of from zero to one hundred with an infinite range of decimals between.

mester of music theory.[20] Had students with RON rankings of 4 been advised into the first semester, Somers' D would have been .517. But with the same group advised to remediate, Somers' D was .519. The minimal difference (.002) in validity is negated by possible ASE of .072 and .059, respectively. Inasmuch as the validity of the instrument would not have been compromised, Table 7 shows that it is clearly in the interest of these students to remediate. Fifty-eight percent of their peers who proceeded to take the first semester of harmony/ear-training, without such treatment, ultimately failed. Table 7 represents grade distributions for this group.

RON was engineered to render assessments after having evaluated at least five objectives. Accordingly, it should be understood that the Somers' D value of .519 discussed to this point represents RON's predictive validity as a composite of five or more learning

Table 7: Grade Distributions. Students with Inconclusive Assessments Overall (RON Rank 4)



[20]Recall that RON reserved the rank of 4 for individuals for whom it could make no reliable estimate of preparedness within the allotted time (see Table 2). In other words, the rank of 4 technically represents no prognosis at all—at least not from RON. But this is not to say that these students could not be ranked. Because this research established that students assessed as inconclusive overall performed at a level between those whom RON had placed into what are now identified as rankings 3 and 5, the category of 4 was created to represent them.

objectives. In other words, discussions of validity, till now, have focussed upon RON as an instrument, and not upon the validity of its individual learning objectives.[21] We examine now the criterion validity of each objective. These shall be presented roughly in order from the least, to the most, reliable. Somers' D values for each objective as predictor of the term grade (and vice versa) may be found in Appendix I.

RON's assessment of the ability to identify the last pitch of a heard phrase in C Major produced the only negative Somers' D value of the study. When assessments of this skill alone were used to predict passes and failures, Somers' D was -.164, suggesting that a student might be even more likely to pass the course if RON assessed him as a non-master of this objective! The high ASE (.235) indicates that this prognosis is suspect. Because omission of the objective effects a slight decrease in Somers' D for the instrument, it would appear that it does indeed contribute somewhat to the overall validity.[22]

Two of the objectives appear, at first glance, neither to have added to, nor subtracted from, the assessment. Of individuals asked to identify correctly paired notes and rests, for example, one was as-

---

[21]Although experimentation with a continuous scale did not produce a more reliable predictor than RON's asymmetrical categories, the symmetrical approach was instructive in that it enabled calculation of predictive reliabilities (as expressed in Somers' D) for individual learning objectives (*vis a vis* the RON instrument overall). These values are detailed in the Appendix I which shows that individual objectives as predictors of the term grades were, in every case, more accurate than the reverse. The scalar predictor allowed, too, for comparisons of hypothesized configurations in which one or more learning objectives had been removed: e.g. to what extent would RON have been more, or less, reliable if Objective No. 1 had not been part of the system.

[22]It is possible that one defect may be the manner in which the objective is contextualized. When told that they will hear a phrase of music in the key of C Major, students are instructed to write the last pitch. The purpose of the problem is to assess the extent to which the testee perceives tonal centricity and scale degree function. No explanation of these functions is given, it being assumed that individuals who possess the skill will understand, intuitively, what they are being asked to show, and show that they can do it by writing the last pitch they hear. It is possible that with additional explanation this line of assessment might yield better results.

sessed as not having mastered the objective, while all others were assessed as inconclusive.[23] Similarly, every person asked to name the basic symbols of notation was assessed as inconclusive.

The 1990 phase of this research showed that both of the above objectives were achievable by a high percentage of persons who eventually failed the first course in the music theory sequence.[24] It was thought, therefore, that persons tested as non-masters of these particular objectives would be poor candidates for music theory. While inconclusive assessments of individual objectives did not count against students, RON assessed 77 percent of those who tested inconclusive in their ability to match notes and rests as unprepared on the basis of deficiencies in other areas (63 percent of these students ultimately failing the course). Similarly, RON ranked 75 percent of those who tested inconclusive in their ability to name the basic symbols of notation as unprepared because of lack in other pre-skills (61 percent of them ultimately failing the course).

That sampled students to whom the above two objectives were directed showed themselves, overall, to be unprepared, indicates that RON functioned as expected. This holds true in spite of the fact that most were assessed as inconclusive with respect to these two objectives. It bears repeating that the sample omitted individuals whom RON advised to remediate and who completed that remediation successfully. Of this latter group, nearly one third who were tested on these two objectives were assessed as not having mastered them. Thus, if students advised to remediate (and who heeded RON's advice by enrolling in the remedial course and passing it) were actually unprepared at the time of assessment, then these two objectives were efficacious.[25]

---

[23]The student assessed as not having mastered this objective earned a grade of D in the sequence.

[24]Of those who failed, 86% could pair notes and rests (compared to 94% of those who passed), and 90% could name the basic symbols of notation (compared to 96% of those who passed).

[25]There was no way, from this study, to have established this empirically. Recall that half of the students advised to take the remedial course followed that advice, and half did not (fn. 8 *supra*). Only the half that ignored the advice (and those who followed it but failed the remedial course) were included in this study. It is this group that was assessed as

RON's assessment of a student's ability to write pitches in treble, alto, and bass clefs predicted student performance with low Somers' D of .083. The ASE of .083, indicates that this assessment may be suspect.[26] If it is reliable, low congruity shows a limited contribution to the overall assessment. Omitting this objective from the analysis, Somers' D for RON assessments increases slightly indicating that the removal of this objective might increase the validity of the system.[27]

Of the forty-eight individuals required to match sounding triads (M, m, +, d) with written triads, all but two tested as masters. The two exceptions tested as inconclusive. The 1990 phase of research found that 22 percent of students who passed, but only 5 percent of those who failed, the first semester of music theory could do this before having had instruction. Unlike easy objectives, which were presented to students whom RON suspected as being unprepared, this difficult objective was presented to students whom RON suspected as being prepared. The purpose of the objective was to confirm the initial prognosis quickly, after which the assessment could be concluded.

A traditional evaluation of the effectiveness of individual test items might have concluded that, whereas the ability to match sounding triads with written did not distinguish between those who were prepared and those who were not (all but two students testing as masters), it should be eliminated. It turns out, however, that 71

---

inconclusive on these two objectives. Of the group that did follow RON's advice, took the remedial course and passed it, nearly one third were assessed as not having mastered these objectives. It was not possible to establish that this latter group was truly unprepared (as RON had said it was) without them having failed something. Instead, they passed the very remedial course that RON had advised them to take.

[26]Although it is unusual that a correlation coefficient would be the same as its ASE, in this instance the values were indeed found, after repeated calculation, to be identical!

[27]Before deleting the objective, further research should be conducted to verify the original estimates of mastery and non-mastery with respect to writing pitches in various clefs. Whereas the 1990 research produced different estimates for each clef, RON was designed to merge these data, selecting clefs at random, and averaging estimates with each problem. If the original estimates were correct, the method used for merging data may be inaccurate.

percent of these students evaluated as having mastered the objective were assessed, on the basis of other testing, as prepared for the music theory sequence overall. Seventy-seven percent of this group passed the course. Once again, the idiosyncratic design of the ExSPRT system was such that this type of problem was given primarily to students who actually were masters. The effectiveness of such a strategy is revealed, ultimately, in the fact that overall predictive validity of the system, as shown in Somers' D, drops when this objective is omitted.

RON's assessment of a student's ability to identify triad roots produced a low Somers' D value of .111. The ASE (.116), being higher than the estimate, allows a possibility that the estimate may be off. When this skill was used to predict actual grade distributions, Somers' D increased to .184 with an ASE of .125. When this objective is omitted, predictive validity of the assessment decreases by two percent. This suggests that the objective contributes to the validity of the system, overall, and that it should be retained in spite of its performance as an individual indicator.[28]

Of the objectives identified in the 1990 phase of research, the ability to write, from dictation, a short diatonic melody promised to distinguish most quickly between prepared and unprepared students. The current research yielded Somers' D of .249 for this objective as indicator of actual grade distributions (ASE.126). Were RON to have omitted this objective Somers' D would have dropped by 1.5 percent. Whereas the objective does contribute to the validity of the system overall, it appears to be less dramatic than originally anticipated.

RON made no inconclusive assessments of any of the 211 individuals it required to notate from dictation a short diatonic melody. Of all learning objectives, this one alone identified every student as either proficient or deficient, indicating that the original estimate of spread between the two was excessive. Wide estimates typically

---

[28]Unlike the previous objective, where too little set-up information may have been problematic, the problem here may have been too much information. Students who did not know what a triad root was were given the opportunity to take a tutorial before attempting the assessment. It is possible that omitting this tutorial may give a more reliable estimate.

cause the ExSPRT system prematurely to conclude assessments, and therefore not as reliably as it should.[29] Another possibility is that RON's diatonic melodies, which were all stepwise, were too easy compared to melodies of the original research (which contained some leaps of a third). RON has subsequently been revised to play melodies including thirds.

The ability to demonstrate understanding of the function of sharps and flats predicted grade distributions to an accuracy of Somers' D .446, with an ASE of .115. It should be noted that, in addition to testing the student's understanding of how accidentals function, this item also evaluated the student's understanding of whole-steps and half-steps. Because the setup information appeared to confuse some students, this research was expected to show that the question was unreliable, or at least required modification. Surprisingly, the data show that this objective is the fourth most reliable predictor of performance.[30]

With Somers' D of .522 and an ASE of .173, the ability of students to identify the longest or shortest of four easy rhythms was the third most reliable objective. After being presented with four rhythms, three of which spanned the same number of beats, students were asked to indicate which of the four was longer (or shorter, as the case may be).

The second most reliable objective was the ability of students to identify the starting pitch of a song. Here the student was told to write, in the key of C, the starting pitch of a folk song. Students were allowed to opt out of tunes they did not know. The correlation coefficient for this objective was .586, with an ASE of .132.

---

[29]The 1990 research showed that 83% of students who passed the first semester of theory could write short diatonic melodies, but that only 33% of those who failed could do it.

[30]The testing of this objective required explanation that appeared to puzzle students and sometimes required verbal clarification. The setup begins with the computer writing a pitch in the student's favorite clef. This pitch may be unaltered, double-flatted, flatted, sharped, or double-sharped. To the left of the staff a grid represents these signs, and the student is instructed as follows: e.g. "Choose the symbol that, when substituted for the given symbol (if any), would produce a pitch one HALF step HIGHER" (or LOWER, depending upon the question).

The most reliable objective was the ability to hear and to write melodic seconds.[31] Here the computer displayed the first pitch in the student's favorite clef then played a second pitch a major or minor second removed. Students were required to notate the second pitch (enharmonic equivalents were counted as correct). Students could play the interval as many times as they wished. With low ASE of .065 and .078 respectively, Somers' D for this objective was .824 as predictor of term grades and .706 as predictor of pass/fail.

Using its asymmetrical rankings, the validity of RON overall was established at Somers' D .519. This value was met or exceeded by only three objectives: student writes M/m melodic seconds (.824), student identifies the starting pitch of a song (.586), and student identifies the longest/shortest of four easy rhythms (.522). The mean Somers' D value for individual objectives was .370. While .370 (average for individual objectives) to .519 (validity overall) is not a sta-

---

[31]It might puzzle some that the three objectives with the highest validity are the ones with the fewest students tested. The objection might be raised, for example, that if writing M/m melodic seconds is so much more reliable than anything else, might one conclude that this is all that is really needed in such a test? The reason so few students were given these particular questions is because of the unique strategy employed by the ExSPRT system. These problems were difficult ones: used only to confirm a high level of competence in instances where prior queries had determined that the student was probably a master. For example, the 1990 pilot study showed that only 27% of those who passed the first semester of theory could write M/m melodic seconds before having had instruction (and a mere 8% of those who failed the first semester could do this). So, what would happen if this were the only objective of the test? Answer: a very low percentage of students would demonstrate mastery of the objective, and one could be quite confident that they were prepared for music theory. But that is not the object of the assessment. The purpose of RON was to identify students NOT prepared for theory. And the 1990 research showed that 73% of even the PREPARED students had not mastered this particular objective before having had instruction. Accordingly, non-mastery of this objective is of very little use in determining which students are unprepared for the sequence. Thus the expert system asks this question only of persons it suspects, from prior questioning, to be masters—hence the low numbers.

tistically rigorous comparison, it does support what some might consider to be obvious—that the cumulative effect of testing across many objectives increases the validity of the system.[32]

Throughout this discussion it has been noted that RON occasionally found it impossible to assess a student as having mastered, or not mastered, one or more objectives. In such instances the student was evaluated as "inconclusive" as pertaining to the objective in question (see fn. 4 & 5 *supra*). While RON currently disregards inconclusives—as if the student had never attempted that type of problem—this research enabled the asking of the question: what, if anything, was measured when RON assessed a student's competence with respect to individual learning objectives, as inconclusive?

It had been assumed, in the creation of the ExSPRT system, that inconclusives were actually masters, or non-masters, but that RON had collected insufficient information to make such a distinction. This research suggests, however, that inconclusive assessments of individual objectives are indicative of a level of competence somewhere between mastery and non-mastery. Mean scores of students testing inconclusive at any given objective bear this out. The system may be improved by factoring inconclusive assessments of individual objectives as somewhere between mastery and non-mastery.

Two final observation are in order. First, the data that makes RON "work," as well as data for this study showing that it worked, were obtained from institutions that attract students of average preparation for university-level music theory.[33] With the exception of the handful of volunteers from Indiana University (in the 1990

---

[32]The comparison is not statistically rigorous because certain objectives were called upon with less frequency than others. For example: 211 individuals were required to write short diatonic melodies (Somers' D of .249), while only 35 were required to write M/m melodic seconds (Somers' D of .824).

[33]Data from the 1990 pilot study were obtained from 141 volunteer freshmen at Ball State University, Indiana University, Biola University, Taylor University, and the University of Oregon. Data for this study in predictive validity were obtained from students at Ball State University. Data continue to be gathered and analyzed from students (since 1995) at Northern Arizona University.

pilot study), the average RON testee might be described as below that of, shall we say, most nationally ranked conservatories and schools of music. In the decade over which RON has evolved, it is likely that this particular clientele has been "hard wired" into the system. In the last eight years the author has directly administered RON, or supervised its administration, to more than 1,300 students at Ball State University and Northern Arizona University. RON has rank ordered these students within standard deviation limits of the normal curve.

It is certain that if the same instrument were used at institutions where the average student is more (or less) prepared, the curve-linear profile of the instrument would be quite different. It is conceivable that this alteration could render the utility of RON, in some instances, ineffective—all students are ranked 1 or 2, for example. This is not to invalidate, however, the testing approach. With some adjustments to its specifications, objectives, and cutoffs, RON could very well be adapted to serve institutions of different size, mission, and student preparedness. Research at unlike institutions would be required, however, for any adjustments to be well-informed.

Second, the notion that some students, who were assessed as unprepared, actually passed the course—two with an A and twelve with a B—might be a little worrisome for those who are not used to these types of assessments. The converse might be easier to comprehend. Of the 140 students whom RON assessed as prepared to begin the music theory sequence, thirty-six did not pass (22 earning D's, and 14 earning F's).

In coming to terms with anomalies, it is important to remind ourselves that students fail courses for reasons other than lack of pre-skills, which explains why no assessment can be expected to be completely reliable. Lack of discipline, or money, or reasons of emotional or physical health, learning disabilities—factors entirely ignored by RON—would explain most of the inconsistencies. Too, RON had no mechanism for detecting that extra ounce of intelligence, that dogged determination, abnormal curiosity, or unusual self-discipline, that might have compensated in instances where it had identified students with insufficient pre-skills (which might explain, though not entirely excuse, those two A's and twelve B's of the preceding paragraph).

Ultimately, we must ask, are the seventy-five individuals whom RON identified as deficient (and who subsequently failed their first music theory course) worth the misidentification of fourteen who proved RON "wrong" by earning A's and B's in the first semester of music theory? The answer to that question is most likely "yes." In view of the fact that many students advised to remediate followed that advice, and passed their first music theory course, the proportion of students whom RON spared "theoretical trauma" (not to mention waste of time and money) outweighs its relatively infrequent misdiagnoses.

This research is significant to music theory pedagogy in five ways. First, it establishes the validity of the intelligent design concept for alerting advisors to students who may not function well in a music theory course unless they receive special attention. Since RON began as an experiment in a radically different concept in assessment, this outcome is very good. With continued work, the prediction error should be reduced and the predictive validity of this instrument, and others like it, improved. Second, this research contributes to our understanding of pre-skills that are more (or less) indicative of success in the music theory sequence. Third, it adds to the growing body of evidence that the rudiments of music theory are hierarchical and that "magic bullet" assessments—where success at higher levels justifies skipping the lower—may be efficient indicators of competence. Fourth, it helps us to know where cutoffs should be made in the advising process. Specifically, we know from this research that students who mix successes with failures while attempting any given objective (especially where that "style" of answering persists over three or four objectives) are probably not prepared.[34] Fifth, while this particular assessment was designed to identify students as unprepared for beginning music theory and ear-training, perhaps of greater importance is the efficacy of computer-adaptive assessments over a wide array of learning objectives throughout the music theory curriculum. Could, for example, an ExSPRT system identify students who might be allowed to skip

[34]This type of response characterized the performance of the "out-of-timers"—those whom RON assessed as inconclusive overall and who were, for the purposes of this research, ranked 4.

lower division courses? This research suggests that, where the purpose is to achieve an accurate assessment in as short a time as possible, and where learning objectives are highly hierarchical (as in music theory) it is possible to design computer-administered assessments that are capable of predicting outcomes with a high degree of validity.[35]

[35]The author thanks Dr. Roy T. St. Laurent, statistics instructor and consultant at Northern Arizona University, for his help in designing and executing this research. Also of invaluable assistance were Dr. Graydon W. Bell (NAU), and Dr. Patricia Sink (the University of North Carolina, Greensboro), who provided most constructive criticisms pertaining to the final report.

APPENDIX I

Somers' D values Per learning objective as predictor of term grade and vice versa.

(ASE in parenthesis)

Objective	Somers' D: Objective Predicts Term Grade — Pass /Fail	Objective Predicts Term Grade — All Gdes	Term Grade Predicts Objective — Pass /Fail	Term Grade Predicts Objective — All Gdes	Students Tested
completes a measure with missing durations	.319 (.100)	.332 (.112)	.297 (.095)	.201 (.069)	95
identifies last pitch of a phrase in C Major	-.164 (.235)	.181 (.235)	-.073 (.109)	.052 (.070)	33
identifies correctly paired notes/rests			See note #1		
identifies longest/shortest of four easy rhythms	.362 (.215)	.522 (.173)	.253 (.163)	.217 (.092)	29
identifies non-equivalent rhythms	.282 (.145)	.377 (.157)	.248 (.130)	.216 (.093)	51
identifies starting pitch of a song	.507 (.126)	.586 (.132)	.507 (.126)	.363 (.084)	41
identifies the root of a triad	.111 (.116)	.184 (.125)	.080 (.084)	.085 (.059)	103
knows number of sharps and flats in key signatures	.281 (.096)	.302 (.105)	.229 (.080)	.156 (.056)	127
matches sounding triad (M m + d)with written			See note #2		
names basic symbols of notation			See note #3		
names pitches in treble alto and bass clefs	.189 (.123)	.160 (.138)	.189 (.123)	.105 (.090)	64
plays pitches of treble and bass clefs on keyboard	.201 (.075)	.304 (.081)	.196 (.073)	.183 (.049)	180
understands the function of sharps and flats	.360 (.111)	.446 (.115)	.357 (.110)	.289 (.075)	71
writes enharmonic equivalents	.272 (.086)	.358 (.103)	.173 (.058)	.142 (.044)	196
hears and writes M/m melodic seconds	.706 (.078)	.824 (.065)	.091 (.087)	.063 (.060)	35
writes pitches in treble alto and bass clefs	.083 (.083)	.151 (.093)	.078 (.079)	.088 (.054)	154
writes sharps/flats of key sig. (correct order and pos.)	.302 (.105)	.380 (.113)	.300 (.104)	.240 (.071)	83
writes short diatonic melody (dictated)	.197 (.109)	.249 (.126)	.074 (.043)	.059 (.031)	211

(notes)

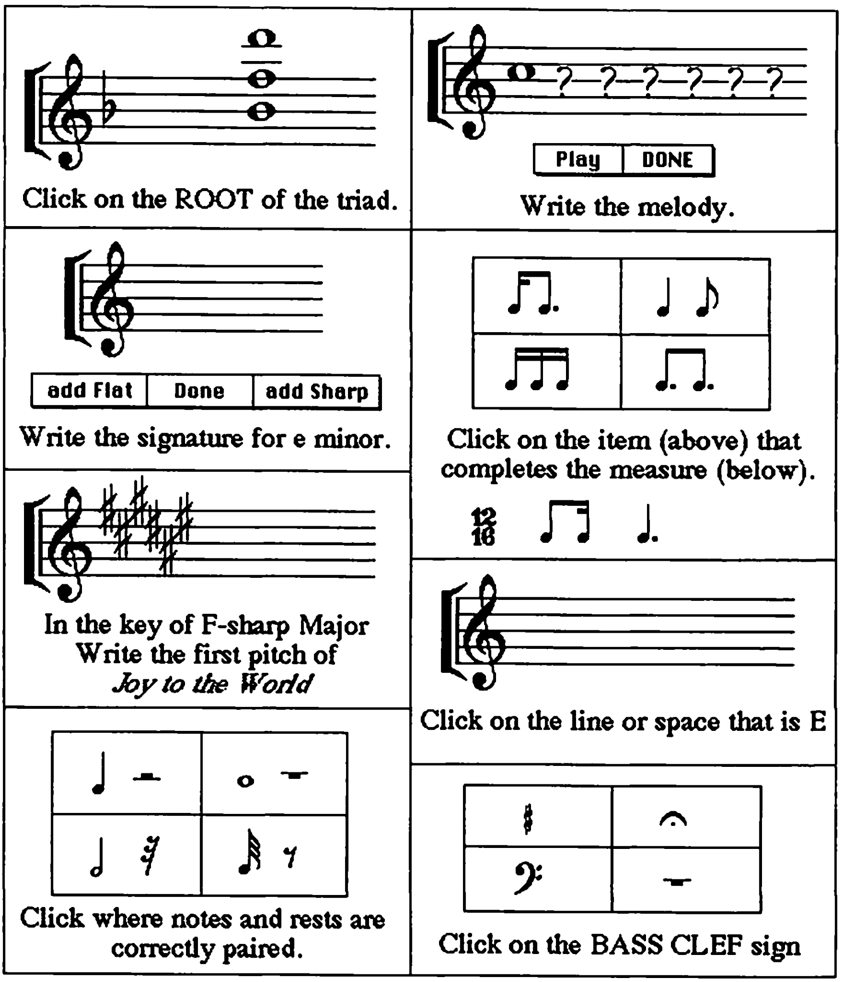
1) It was not possible to compute statistical values for this objective as most individuals fell within the same cell. Of the 39 individuals tested all but one were inconclusive. The remaining person tested as not having mastered it.

2) Not possible to compute statistical values: of the 48 individuals tested, all but two tested as masters. The remaining two tested as inconclusive.

3) Not possible to compute statistical values: all 44 individuals tested in this objective were inconclusive.

APPENDIX II

Stimuli for Selected Objectives



Teaching Harmonic Rhythm

Joseph P. Swain

The Problem

Harmonic rhythm is one of those essential aesthetic aspects of traditional western harmony that begs for attention from scholars, demands to be taught to students, and yet frustrates teachers who attempt to explain it. The concept seems simple and a few well-chosen examples can demonstrate its obvious importance, but analytical technique for harmonic rhythm has never had sufficient precision and consistency to be of much practical value since Walter Piston first defined the term in 1944.[1] We have had precious little theory for it, and so no advice, no specifics such as those for voice-leading, that we can give to our students. In the following pages I would like first to explain how a new analytical technique for harmonic rhythm called dimensional analysis can address these problems and second to describe my experience using this technique to teach harmonic rhythm in the second semester of a traditional harmony course.[2] Dimensional analysis, the students agreed, provided a precise approach to harmonic rhythm that they could easily grasp, and at the same time amplified other fundamental aspects of harmonic practice by integrating them with rhythmic effects.[3]

---

[1]*Harvard Dictionary of Music* (Cambridge, Massachusetts: Harvard University Press), p. 319. For a good survey of theoretical treatments of harmonic rhythm, a meager literature indeed, see Mary Irene Arlin, "Harmonic Rhythm in Selected Fugues from The Well-Tempered Clavier, Book I." M. M. Thesis. Indiana University, September 1965.

[2]For fuller theoretical discussion of this technique, see Joseph P. Swain, "Dimensions of Harmonic Rhythm" *Music Theory Spectrum* 20.1 (Spring 1998): 48-71.

[3]The citation of student reactions to dimensional analysis comes from a survey administered at the conclusion of the course, following the final

As it developed through the editions of his textbook *Harmony*, Walter Piston's concept of harmonic rhythm focused on the durations of chord roots. The first problem is that it is not always clear what is to count as a change of root and the beginning of a new duration. Should the third beat of Example 1, for instance, be tallied as two eighth-note durations (E diminished, B-flat major), or one quarter-note duration that includes a non-harmonic tone? If the latter, which triad? The same dilemma recurs in measure three, and in countless other instances of traditional western music.

Example 1: Handel, "Thou art gone up on high, (No. 36) mm. 1-5, from *Messiah*.

---

exam. All responses were anonymous. The questions were: 1. Do you believe that the dimensional analysis increased your understanding of traditional harmony? If so, how? 2. Do you think that aural skills (ear training) tied into the graphing process in any way? If so, how? 3. Do you believe that the class presentations by students were worthwhile? Did they increase your understanding of harmonic rhythm and *Messiah* as the semester progressed? 4. Did you feel adequately prepared to make the graph with the materials given to you and the two preparatory classes? If not, what would have helped you to be better prepared? 5. In making the graph, which dimensions seemed easy to develop, if any, and which were difficult, if any? 6. About how many hours did it require to make the graph?

Besides voice-leading ambiguity, should not changes in other harmonic aspects affect perceptions of rhythm? Is there not a perceptual change, therefore a rhythmic event, of harmonic character when the G minor triad moves from root position to first inversion in measure three? Or even more so when a chord function changes, as when the leading tone is canceled in measure five, negating its dominant tension, again without any new root? The point is quite simple: a chord has many facets; therefore harmonic rhythm is likewise multi-faceted. Representing rhythmic changes on one dimension only—the root—must oversimplify what is a complex musical perception.

The second problem with Piston's concept is that, even if harmonic rhythm were limited to root changes and even if that were a simple matter, there is no theory about how it informs the dynamics of a harmonic phrase, its tensions and resolutions. That is why it is easy to show students what harmonic rhythm is, but difficult to show them how to apply it either in analysis or composition.

And that is why Piston's own treatment of the issue is limited to a collection of examples showing only how varied harmonic rhythm can be, and why its application in analysis and criticism has been limited to comparisons. Now, comparisons can indeed inform and enlighten us, as they do in Example 2. Here, it is important to show students that even though the basses in the first quotation from "For unto us" sing a run of sixteenth-notes, the sense of pacing is faster in the second quotation because the harmonic progression moves in quarter-notes rather than half-notes, a sense that contributes mightily to the climax at "Wonderful, counsellor . . .". But from this we can only conclude that the function of harmonic rhythm is to lend one phrase some certain monolithic sense of speed which may then be contrasted with a sufficiently faster or slower sense in another phrase. A teacher can ask students to experiment with this idea; I assign contrasting arrangements of the same tune. But any subtler application of harmonic rhythm seems *ad hoc*, without general principles.

Example 2: Handel, "For unto us," (No. 12) mm. 20-22, 26-28 from *Messiah.*

A Solution

Dimensional analysis attacks the problem of multiple harmonic facets and effects simply by making graphs of each aspect—each "dimension"— and aligning them underneath the passage (see Example 3). This allows the retention of more than one interpretation, more than one harmonic aspect, of a single event. The resulting dimensional analysis can then address the second of the problems outlined above: what can harmonic rhythm tell us about the dynamic tensions and resolutions within a single passage? The following principles guide the interpretation: that rhythmic tension of one kind rises with increase in speed of changes (speed hypothesis); that rhythmic tension of another kind rises with increasing divergence—

Example 3: Dimensional analysis of Handel, ritornello (mm. 1-11) from "Thou art gone up on high" (No. 36) from *Messiah*.

independence of motions—among the dimensions (independence hypothesis).[4]

Both of these commonsense hypotheses above derive from the simple fact that at the end of a composition, when tension is resolved, the motion in all dimensions ceases and therefore convergence among them is at a maximum.

Using the Technique in Class

Each student was assigned to make a dimensional graph of a single ritornello from Handel's *Messiah* and to give a five-minute oral presentation of the results, which should point out particular problems in making analytical decisions and then some critical comment about Handel's harmonic rhythm in the excerpt. (Baroque ritornellos are especially appropriate for applying this technique because they are in some sense complete, miniature compositions where harmonic rhythm plays a decisive role in their articulation.) I prepared the students over two fifty-minute class periods by elaborating on the following instructions:

Brief Instructions for Making A Dimensional Graph of Harmonic Rhythm

Step 1. Transcription. Make a transcription of the ritornello, showing all contrapuntal voices. One system per page.

Step 2. Rhythm of the texture. This graph shows all the rhythmic attack points regardless of where they occur in the texture.

---

[4]In his *Structural Functions of Music* (Englewood Cliffs, N. J., 1976), Wallace Berry, whose principal concern is the description of the course of "intensity" in music, essentially makes the same assumptions. See pp. 11, 86.

These principles are presented here in simplified formulation. They are refined significantly in Chaps. 9 and 10 of my *Harmonic Rhythm: Analysis and Interpretation* (New York: Oxford Univerisity Press, 2002). There human attention capacities account for various tensions of harmonic rhythm and the "coloring" of the musical texture that occurs when we attend to one focal dimension more than the others. The assumption that musical tension has differences in kind, as well as amounts, is developed at length in Joseph P. Swain, "The Concept of Musical Syntax," *The Musical Quarterly* 79.2 (1995):281-308; and in chap. 2 of Swain, *Musical Languages* (New York: Norton, 1997).

Step 3. Phenomenal harmonic rhythm shows the duration of every new vertical combination of pitches. An octave shift counts as a new pitch. The simple repetition of a chord (without changing any pitches in any voice) counts as no change. Before beginning to graph, inspect the music to see if all phenomenal changes might coincide with new attack points (textural rhythm changes). Often in Baroque music the two graphs are identical.

Step. 4. Bass pitch rhythm. Graph the rhythm of the bass part, adding together repeated notes. Octave shifts are new notes. Do not omit rests.

Step 5. Root/quality rhythm (level 1). Graph the duration of each harmonic root/quality change. For this first level, graph every reasonably audible change of root or quality. Below the rhythm graph, write in the letter pitch name for the root and symbol for quality (upper case = major triad (add + for augmented), lower case = minor (add - for diminished).

Step 6. Density. Graph the number of contrapuntal voices that create each new root. As long as a voice changes its note, it counts toward the density, even if that pitch was a factor of the previous chord. Octave shifts are new pitches.

Step 7. Root/quality rhythm (level 2). Graph the duration of each higher-level root/quality change.

Step 8. Functional harmonic rhythm. Use the first level Root/quality graph as a starting point. Decide on the prevailing key and each chord's function within that key. Declare the key in effect on a given level. Graph the duration of each function and label each duration below (I = tonic function; IV = subdominant; V = dominant). Secondary dominants establish a level embedded within the main key. The function of the secondary dominant area in the principal key depends on the function of the object chord. There may be two, or even three levels of harmonic functions.[5]

---

[5]This is the verbatim instruction given students in Spring 1998. Since then these rules have been refined. See *Harmonic Rhythm*, chaps. 2-7.

Despite the brevity of these preparations, students reported almost no difficulty in generating the graphs. The average time spent was about four hours, and understandably, more care was needed with the more abstract, higher-level graphs whose values could not simply be transcribed from the score.

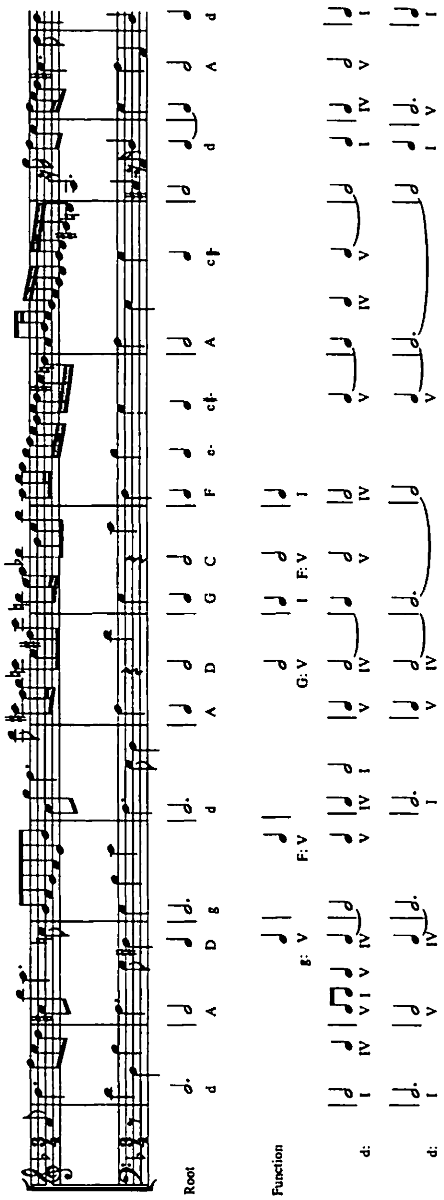
Analysis of the More Abstract Dimensions

Indeed, a good deal of class discussion arose from problems and inconsistencies in the students' graphs of the more abstract dimensions. To get them started, I reminded them of two familiar principles fundamental to traditional harmony. One is the principle of voice-leading, in particular that chords formed in conjunction with step motion can also be considered as purely melodic elaborations of a longer-lasting structural chord. The second is the secondary dominant principle, that it is possible to embed a dominant function chord borrowed from a foreign key whose object of resolution is a functioning member of the main key of the phrase.

How well did they do? Example 4 shows one student's analysis of the higher-level root/quality and function dimensions of "Thou art gone up on high." In the root dimension many of the differences, compared with my own analysis (Example 3) are reasonable alternatives to hearing the voice leading, such as in m. 1. Others are less justifiable (mm. 5-6, 10). Most of the problems arose in the function graphs, particularly in connection with embedded functions. Many students, including this one, neglected to retain a single function on the higher level for the entire length of the embedded key (mm. 3-4, 5-6). Another common error was a too slavish translation of the roots into functions without taking into account the context (example: a D minor chord in this ritornello always coming out as a tonic function even when it was inverted and led to a strong dominant.) Occasionally, as in this student's case, an entire functional level was graphed when none seemed required.

Nevertheless, the benefit of dealing in concrete fashion with the whole concept of different levels of harmonic syntax far outweighed the cost of errors made in students' first attempts. "The most interesting thing about traditional harmony that this analysis has taught me must certainly be the interplay between higher and lower level

Example 4: Student analysis of ritornello of Example 3, showing higher-level root and function dimensions only.



function and how they affect the overall nature of the piece," wrote one student on his survey and his reaction, while seldom expressed so succinctly, was echoed on many other responses. One reason why this notion of harmonic abstraction went over so convincingly, I believe, is that students could put flesh on it without having to learn and accept any radically new analytical concept or jargon. Another reason is the sheer perceptibility of the graphs. Many students commented that the higher-level analyses were the ones that connected most usefully with their aural training: "The better you can hear the music (chord changes) the better off you are in making root and function graphs." Multi-level analysis, a concept more common in advanced theory courses, can be introduced through the back door, as it were, since dimensional technique depends only on the concepts and terminology that they have mastered in the course of traditional harmony. At the same time, the interpretative results affirm and give value to those concepts.

And it is perhaps not too early to expose undergraduates to the war between bottom-up and top-down perceptual perspectives, the general role of musical context, one of the most truculent issues in analysis. Although it may appear at first glance that the dimensional graph develops logically from the "data" of the textural and phenomenal analyses in the step-by-step procedure given above, I advised the students to be open to revising analyses because of how they hear the higher levels. In Example 3, calling the third beat of the second measure a D major triad when no D pitch is sounded depends on the sense of the chord in the D minor context and the instant recognition of the F-sharp as a local leading tone moving to a G minor triad immediately.[6] Similarly, the decision to characterize the entire fourth measure as a single D minor chord in the higher-level root dimension depends on the perception of the overall key of D minor outweighing the local function in F major. Yet, the arrival of tonic *function* waits until the third beat because of the melodic motion in the bass. Of course, contextual issues come up regularly in traditional harmonic analysis, too; what dimensional

[6]The choice between the named D triad and the alternative, F-sharp diminished, is academic here since both would have the same function indicated below. In harmonic rhythm, the crucial fact is that the chord changes; sometimes it doesn't matter exactly what it changes into.

graphing seems to encourage is more precise articulation of reasons for making analytical decisions. "I had previously understood that these issues existed and were utilized, however I did not have the vocabulary to quantify them."

Interpretation

The students' oral presentations tended to be long on technique and short on critical interpretation. Despite the simplicity of the interpretative hypotheses, only a few took that step back to view the graph as a whole. This suggests that more emphasis on reading "vertically" as well as horizontally is probably required during class discussions. Nevertheless, the few students who attempted interpretation applied the hypotheses fairly well. The independence hypothesis supplies particularly clear results in the conclusions of ritornello after ritornello, as the various dimensions, after having proceeded at varying speeds up to the last cadence, suddenly converge in their motion (see Example 3). In any case a teacher can use even rather innocuous student observations to begin the exploration and discussion of sophisticated critical points.

Example 3 can tell us something about the subtlety of Baroque phrasing in Handel's ritornello. The downbeat of m. 4 suddenly brings a convergence of half-note durations in four dimensions (two root and two function levels). This lowering of a kind of textural tension is confirmed immediately by a lowering of harmonic tension: the weak tonic function on the third beat, just enough to make the phrase barely articulate without undermining the essential Baroque motor rhythm that continues unabated in the texture/phenomenal graphs. The feature that catches the eye (and, we hope, the ear) is the prolonged subdominant function that subsides into the weak tonic. Thus the phrase is articulate in one sense but incomplete in another, since it lacks the complete syntax of subdominant-dominant-tonic. The next phrase, much longer, resumes that prolonged subdominant, but this time progresses to a prolonged dominant. Its resolution on the weak third beat recalls m. 4, but then confirms that resolution decisively with a much faster, complete syntax in which nearly all the dimensions converge in speed for the first time.

One student wrote, "The dimensional analyses makes the concept of musical tension much clearer and easier to understand." The control of musical tension in its various forms is a key theoretical issue, and these graphs can help to make its marvelous complexity real for students.

Some kinds of musical tension can be understood by direct inspection of the score. In this ritornello it is obvious that the faster motion is reserved for the end of the second phrase in a well-crafted, composed accelerando which comes to a halt at the end of the prolonged dominant. This feature is reflected in the texture/phenomenal graph. But students learned to read tensions and resolutions in the graphs that are not immediately available from the score: the greater number of longer durations at beginning and end, the convergence of dimensions at the final cadence, and the areas of greatest textural divergence—dimensions moving at different speeds—in the center.

One could ask students to notice how the ritornello, while increasing its tension in various ways, does so fluidly, without sudden dissonances or any other abrupt articulations. Could this be due to the gentle process of acceleration, spread over various dimensions, and perhaps as well to the principal harmonic functions invariably beginning on offbeats?

One could point out, and ask students to verify by hearing, that the two sixteenth-note scales in the violin (mm.7-8), identical save for the octave transposition, have different effects. The first is the locus of greatest tension by pure speed (hypothesis 1 above) while the second has greater tension of textural divergence and prolonged dominant function.

One could even teach that music theory can be logically rigorous at times: an embedded key, as in measures 5-6, must produce a tension of divergent motions if we assume that its harmonic function, deriving from its object, must be prolonged for the length of the embedded key. Thus the quarter-note functional changes throughout the embedded keys of G major and F major translate into a seven-beat higher-level subdominant function because G major and F major themselves have subdominant function in relation to D minor, the home tonic.[7] Thus secondary dominant regions

[7]Subdominant function comes from the fact that G major and F major represent tonicization of triads IV and III in a D minor context.

have a rhythmic effect, in addition to the harmonic, that is precisely quantifiable.

Conclusions

Such are some analytical possibilities arising from the dimensional analysis of a single ritornello. Despite its infancy, at least three pedagogical advantages of the dimensional technique seem compelling enough to try it in undergraduate teaching:

1) Dimensional analysis makes Piston's concept workable in class. Much of its technique is specific and objective, which undergraduates like, so teachers can substantiate the proclaimed importance of harmonic rhythm with directions, advice, and a full-blown method of studying it. Yet the more abstract dimensions, to say nothing of the analysis taken as a whole, allow for interpretations that can make every student graph a unique creation. Because of this, my students felt satisfaction and a sense of achievement in making these analyses of Handel ritornellos, and this was especially effective coming at the end of the year-long course. Harmonic rhythm could not only review all of the harmonic concepts they had learned, but could put them into practice and show them that all their labors could indeed bear fruit.

2) Dimensional analysis forces students to deal with musical texture with a thoroughness rarely found in harmonic theory. Students found density most interesting because it called to their attention the contrapuntal weave that created the rhythms they were studying. Example 3 is not the best demonstration of this, being a two-voice composition, but students who analyzed more fully orchestrated ritornellos found things to say about density patterns, particularly when Handel drops the bass voice altogether as he is wont. Texture is also behind any consideration of convergent rhythms among the dimensions, since it is the handling of the counterpoint that affects the perception of the more abstract levels.

3) Finally, dimensional analysis forces students to engage their own musical experiences of a piece in a multiplicity of ways which, if they were considered heretofore, were probably not considered together, as they might affect one another. The most prominent among these is the realization of harmony as a multi-level phenomenon.

Other pedagogical strategies will certainly be discovered by other teachers who try the method. One that comes to mind from the compositional point of view, rather than the analytical, is to have students analyze their own composed phrases and arrangements. This exercise may make more palpable some very practical issues, such as when to use root position and when to invert, how to choose chords for their function within the phrase, and how to construct the cadence for the desired rhythmic effect. Another possibility is the comparison of harmonic rhythm in different musical languages and styles.

In any event, harmonic rhythm need no longer be one of those principles introduced with appropriate gravity and fervor only to be given short shrift for lack of anything specific to say. Dimensional analysis gives teachers plenty to say and, I expect, all of us plenty to learn.

The Evaluation and Design of an Undergraduate Music Theory Placement Exam

Barbara Murphy

According to the National Association of Schools of Music 1999-2000 Handbook, all music majors must have certain competencies in aural skills and analysis, composition and improvisation, repertory and history, technology, and synthesis of ideas.[1] Generally, students meet these requirements by taking courses in Basic Musicianship which include courses in Theory, Ear-training, and Musicology. These courses comprise 20-35 percent of their curriculum (24-42 hours of a typical 120 semester hour curriculum).[2] Even though NASM states that to be admitted to a baccalaureate degree in music, students should possess a musical background that allows them to "relate musical sound to notation and terminology both quickly and accurately enough to undertake basic musicianship studies,"[3] many students enter college without this prerequisite knowledge. Some students enter music study having had no theory background; their only exposure to music has been in band or choir where little or no explanation of theoretical concepts has taken place. Other students may be able to read multiple clefs and play certain scales and intervals but may not know what they are playing. Still other students have taken theory in school or summer camp and they may be knowledgeable in theoretical concepts through chords or even part-writing.

To evaluate entering students, most schools of music give placement exams on rudimentary music topics. Because existing stan-

[1]National Association of Schools of Music. 1999-2000 Handbook. (Reston, VA: National Association of Schools of Music, 1999), 78-79.

[2]Ibid., 81-91.

[3]Ibid., 72.

dardized exams do not measure the material covered in undergraduate music theory courses, most schools create in-house, or criterion-referenced, exams. Although these exams have content validity, very few evaluations have been conducted on their results to determine the reliability and predictive validity of the exams.

The purpose of this article is to discuss the steps taken to evaluate the Theory Placement Exam used at the University of Tennessee, Knoxville and the steps taken to create a new Music Theory Placement Exam that could be used to place incoming music majors and minors into Theory or Fundamentals of Music classes based on their knowledge and background in theory. The new exam consists of items in a SuperCard stack and is administered via the computer. This article describes an evaluation procedure that determined the extent to which the exam appropriately and consistently placed students in a music theory class; that is, the procedure used to determine the predictive validity and reliability of the Undergraduate Music Theory Placement Exam.

Related Research

In recent years, there has been much development of computer assisted instructional programs in music and in theory in particular. However, only a couple of placement programs have been written and evaluated. These programs are described below.

James P. Colman's Program

In 1990, James Peter Colman reported on his development and validation of a theory placement exam for use at Michigan State University.[4] His goals were (1) "to create a diagnostic test capable of measuring the current music theory achievement of incoming students so that statistical data could replace the assumptions made by college theory professors," (2) "to determine whether the newly created test could function as a predictive variable in the evalua-

---

[4]James Peter Colman, "The Development and Validation of a Computerized Diagnostic Test for the Prediction of Success in the First-Year Music Theory Sequence by Incoming Freshman at Michigan State University." (unpublished Ph.D. Dissertation, Michigan State University, 1990).

tion of future success in music theory," and (3) "to provide a musically specific test that would give advisors of college music majors another aid for proper advising discussions."[5]

He developed the test by defining content areas, clarifying the behavioral objectives, and creating 90 test items which he then submitted to several teachers of theory (at MSU) for clarification and measurement of the objectives.[6] The test included items on aural skills as well as written theory skills. Topics included scales (27% weight), pitch notation (20%), notes and rests (13%), intervals (13%), triads (13%), key signatures (8%) and time notation (6%).[7] The test was delivered on Macintosh computers via HyperCard. It was presented in a linear fashion, although students were able to pass on any item and come back to it at the end of the test. The program created an answer sheet for each student with the student's score, as well as a file of the results for statistical analysis. Fifty-nine incoming freshmen took the 50-minute test during registration. The mean score was 56.75 with a standard deviation of 11.97. Colman used a test-retest method for assessing reliability resulting in a reliability estimate of .86.[8]

In addition to the results of the placement test, Colman looked at the grades the students received in the three quarters of theory classes that year. The sample size, therefore, dwindled from 59 to 32 students who completed the entire year of theory classes. From the results, Colman noted that the students' grades in the theory classes improved over the three terms, a fact which he attributed to practice by the students or to attrition. He found that students' grades on the test and their term grades correlated well, with a decline in the correlation in Spring term. The lowest correlation was between the test and the aural skills class grade. This low correlation could be the result of the low number (10%) of test items dealing with aural skills. Colman noted that his most striking result was "the discovery that the theory test correlated more strongly with the less sensitive grade point scale than it did with percentage grade scale."[9]

[5]Ibid., 1-2.
[6]Ibid., 41.
[7]Ibid., 40.
[8]Ibid., 53.
[9]Ibid., 67.

The final result was that, although the items on the test predicted course grades fairly well, some test items needed to be revised or discarded. However, this step was apparently never taken.

Timothy Smith's Program

In 1994, Tim Smith reported on his development of an adaptive computer test, called Ready or Not (RON), that the School of Music at Ball State University is now using as part of its "enrollment management program."[10] Smith defines an adaptive computer test as one that "could omit items that are too hard or too easy, compare student performance with statistical norms, and generate problems algorithmically."[11] This type of test "interacts with the user in a nonlinear fashion based upon responses as they are integrated with, and interpreted by, data."[12]

The goal of such a test is "more efficient and accurate identification of students at risk," where "efficiency is a measure of the time needed to accomplish that goal and accuracy is a measure of the validity of the assessment."[13] Smith's idea was that, if students show a high level of mastery, they should not be required to take test items lower in the hierarchy of difficulty, and so their test would be shorter than other students. Therefore, he decided to design his test "beginning with items of high discrimination [so that] the system would more quickly detect a trend toward mastery or non-mastery. After a trend had been established, substitution of criterion-referenced items would more efficiently and more accurately diagnose competence."[14]

Smith borrowed an equation from Bayesian mathematics called the Sequential Probabilities Ratio (SPR) to develop his program. By using this equation, the computer could be set to recalculate the probability ratio after each item and determine whether the student is a master or non-master of the topic in question.

---

[10]Timothy A. Smith, "An ExSPRT System Approach to the Assessment of Students Needing Remediation in Music Theory," *Journal of Music Theory Pedagogy* 8 (1994):179-200.

[11]Ibid., 180.

[12]Ibid., 186.

[13]Ibid., 189.

[14]Ibid., 187.

The resulting computer program tested twenty-two learning objectives, both written and aural. It began by posing items with high discrimination values and proceeded to items with lower discrimination values. These values were determined by analyzing items on a test previously given to 141 students at various universities. The computer program "begins by selecting one of the five most discriminating objectives and generating random problems within that objective. As each student demonstrates mastery or non-mastery of the objective, RON advances to one of the top five remaining objectives."[15]

At first, the students completed the test very rapidly, sometimes in less than ten minutes, and after being tested on only two or three objectives. In the second year, RON was told to withhold its assessment until the student completed at least five objectives.[16] But still RON often detected a need for remediation within fifteen minutes. One aspect of RON not discussed much in this article is the makeup of the test items themselves. Smith talks of the objectives and their discrimination values, but, unlike Colman who discusses the items at great length, Smith discusses them very little.

The Current Study

The aim of this project was to create a exam that would assess student achievement and place students into the proper class as well as keep data for analysis. This exam would not adjust to student responses like Smith's. The aim here was similar to Colman's study with the exception that this research would include a revision of the exam and an analysis of this revision.

The exams

The placement exam used at the beginning of this study, subsequently referred to as Placement 96, was a revision of an exam that had been used since 1993, subsequently referred to as Placement 93. Both Place-

---

[15]Ibid., 193.
[16]Ibid., 194.

ment 93 and Placement 96 covered written theoretical topics only; no ear training topics were included on the exams.[17]

Placement 93 consisted of twenty items. There were two notation items; for these items, the students were asked to name the lines and spaces of treble and bass clefs by writing the name of notes under the notes on the staff. For the one rhythm item, the students were to place bar lines in an example of 4/4 meter. Four items on scales followed; for each, students identified the type of scale written (major, natural, harmonic or melodic minor). Seven items on intervals were included, for which students identified the quality of the written interval. (For all but one item, all three responses used the same interval size.) On each of the last six items, students were to identify the quality of a triad presented in both treble and bass clefs.

Placement 96 consisted of twenty-five items, five each in the categories of notation, rhythm, scales, intervals and triads. All items were multiple choice questions with four possible answers. The items on Placement 96 were much more heterogeneous; the items were designed to test many types of knowledge in each category. The five notation items included one item testing knowledge of anchor notes (i.e., Which note is middle C?), two items testing the students' ability to identify pitches in treble and bass clef, one item on notation (i.e., Which is the correct notation for a Eb eighth-note?), and one item on the meaning of accidentals. The items on rhythm included two on note values (i.e., a given note is equal to how many beats or what combination of notes), one item on the number of beats in a given time signature, and two items for which students completed a measure with a note or rest. The items on scales and key signatures included two items on scale identification, two items on key signature identification and one item on parallel and relative major and minor scales. The interval items included one item

[17]At the University of Tennessee, like many other schools, ear training and written theory are two different classes. Most of the ear training at UT is delivered and tested via computer using a program called the Curriculum for Aural Training (C.A.T.) written in-house by Don Pederson and Mark Boling. Since these programs have their own placement exam which has been used successfully for several years, the placement exams used in this study covered only written theory topics.

on interval quality, one on interval size, one item on both quality and size, and two items for which the student chose which interval was the one specified. The five triad items included four on triad quality identification and one item on inversions.

Evaluation of Placement exams

Before any revision could be undertaken, the exams were evaluated to determine how they were performing as indicators of student knowledge and student success in theory classes. In addition, each exam item was evaluated to determine if it was separating those who knew the information from those who didn't (i.e., an item analysis). For these analyses, a macro entitled "Item Analysis for Multiple Choice Tests" from the SAS Sample Library was used. The macro computes descriptive statistics for analyses of data from multiple-choice exams. The macro also provides an item analysis. For the analyses, each observation (i.e., line of data) contains the answers of one Subject (S) to the exam items. The descriptive statistics resulting from the analysis on Placement 93 are shown in Table 1.

The statistics of most importance here are the mean score, the standard deviation, the KuderRichardson 20 (KR-20), and the Spearman-Brown Prophecy number. The mean score is the average score. A standard deviation is ". . . related to the distances . . . of scores . . . from their mean. . . . It is a measure of variability or dispersion such that a distance from one standard deviation below

Table I: Descriptive Statistics for Freshman Placement Exam 1993

Number of Items	20
Number of Subjects	115
Mean score	11.383
Variance of Scores	26.537
Standard Deviation	5.151
Kuder-Richardson 20	0.880
Standard error (from KR20)	1.783
Spearman-Brown Prophecy (from KR-21): To obtain a reliability of 90 the test should contain this many items	31

31

the mean to one standard deviation above the mean includes about 2/3 of the cases in most distributions of scores."[18]

Kurtz and Mayo explain the Spearman-Brown Prophecy formula as one that depicts the relationship that the "correlation between scores on one test and an alternate form of it will then be high."[19] Further, they explain that in deriving the formula, "it is assumed additional forms of the test exist or could exist that (1) all have equal standard deviations and that (2) all correlate equally with each other. In practice, it seems not to matter much whether or not these assumptions hold."[20] In addition, the formula can be used, as it was in this study, to determine how much an exam would need to be lengthened to achieve a specified reliability.[21]

To explain the Kuder-Richardson 20 formula, Kurtz and Mayo state that "Some time ago, a common procedure for determining the reliability coefficient was to split a test into comparable halves, to correlate scores on these halves, and then by the Spearman-Brown Prophecy formula . . . to estimate what the correlation would be between whole tests rather than between half tests. It was recognized that different ways of splitting a test into halves gave different results, but nobody did anything about it until Kuder and Richardson independently saw how to solve the problem. . . ."[22]

The Kuder-Richardson 20 formula solves the problem by making the following assumptions: " . . . Each [test] item measures exactly the same composite of factors as does every other item."[23] Secondly, it assumes that " . . . all inter-item correlations are equal."[24] Lastly, it assumes that " . . . all items have the same squared standard deviation or variance. . . ."[25] The resulting formula, the KR-20, "gives a good approximation to the values computed by the more rigorous formulas [the K-R (8) and K-R (14)]. . . ."[26]

[18]Albert K. Kurtz and Samuel T. Mayo. *Statistical Methods in Education and Psychology*. (New York: Springer-Verlag, 1979), 47.

[19]Ibid., 271.

[20]Ibid.

[21]Ibid., 272.

[22]Ibid., 265-266.

[23]Ibid., 267.

[24]Ibid., 268.

[25]Ibid.

[26]Ibid.

As shown in Table 1, 115 Ss took Placement 93 and averaged 11.383 items correct (56.9%) with a standard deviation of 5.151. The Kuder-Richardson 20 (KR-20) result was 0.880. The Spearman-Brown Prophecy result indicates that to obtain reliable results at the 0.90 level of reliability the test should contain at least 31 items instead of the 20 items on the exam.

The same analyses (i.e., descriptive statistics and item analysis) were also conducted on the subtests for notation, scales, interval quality and triad quality. The analysis was not run on the subtest for rhythm, since this subtest consisted of only one item. The descriptive statistics from this analysis are shown in Table 2.

Table 2: Descriptive Statistics for the Subtests of Freshman Placement Exam 1993

Subtest	1-Notation	2-Scales	3-Intervals	4-Triads
Number of Items	2	4	7	6
Number of Subjects	115	115	115	115
Mean score	1.8	1.983	3.765	2.965
Variance of Scores	0.214	1.859	4.479	4.385
Standard Deviation	0.463	1.364	2.116	2.094
Kuder-Richardson 20	0.383	0.644	0.74	0.80
Standard Error (from KR20)	0.363	0.813	1.079	0.936

The fact that the KR-20 result (0.80) is higher on Subtest 4-Triads than all the other subtests may be the result of the homogeneous nature of the relatively large number of items in this group (6). All items on Subtest 4-Triads dealt with triad quality and all items had the same four answer choices. Likewise, the low KR-20 result of Subtest 1-Notation (0.383) may be the result of its having only two fairly different items in this category. The conclusion that the homogenous nature of the test items has an effect on the score is supported by Anastasi in her description of Kuder-Richardson reliability as a: "method for finding reliability, . . . utilizing a single administration of a single form, [that] is based on the consistency of subject's responses to all items in the test. This interitem consistency is influenced by two sources of error variance: 1) content sampling (as in alternateform and split-half reliability); and 2) heterogeneity of the behavior domain sample. The more homogenous the domain, the higher the interitem consistency. For example, if one test includes only multiplication items, while another comprises addition, subtraction, multiplication, and division items, the former test will probably show more interitem consistency than the latter. . . . It is apparent that test scores will be less ambiguous when derived from relatively homogeneous test."[27]

The statistics for Subtest 3-Intervals also follow the above assumption that the homogeneous nature of the items results in higher scores. The interval subtest had seven items—the largest number of items in a subtest—and all items tested interval quality; size was the same in each answer in all but one instance. The KR20 result is .74, the second highest, again indicating that the more homogenous the items and answer choices, the more reliable the exam is in indicating student knowledge.

Next, analyses were conducted on the answers from Placement 96. The descriptive statistics for the exam as a whole and the five subtests are shown in Table 3.

Table 3: Descriptive Statistics for Placement 1996

	Whole	Subtest 1 Notation	Subtest 2 Rhythm	Subtest 3 Scales	Subtest 4 Intervals	Subtest 5 Triads
Number of Items	25	5	5	5	5	5
Number of Subjects	90	90	90	90	90	90
Mean score	15.811	3.722	3.833	2.833	2.633	2.789
Variance of Scores	28.874	1.237	1.713	2.096	2.100	2.550
Standard Deviation	5.373	1.112	1.309	1.448	1.449	1.597
Kuder-Richardson 20	0.862	0.471	0.611	0.631	0.582	0.677
Standard Error (from KR20)	1.995	0.809	0.816	0.880	0.937	0.908
Spearman-Brown Prophecy (from KR-21): To obtain a reliability of .90 the test should contain this many	47					

[27]Anne Anastasi. *Psychological Testing*, 3rd ed. (New York: MacMillan Publishing Co, 1968), 84.

Although the mean score on the exam, 63% (15.811 of 25 items correct), is not an encouraging percentage, the exam does seem to be fairly reliable as evidenced by the Kuder-Richardson 20 result (.862). The KR-20 result on this exam is lower than the previous exam (.862 versus. 0.880) but not by much (0.18). It can, therefore, be assumed that both exams are fairly reliable and are testing at the same level. However, the number of items needed on this exam for a reliability level of 0.90 (47) is much higher than the 31 of the previous exam. This is most likely the result of the heterogeneous nature of the items on this exam versus the previous exam. This conclusion holds since on Subtest 5-Triads four of the five items test the same concept and this subtest has the highest KR-20 result (.67). This assumption is again borne out by the KR-20 results on Subtests 3 and 2 (.63 and .61 respectively). In each of these subtests, few different topics were covered; both Subtest 2 and Subtest 3 test only three different topics. The items on Subtest 1-Notation are the most diverse and, consequently, the KR-20 result on this Subtest is the lowest (0.471).

Revision of the Music Theory Placement Exam

After completing the item analyses, Placement 96 was revised. The number of items in each subtest was increased to ten, therefore increasing the number of items on the exam to fifty, following the guidelines presented by the Spearman-Brown prophecy. The statistics for each item of Placement 96 (i.e., the item analysis) were studied to determine which items needed to be changed or replaced. Items that separated the Ss in the top third of the group from those in the bottom third were kept and the other items were dropped or revised. Five similar items were added to each subtest. An attempt was made to make the items in each area more homogeneous. The resulting exam is now known as the UT Music Theory Placement Exam (or Placement 97). The general topic areas remained the same; specific topic areas included on the exam are listed in Table 4. An example item from each specific topic area is shown in Appendix 1.

As can been seen in Table 4, the items in the sections on Intervals and Triads were more homogenous—the items test the same information (i.e., eight items on triad quality) in the same format

Table 4: Topics Covered with Placement 97 exam items

General Topic Area	Specific Topic	Num.of items
Notation	Anchor Notes (middle C; A440)	2
	Clefs	4
	Enharmonic Notes	1
	Aspects of Notation (flags, accidentals placement)	3
Rhythm	Note Values	2
	Note/Rest Values - Simple time	3
	Note/Rest values - Compound time	3
	beats in simple time	1
	beats in compound time	1
Scales/KeySignatures	Types of scales	3
	Modes	1
	Key signatures	4
	Relative major and minor	1
	scale degree names	1
Intervals	Intervals - Quality only	2
	Intervals - Quality and size	4
	Identification of intervals	4
Triads	Triad - Quality only	8
	Triad Inversion	2

(i.e., all items have the same four possible answers in the same order). Items in other areas, such as notation, test many more aspects of the topics; these items are more heterogenous.

This new exam was entered into a SuperCard stack for delivery on Macintosh computers. The Supercard program consists of three windows: an introduction and instruction window, a exam item window, and a results window. The program works as follows from the students' perspective: The students see the introduction screen and, upon clicking the Continue button, are asked a series of questions about their name, social security number, major, and primary instrument. A file is created for each student to store this information as well as the answer to each item and the results on the exam. The exam then begins. The exam is presented in a linear fashion; all students are presented every item in the same order. After the students choose an answer, a dialog box indicates whether the answer is correct or not, and the answer is recorded. Although telling the students that their answer is correct or incorrect may lead to some learning, there is probably little learning taking place since no explanation of why the answer was wrong is given.

At the end of the exam, the students are shown their results in the form of the percentage correct for each subtest and for the exam as a whole. In addition, the program indicates which class the students are recommended to take. If the students score at least 60% on each subtest and on the exam as a whole, they are recommended to take Music Theory 110B (a regular section of Theory I). If the students score lower than 60% on any subtest or the entire exam, they are recommended to take Music Theory 110A (a remedial section of Theory I) if they are music majors or Music General 100 (Fundamentals) if they are not music majors. Students may then print the results and take their recommendation with them. The scores and recommendations are stored in a computer file. A program is available to format and print the names of all the students who have taken the exam, their social security numbers, and their results for distribution to advisors and for their student files.

Placement 97 was given to students already enrolled in Music Theory 110A in the Fall of 1997 as a trial and subsequently to all students auditioning for the School of Music in the Spring of 1998 (119 students). An analysis was run on all 133 subjects' answers. The results are presented in Table 5.

The results of the analysis shows that the mean score on this exam was 32 (64.5%) with a standard deviation of 9.551. The KR-20 result for the exam was .909, the highest KuderRichardson result of all three placement exams. The Spearman-Brown prophecy number of 55 indicates that only five more items need to be added to this exam to obtain a reliability of .90, a great change over the almost doubling of the number of items on the last exam.[28] This new placement exam is thus an improvement over the older exams.

---

[28]In Spring 1999, five more items were added to the Placement 96 exam, one item in each of the five categories, making a total of 55 items on the exam. This new exam was given to 159 prospective music majors on their audition days. The KR-20 for this new exam indicated a reliability of .923.

Table 5: Descriptive Statistics for Placement 1997

	Whole	Subtest 1 Notation	Subtest 2 Rhythm	Subtest 3 Scales	Subtest 4 Intervals	Subtest 5 Triads
Number of Items	50	10	10	10	10	10
Number of Subjects	133	133	133	133	133	133
Mean score	32.241	7.692	7.233	6.248	5.271	5.797
Variance of Scores	91.230	2.927	4.286	5.703	6.760	9.284
Standard Deviation	9.551	1.711	2.070	2.388	2.600	3.047
Kuder-Richardson 20	.909	.585	.639	.726	.714	.827
Standard Error (from KR20)	2.884	1.103	1.243	1.251	1.392	1.267
Spearman-Brown Prophecy (from KR-21): To obtain a reliability of .90 the test should contain this many	55					

The KR-20 results on the subtests are also encouraging, since each one is an improvement over the KR-20 subtest scores on Placement 96. It is also interesting to note that the KR-20 scores increase with each subsection of the exam, again suggesting that the more homogenous the items, the higher the score.

Although the KR-20 scores change for the better on each subsequent subtest, the mean scores on the subtests fall from a high of 7.692 (76%) on notation to a low of 5.271 (52%) on intervals. (The mean score of the Triad subtest, 5.797, is higher than the intervals subtest but lower than all other subtests.) These mean scores are similar to those of Placement 96 (see Table 6), indicating that the students' knowledge has not differed much from one year to the next. All mean scores improved except for the score on the notation subtest. It is not yet certain why this decline occurred, but the change of score is minimal.

Correlation of Tests Scores to Grades

A very important factor in the evaluation process is to determine how well the exams serve as indicators of the students' performance in the theory classes they were recommended to take. To investi-

Table 6: Comparison of Mean Scores on Placement 96 and Placement 97

	Whole	Subtest 1 Notation	Subtest 2 Rhythm	Subtest 3 Scales	Subtest 4 Intervals	Subtest 5 Triads
Placement 96 Mean Scores	63.2%	74.4%	76.6%	56.6%	52.6%	55.7%
Placement 97 Mean Scores	64.4%	76.9%	72.3%	62.4%	52.7%	57.9%
Difference	+1.2%	+2.5%	-4.3%	+5.8%	+.01%	+2.2%

gate the relationship of exam scores to performance, the total correct scores on Placement 96 and Placement 97 were correlated to the grades the students received in their courses. Correlation, as Ott explains, is "One measure of the strength of the relationship between two variables x and y . . . . Given pairs of observations. . ., we can compute the sample correlation coefficient r . . . . r lies between -1 and +1. r > 0 indicates a positive linear relationship and r < 0 a negative linear relationship between x and y. r = 0 indicates no linear relationship between x and y."[29] What is hoped for, therefore, is a positive correlation coefficient close to 1.

To calculate the correlation coefficients, the students' grades for their theory classes had to be determined. For Placement 96, there were 69 students whose grades were found. Since Placement 97 has been given only one year, only 14 students had grades. It is important to remember that these 14 students were already taking theory classes and had been presented information on the exam prior to their taking the placement exam; their scores should thus be higher and correlate better with their grades.

The correlation of the total correct scores on the exam to grade point (e.g., 4.0 for A, 3.5 for B+) for Placement 96 was .4089 with a significant probability of .0005. Although this result is greater than 0 and therefore indicates a positive relationship, the relatively low score of .4089 indicates that the relationship is minimal. The same correlation for the 14 students taking Placement 97 was .7821 with a significant probability of .0009. Such a high correlation coefficient indicates that the exam better predicts students' grades in theory class and makes the exam much more useful for both the student and the teacher.

[29]Lyman Ott. *An Introduction to Statistical Methods and Data Analysis.* (North Scituate, Mass: Duxbury Press, 1977), 219-220

Conclusions and Future Plans

After studying the results of the analyses of the various exam scores and the correlation of exam scores to grades, one can conclude that the new exam, Placement 97, is an improvement over the older exams. One can also conclude that this improvement is due to the higher number of items on the exam and the homogeneous nature of the items on the subtests. Although a large amount of work has been completed to this date, more remains to be done in the areas of evaluation, revision, and programming projects for both the Placement exam and theory classes.

Once the students who took Placement 97 have completed their theory courses, their grades will be correlated to the total correct score on Placement 97. From the results of these correlations and the item analyses reported above, Placement 97 will be revised, improved and added to. The goal is to create a bank of exam items that the computer program can draw from so that students will not get the same items each time they take the exam. This process of testing, evaluation, and revision will continue for many years in order to create such a bank.

In addition, several improvements will be made to the exam. First, the items will be randomized, so that each student will receive the total number of items but not in the same order. Randomizing will help to prevent cheating as well as learning by the students since exam items on the same topic may not be given together. Other improvements to Placement 97 include improvements to the graphics and the reporting programs and the addition of a sample exam to be created for the Web. The sample exam would include items similar to those on the actual test so that students will know what to expect. Finally, mastery based tutorial programs on topics covered by the Placement exam are being created.

It is not the author's intent to make this exam a computer adaptive exam such as Tim Smith's "Ready or Not" for two reasons. First, the Placement 97 exam is performing well; its predictive validity is high (.78 as compared to Smith's RON at .52). Second, the author would like to control the minimum number of items on the exam and the exam makeup. The statistics used in this study (specifically the Spearman-Brown prophecy) indicate that at least 55 items are needed to make the exam reliable. A computer adaptive exam may not give the students the minimum number of items; it may make its decision based on a smaller sample of items. Smith reported that he found this to be true; in the second year of testing of his program, he instructed RON to withhold a recommendation until it had tested at least five objectives. The new placement exam should contain at least 55 items covering the five topic areas on the exam.

The second aspect the author would like to control is the exam makeup. Smith's adaptive exam begins by choosing items in areas that are high in discrimination value. As a student demonstrates mastery or non-mastery, the student is moved on to another objective. This type of item generation seems to assume that students who do not do well on one topic (such as intervals) will not do well on topics that rely on this information (such as triads). This is not always the case. While analyzing the results of the exams in this study, it was found that some students score well on topics seeming to require more knowledge (such as triads) even though their scores on "easier" topics (e.g., intervals, rhythm) were low. The new placement exam should test the student on all areas and make a recommendation on their cumulative knowledge.

Once all this work has been completed, an entire package will exist for theory students similar to the many computer packages that exist for ear-training students. This package will consist of a exam of theory skills and reports of these results for instructors and administrators. The package will also include a system to route students to lessons on which they need more work. Reports of the students' results on the lessons will be available to teachers for the purpose of student evaluation and further research. It is hoped that this.package of programs will be commercially available once it is completed. By having such a package, student skills may improve to the point that instructors can spend less class time on rudiments and more class time on issues of analysis and synthesis—the theoretical part of Theory class.

SELECTED BIBLIOGRAPHY

Anastasi, Anne. *Psychological Testing,* 3rd ed. New York: Macmillan Publishing Co., 1968

Colman, Peter James. "The Development and Validation of a Computerized Diagnostic Test for the Prediction of Success in the First-Year Music Theory Sequence by Incoming Freshman at Michigan State University." Unpublished Ph.D. dissertation, Michigan State University, 1990.

Kurtz, Albert K. And Samuel T. Mayo. *Statistical Methods in Education and Psychology*. New York: Springer-Verlag, 1979.

National Association of Schools of Music. *1999-2000 Handbook.* Reston, VA: National Association of Schools of Music, 1999.

Ott, Lyman. *An Introduction to Statistical Methods and Data Analysis.* North Scituate, Mass: Duxbury Press, 1977.

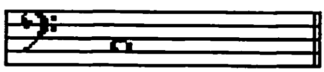
Smith, Timothy A. "An ExSPRT Systems Approach to the Assessment of Students Needing Remediation in Music Theory." *Journal of Music Theory Pedagogy* 8 (1994), 179-200.

APPENDIX 1:
SAMPLE ITEMS FROM PLACEMENT 97

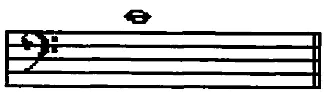
Notation

<u>Topic</u>: Anchor Notes
Which of the following is Middle C?

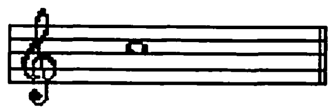
a.



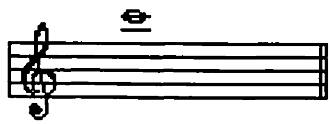
b.



c.



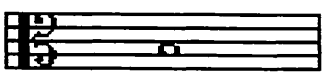
d.



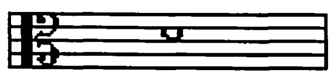
<u>Topic:</u> Clefs
Which note is a D?

a.

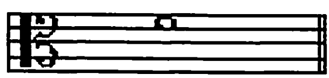
b.



c.



d.



<u>Topic:</u> Enharmonic Notes
Which pitch would sound the same as a C?

a. Db
b. Bx
c. B
d. Dbb

Topic: Aspects of notation
The note marked with an asterisk (*) below is a :

*

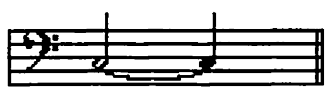
a. Bb
b. B
c. A#
d. D

Rhythm

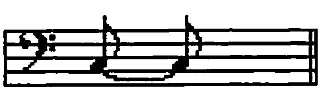
Topic: Note values
Which of the following is equal to the note below?

a.

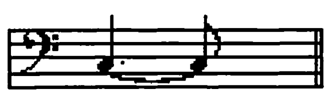
c.



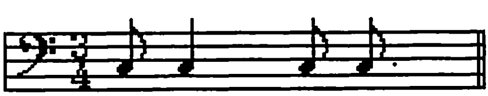
b.



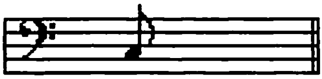
d.



Topic: Note/rest values - simple/compound time
Which note completes the following measure?



a.



c.

b.

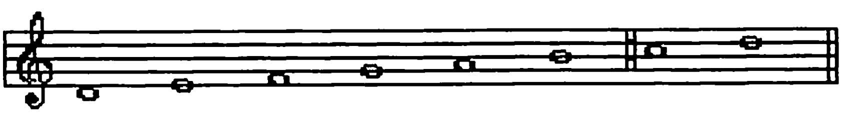
d.

Topic: beats in simple/compound time
The note below would get how many beats?

a. 1/2 beat
b. 1 beat
c. 1 1/2 beats
d. 2 beats

Scales/Key Signatures

Topic: types of scales
What scale is shown below?



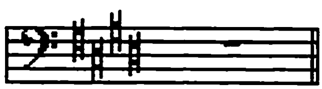
a. D major
b. D natural minor
c. D harmonic minor
d. D melodic minor

Topic: Modes
What scale is shown below?

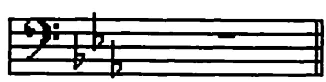
a. B Ionian
b. B Dorian
c. B Lydian
d. B Aeolian

Topic: Key Signatures
The key signature for E minor is which of the following?

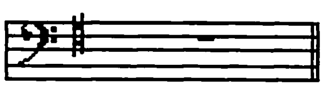
a.



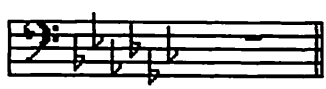
c.



b.



d.



Topic: relative major and minor scales
The relative major to A minor is:

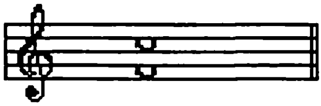
a. F major
b. F# major
c. C major
d. C# major

Topic: Scale degree names
The fifth scale degree is called the:

a. Tonic
b. Subdominant
c. Leading Tone
d. Dominant

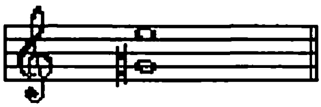
Intervals

Topic: Interval quality
The quality of the interval shown below is which of the following?



a. Major
b. Minor
c. Perfect
d. Augmented

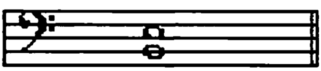
Topic: Interval quality and size
The interval shown below is which of the following?



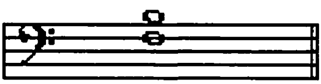
a. Major 3
b. Minor 6
c. Augmented 5
d. Diminished 4

Topic: identification of intervals
Which of the following intervals is a Perfect 4?

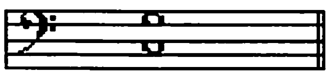
a.



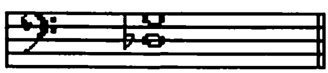
b.



c.



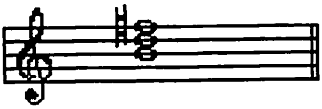
d.

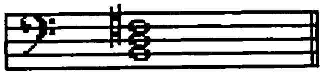


Triads

Topic: Triad quality

What is the quality of the following triad (shown twice in different clefs)?





a. Major

b. Minor

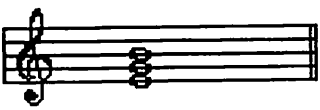
c. Diminished

d. Augmented

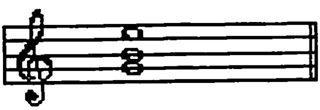
Topic: Triad inversion

Which of the following chords is in first inversion?

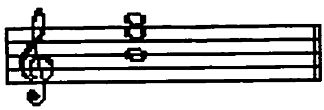
a.



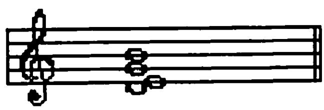
b.



c.



d.



Part Writing, Prose Writing: An Investigation of Writing-to-Learn in the Music Theory Classroom

Bruce C. Kelley

Introduction

In the last twenty years, educators have advanced the premise that writing, in various forms, can foster learning in every class in every discipline. While we routinely use writing in many ways to communicate, define a position, or make new connections, some forms of writing, such as lists, reminder notes, and diaries, tend to be mental scribbles indecipherable to all but the writer. These scribbles constitute one of language's most private forms: writing for the purpose of bringing personal meaning to a situation that previously had none.

If writing-for-oneself can be intricately bound with clarifying or remembering ideas, might this type of writing be used in the college classroom? In what ways might the use of language promote the learning of particular ideas or subjects—in this case, music theory? Since the 1970s, these questions have been underscored by a growing body of work that has discussed the impact of writing on learning, and turned attention to classroom writing in college curricula. Two major educational movements, writing-across-the-curriculum (WAC) and writing-to-learn (WTL), have evolved into philosophical and political rallying points for those who believe that writing is a necessary component of every academic course.

Focus on the pedagogical importance of writing has led to ground-breaking research on the writing process[1] and on the ef-

[1]See especially Emig, J. (1971). *The Composing Processes of Twelfth Graders*, Research Report no. 13. Urbana, IL: National Council of Teachers of English; and Britton, J. *et al.*, (1975). *The Development of Writing Abilities* (11-18). London: Macmillan Education Press.

fects of various forms of writing in the classroom.[2] These studies assume that writing can play an important role in a variety of disciplines, at all levels of schooling. The writing-to-learn philosophy has even been applied to learning music theory and aural skills in the last few years.[3] Rarely, however, has rigorous investigation been applied toward determining the effectiveness of writing-to-learn, especially in an essentially non-linguistic subject such as music theory. While many hypotheses have been formulated, few empirical studies have carefully measured writing's effects on learning, and none have examined the role that writing can play in the music theory classroom. In my own classes, I have discovered that certain writing exercises help students grasp new concepts. This experience has led me to explore more formally the effect of writing-to-learn on learning music theory constructs.

This article investigates the theories and claims of the writing-to-learn movement, and their effectiveness in the music theory classroom. The psychological roots of writing-to-learn are examined, as are some current theories about how and why writing helps students learn across all disciplines. The article then explores the results of an investigation at the Ohio State University that compared

---

[2]See especially Emig, J. (1977). Writing as a Mode of Learning. In D. Gaswami & M. Butler (Eds.), *The Web of Meaning: Essays on Writing, Teaching, Learning, and Thinking* (pp. 122-131). Upper Montclair: Boynton/Cook Publishers; Applebee, A. (1984). Writing and Reasoning. *Review of Educational Research* 54 (4), 577-596; Newell, G. (1984). Learning from Writing in Two Content Areas: A Case Study/Protocol Analysis. *Research in the Teaching of English* 18 (3), 265-287; Langer, J. A., & Applebee, A. (1987). *How Writing Shapes Thinking: A Study of Teaching and Learning.* Research Report no. 22. Urbana, IL: National Council of Teachers of English; Tierney, R. *et al.* (1989). The Effects of Reading and Writing Upon Thinking Critically. *Reading Research Quarterly* 24 (2), 134-173; and Menon, R. (1992). Writing to Learn Mathematics: Student Journals and Student-constructed Questions. (Doctoral dissertation, The University of British Columbia, 1992). *Dissertation Abstracts International*, 54, 2872A.

[3]See, for example, Davidson, L., Scripp, L., & Fletcher, A. (1995). Enhancing Sight-Singing Through Reflective Writing: A New Approach to the Undergraduate Curriculum. *Journal of Music Theory Pedagogy* 9, 1-30; and McGee, D. (1993). The Power of Prose: Writing in the Undergraduate Music Theory Curriculum. *Journal of Music Theory Pedagogy* 7, 85-104.

short, ungraded writing-to-learn exercises with creative part-writing exercises and visual error-detection exercises. Writing has been used in the music theory curriculum to record thoughts and communicate information, but the use of short, ungraded writing assignments, addressed primarily to the writers themselves, is relatively unknown. This specialized kind of writing is most favored by educators interested in using writing in all the disciplines because it creatively engages the students' thought processes. Part writing exercises required the student to notate an original setting that illustrated a specific problem or concept. Error-detection exercises required the student to circle errors or star correct solutions.

The investigation into these varied exercises showed that part writing, but not prose writing, was a significant factor in learning theory constructs, as measured by the pre- and post-tests. This article concludes by re-examining writing-to-learn theory and drawing some conclusions about writing and the process of learning music theory.

Instructional Paradigms

Writing prose does not often play an important role in the music theory classroom, where notes and the symbols of analysis provide the focus of communication. Typically the classroom teacher relies on the short analytical exercise and the workbook composition. Little emphasis is placed on prose writing except as a means of relaying information in analytical essays.

But might writing— specifically the short, ungraded writing assignment— help solve some of the challenges that face music theory pedagogy? Are prose writing exercises as effective for learning as exercises which involve written notation? Michael Rogers states that

> Theory, then, is not just something to learn but is also something *to do*. It represents not just a cluster of answers but a range of options for thinking about and listening to music. Music theory, in my opinion, is not a *subject* like pharmacy with labels to learn and prescriptions to fill, but is an *activity*—more like composition or performance. The activity is *theorizing*: i.e., thinking about what we hear and hearing what we think about—and I would include even thinking about what we think.[4]

Certainly there are times when workbook exercises are vitally necessary, and when lecture and example are more efficient than any prose writing assignment could be; but there may also be times when short writing assignments are the most efficient way to teach music theory. Psychologist Howard Gardner claims that:

> Much of the teaching and learning occurs through language—at one time, principally through oral instructions, employing verse, collections of adages, or simple explanations; and now, increasingly, through the word in its written form. A compelling example of this aspect can be found in the sciences. Despite the evident importance of logical-mathematical reasoning and symbol systems, language remains the optimal means for conveying the basic concepts in textbooks. In addition, language supplies the metaphors that are crucial for launching and for explaining a new scientific development.[5]

Prose writing may be especially relevant to helping students understand the philosophical side of the theoretical equation, and may also prove potent at helping students understand how they think about music theory itself. Paul Connolly, an advocate of writing-to-learn, states that the primary source of constructed knowledge, regardless of the discipline, is language—spoken or written:

> While musicians may speak to one another through their notes or mathematicians may spring to a blackboard to talk in numbers, natural language remains the most important mediator of concepts we do not yet fully hold. It negotiates all the necessary integrations between the experiences we have stored in our personal stories and the more abstract lessons that are filed in a culture's various systems of thought.[6]

Language is integral to the learning process. According to Connolly, "local" symbol systems, such as notation and numbers,

---

[4]Rogers, M. R. (1984). *Teaching Approaches in Music Theory.* Carbondale and Edwardsville, IL: Southern Illinois University Press, p. 7.

[5]Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences.* New York: Basic Books, Inc., p. 78.

[6]Connolly, P. (1989). Writing and the Ecology of Learning. In P. Connolly & T. Vilardi (Eds.), *Writing to Learn Mathematics and Science* (pp. 1-14). New York: Teachers College Press, p. 4.

are, for learning, subordinate to language, which is the primary conveyor of meaning. Does prose writing, an exercise completely separate from the activities commonly used to learn basic concepts of music theory, actually provide a new, more effective way to teach? Does the written "language" of music—notation—help students learn more effectively than written prose, or is the reverse true? At a deeper level, there lies this fundamental question: Just how do students best learn music theory constructs?

Writing-to-Learn: A Psychological Background

The writing-to learn-movement is based heavily on the learning theories of Russian psychologist Lev Vygotsky. Vygotsky's most influential work, *Language and the Mind*, examined the potential relationship between language and cognitive development in great detail. Vygotsky suggested that language was not just a window on the thought process, but an integral component of the thought process itself, a tool that actively influences cognition. There is a constant interaction between thought and language, the two becoming intricately and inexorably linked:

> Schematically, we may imagine thought and speech as two intersecting circles. In their overlapping parts, thought and speech coincide to produce what is called verbal thought.[7]

The space shared in common between the two circles represents what Vygotsky termed the *inner speech*, where thought and speech activity meet.

Vygotsky's description of inner speech was influential to later researchers, for it delineates the interaction of language and thought. Inner speech is more than just an internalized form of external speech; it is unique, with a special relationship to the other speech formations. It has a syntax that is disconnected and incomplete when compared to external speech:

---

[7]Vygotsky, L. (1992). *Thought and Language* (A. Kozulin, Trans.). Cambridge, MA: The MIT Press. (Original work published 1934), p. 88.

> It [inner speech] still remains speech, i.e., thought connected with words. But while in external speech thought is embodied in words, in inner speech words die as they bring forth thought. Inner speech is to a large extent thinking in pure meanings. It is a dynamic, shifting, unstable thing, fluttering between word and thought, the two more or less stable, more or less firmly delineated components of verbal thought.[8]

Vygotsky observed that inner speech is a distinct plane of verbal thought. As inner speech becomes uttered speech, a complex transformation occurs. Inner speech must be syntactically reinterpreted and organized before it can be understood as external speech.[9] Inner speech is the translator between thought and word, between outgoing and incoming messages. It becomes a tool that one uses to bring words to a thought, or thoughts from another's words. Vygotsky assumed that language, translated through the processes of the inner speech, becomes the means through which we learn.

Writing as a specific form of language is an important cognitive agitator in Vygotsky's theories. The relationship between writing and inner speech is in many ways a study in contrast:

> Inner speech is condensed, abbreviated speech. Written speech is deployed to its fullest extent, more complete than oral speech. Inner speech is almost entirely predicative because the situation, the subject of thought, is always known to the thinker. Written speech, on the contrary, must explain the situation fully in order to be intelligible. The change from maximally compact inner speech to maximally detailed written speech requires what might be called deliberate semantics—deliberate structuring of the web of meaning.[10]

The "web of meaning" is a description of how word and thought interact continually to produce a network of ideas and feedback:

> The relation of thought to word is not a thing but a process, a continual movement back and forth from thought to word and from word to thought. In that process, the relation of thought to word undergoes changes that themselves may be regarded as developmental in the

---

[8]Vygotsky, 249.
[9]Vygotsky, 249.
[10]Vygotsky, 182.

> functional sense. Thought is not merely expressed in words; it comes into existence through them. Every thought tends to connect something with something else, to establish a relation between things.[11]

Conceptualizing, creating meaning, constructing reality: these are all facilitated by the word, especially in its written form. The idea that writing interacts strongly with the inner speech to create meaning is one of Vygotsky's greatest contributions; it is an idea that polarized an educational movement dedicated to improving student's learning through writing.

Writing-to-Learn

The writing-to-learn movement began in the early 1970s as educators gained access to the first English translation of Vygotsky's *Language and the Mind.* The studies of Janet Emig,[12] James Britton *et al.* ,[13] and Douglas Barnes,[14] initially described what was known then as the Writing-Across-the-Curriculum movement.[15] In the early

[11]Vygotsky, 218.

[12]Emig, J. (1971). *The Composing Processes of Twelfth Graders.* Research Report no. 13. Urbana, IL: National Council of Teachers of English.

[13]Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The Development of Writing Abilities* (11-18). London: Macmillan Education Press.

[14]Barnes, D. (1976). *From Communication to Curriculum.* London: Hazell Watson & Viney Limited.

[15]The historical development and spread of the Writing-Across-the-Curriculum movement is best presented in the following articles: Freisinger, R. (1980). Cross-Disciplinary Writing Programs: Beginnings. In T. Fulwiler & A. Young (Eds.), *Language Connections: Writing and Reading Across the Curriculum* (pp. 3-14). Urbana, IL: National Council of Teachers of English; Maimon, E. (1982). Writing Across the Curriculum: Past, Present and Future. In C. W. Griffin (Ed.), *Teaching Writing in All Disciplines* (pp. 67-74). New Directions for Teaching and Learning, no. 12. San Francisco: Jossey-Bass; Petersen, B. (1982). Additional Resources in the Practice of Writing Across the Disciplines. In C. W. Griffin (Ed.), *Teaching Writing in All Disciplines* (pp. 75-82). New Directions for Teaching and Learning, no. 12. San Francisco: Jossey-Bass; and Tchudi, S. (1986). *Teaching Writing in the Content Areas: College Level.* Washington, DC: National Education Association of the United States.

1980s the movement shifted focus from curricular change to describing how writing affected learning, and adopted the following basic theoretical tenants:

(1) Writing can help students learn and think about content in any discipline, thus helping to achieve the goals of the instructor.
(2) Writing used for learning does not require explicit *teaching* of writing—only use of writing as a pedagogical tool. . . . This model highlights not writing for its own sake, but writing for the discipline's sake, i.e., writing for content mastery.[16]

In other words, the instructor of history or mathematics or music has no obligation to teach students how to write, but only to use writing as a tool for learning.

Many hypotheses of how writing—the symbolic representation of language—might enhance learning have been proposed. Some have claimed that writing manifests and enlarges the capacity to discover connections,[17] or that writing and learning share similarities because both are multifaceted, may be enriched by self-provided feedback, serve an analytical and connective function, and engage and commit the mind in their own particular rhythm.[18] Writing forces students to organize, to look for patterns, and to explain how and why things work.[19] The physical limitations on thinking required by writing force the writer to concentrate on a subject: reviewing, revising, and actively participating in the communication process itself.

---

[16]Kurfiss, J. (1985). Do Students Really Learn from Writing? *Writing Across the Curriculum* 3 (1), 3-4. (ERIC Document Reproduction Service No. ED 293 123)

[17]Knowblauch, C., & Brannon, L. (1983). Writing as Learning Through the Curriculum. *College English* 45: 473.

[18]Emig, Writing, 5-6.

[19]Boone, B. (1983). Great Teaching Ideas: Writing to Learn. *Virginia English Bulletin* 33 (2), 32-36.

The research of authors such as Emig,[20] Dowst,[21] Boone,[22] Applebee,[23] Gere,[24] and Archambeault[25] provides a basis for estimating writing's potential impact on learning music theory. The writing-to-learn principles emphasize teacher and class participation in a shared exchange of ideas so that students will gain a deeper understanding of topics than the simple lecture format can provide. Students learn through writing because they are forced to manipulate information in a personal context:

> Writing requires a high level of active processing of information. Unless students are merely copying word for word from printed text, they are translating the new information into their own words, forming connections with previously learned information, interpreting and applying it in new contexts. Writing is also the concrete visualization of content material. The learner is not only making connections, but creating new connections while developing a personal organizational schema.[26]

Creating new connections, constructing meaning and knowledge, and exploring implications are ways to describe writing's effect on learning. These ideas are important to the theory of writing-to-learn, for it seems plausible that writing, which is the most specific form of language, can enrich learning in every discipline. Writing, in general, enables students to grasp new concepts, and the short, personal, writing exercise, specifically, seems to show the most peda-

---

[20]Emig, Writing (1977).

[21]Dowst, K. (1987). The Epistemic Approach: Writing, Knowing, and Learning. In T. R. Donovan & B. W. McClelland (Eds.), *Eight Approaches to Teaching Composition* (pp. 65-85). Urbana, IL: National Council of Teachers of English.

[22]Boone, Great Teaching Ideas (1983).

[23]Applebee, A. (1984). Writing and Reasoning. *Review of Educational Research* 54 (4), 577-596.

[24]Gere, A. R. (Ed.). (1985). *Roots in the Sawdust: Writing to Learn Across the Disciplines*. Urbana, IL: National Council of Teachers of English.

[25]Archambeault, B. (1991, November). *Writing Across the Curriculum: Mathematics*. Paper presented at the 81st Annual Meeting of the National Council of Teachers of English, Seattle, WA.

[26]Archambeault, 3.

gogical promise.[27] Written language can increase the number of strands in the web of meaning (to use Vygotsky's analogy). How, then, could writing be used in the music theory classroom? A study that I conducted at the Ohio State University attempted to answer this question.

The Investigation

This investigation examined the effectiveness of the writing-to-learn strategy in the music theory classroom, and compared it to traditional exercises that also engage the students' thought processes. Three paradigms were used as treatments: a) short writing-to-learn exercises; b) part writing exercises; and c) visual error detection exercises. These exercises were chosen for comparison because each involves the physical act of putting pencil to paper. Quantitative and qualitative data were gathered from the sophomore music theory class through a series of in-class treatments, pre-, and post-tests. The treatments were administered through a crossover design with the intervention based on an a mixed effects model. Factors analyzed included treatment mode (error detection, part writing, and writing-to-learn), student grade history for the freshman music theory courses (in GPA, 0.0-4.0), student's musical instrument (instrumentalist, keyboardist, or vocalist) and gender, resulting in a 3.4.3.2 factorial design. The qualitative data were subjected to a protocol analysis. Part writing and error detection activities are common teaching tools in the music theory classrooms while writing-to-learn activities are not. This study explored the efficiency of each treatment in an effort to determine whether or not the short writing-to-learn exercises compared favorably with the traditional classroom practices.

---

[27]See, for example, Applebee, A. (1984). Writing and Reasoning. *Review of Educational Research* 54 (4), 577-596; Marshall, J. D. (1987). The Effects on Writing on Students' Understanding of Literary Texts. *Research in the Teaching of English* 21(1), 30-63; and Durst, R. K. (1987). Cognitive and Linguistic Demands of Analytic Writing. *Research in the Teaching of English* 21(4), 347-376.

Data on the three types of in-class treatments were collected to assess their effect on learning three harmonic constructs (the Neapolitan sixth chord, the augmented sixth chords, and the common-tone diminished seventh chord), specifically to see if any of the treatments generated either (1) higher improvement scores (post-test minus pre-test), or (2) higher post-test scores than the others. Performance instrument, music theory GPA, race, and sex were also examined to see if these factors interacted with the treatments.

A mixed-effects general linear model served as the means to analyze improvement and post-test scores. Since this study used a cross-over design, multiple responses (one set for each treatment) were collected from each subject. The effect of using multiple responses was included in the analysis to discover if responses from the same subject were similar. The analytical model had to account for missing data, and this was accomplished by using the following model from Milliken and Johnson's *Analysis of Messy Data*:[28]

$$y_{ijkm} = \mu + \rho_{im} + \Pi_k + \tau_j + \lambda_{i(k-1)} + \varepsilon_{ijkm}$$

Throughout this analysis, a factor was considered significant if its p-value (denoted Pr > F) was less than 0.05.

The pre- and post-tests for each topic were identical, and measured the students' knowledge of a specific topic before and after it was taught in the classroom (Figure 1).

[28]Milliken, G. A. & Johnson, D. E. (1984). *Analysis of Messy Data*. Belmont, California: Lifetime Learning Publications, p. 446-448.

Figure 1: Sample of Pre/Post Test (Subject: Augmented Sixths)

Augmented Sixth

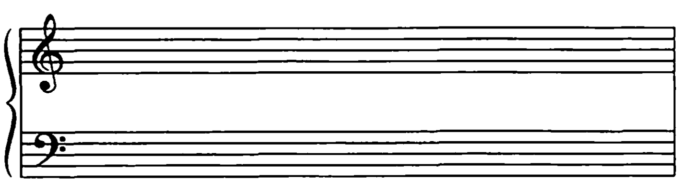
Code Name__________

Teacher_____________

1. Describe the augmented sixth chords in as much detail as possible.

2. Provide a Roman numeral analysis for the following progression:

3. Create a four-voice progression in g minor that starts and ends on I and includes an augmented sixth chord:



The pre- and post-tests included three independent tasks: a part-writing exercise, a prose writing exercise, and a short analysis exercise. The pre- and post-tests were analyzed through a point system developed for each of the harmonic constructs. Generally, the tests measured whether a student could a) recognize the specific concept in an analysis, b) discuss the concept in his or her own language, and c) create the concept in the context of a four-part voice-leading example. Specific grading rubrics were designed for each test (Figure 2).

Figure 2: Scoring rubric for augmented sixth chords

Analytical Identification:	
Did the student recognize the chord in a musical context? (Two examples @ 1 point each—student must have labeled as some sort of augmented chord.)	2 pts.
Did the student identify the chord appropriately? (Two examples @ 1 point each.—student correctly identified chord as German, French, or Italian)	2 pt.
Description of Chord	
Is the Chord construction explained? (list three types, mention +6, mention root, mention additional notes—.5 points each)	2 pts.
Is the voice leading explained? (Augmented sixth resolving—.5, other voices—.5)	1 pt.
Is context explained? (Preparation and resolution, .5 each.)	1 pt.
Four-part harmonization:	
Is the chord approached correctly in context? (Chord selection, voice leading .5 each)	1 pt.
Is the chord built correctly in context? (Augmented sixth, other voices—.5 each)	1 pt.
Are special voice leading concerns taken into account? (#4, b6 .5 each)	1 pt.
Are the rest of the voices resolved correctly? (.5 each for other two voices)	1 pt

These tests were analyzed in several ways. First, each of the three independent tasks was individually analyzed for significance. The tests were then examined in total to see if any of the variable factors had an effect on the overall score. Each analysis was conducted twice, once using the improvement scores (the difference between each student's pre- and post-tests), and once using just the scores generated by the post-test (ignoring the student's performance on the pre-test). This process generated eight separate analyses (Figure 3).

Figure 3: List of Analyses

- Improvement Scores:
  - Part-writing Task
  - Writing Task
  - Analysis Task
  - Total Test Score
- Post-Test Scores:
  - Part-writing Task
  - Writing Task
  - Analysis Task
  - Total Test Score

The full statistical analysis examined the following variables: time period of the treatment, the treatment itself, order effect, instrument, treatment as it interacted with instrument, GPA, treatment as it interacted with GPA, sex, and sex as it interacted with GPA. If the full statistical analysis showed that a factor was significant at the .05 level in the full model, a reduced model of variables was run for computational ease. This model included only the following variables: time period of treatment, treatment, order effect, instrument, GPA, and sex. If a factor was still significant in the reduced model, Bonferroni multiple comparisons were performed to determine which factor levels were different.[29]

[29]Readers interested in greater detail of this investigation are referred to Kelley, B. (1997). *Part Writing, Error Detection, and Writing-to-Learn: Three Strategies for Learning Music Theory Constructs* (Doctoral dissertation, The Ohio State University, 1997). Dissertation Abstracts International, UMI Number 9721115.

Statistical Implications

Statistical analyses of data indicated that significant factors influenced students' scores. These include music theory GPA, performance background, and treatment. Perhaps to no one's surprise, GPA in music theory was a significant factor. Students with a higher GPA scored better than students with a lower GPA.

The student's applied background was a significant factor in the statistical analyses. Several different analyses showed that musical instrument was a significant factor (see Table 1):

Table 1: Musical instrument as a significant factor

Key to identification numbers:
1 = woodwind, brass, and string players
2 = piano and percussion players
3 = vocalists

Analysis:	Average Improvement:*
Improvement Score Analyses	
Analysis of Writing Task**	1 scores higher than 3 by 0.55
Post-Test Score Analyses	
Analysis of Writing Task	1 scores higher than 3 by 0.66
Total Post-Test Scores	1 scores higher than 3 by 1.68
Total Post-Test Scores	2 scores higher than 3 by 1.73

*The average improvement is a measure of how test results compare between two variables. The analyses of independent tasks show average improvement out of a total of four points, while the analyses of the total test scores show average improvement out of a total of twelve points. Thus, the very first line indicates that woodwind, brass, and string players scored an average of .55 (out of four) points higher than did vocalists on the writing task.

**Significant at the .06 level rather than the .05 level.

Wind and string players were able to answer the prose writing task on the pre- and post-tests (note that this is different from the writing-to-learn treatment) more effectively than vocalists, regardless of treatment.

Vocalists also scored significantly lower on the total scores when compared to the other instrument groups. These results indicate that vocalists represent a population of special concern for the music theory teacher. This may be an artifact of earlier training. Because vocal education at the high school level seems not to focus on good music reading skills, many vocalists come to the college level with an underdeveloped understanding and comprehension of music fundamentals. While there are certainly exceptions to the rule, it seems that the average vocalist needs an extra push to learn the concepts of music theory.

The statistical analyses failed to indicate conclusively which treatment helps vocalists learn most easily, but the post-test analysis of the analytical task provided a possible answer. Instrument and treatment are intertwined for significance in this statistical measure, which showed that vocalists who received the writing-to-learn treatment scored an average of 1.57 units *lower* on the analysis task than vocalists who received the part-writing treatment, and that vocalists receiving the writing treatment scored 1.24 units *lower* than vocalists who received the error-detection treatment (see Table 2).

Table 2: Treatment as a significant factor for vocalists

Key to treatment numbers:
1 = writing
2 = error detection
3 = part writing

Analysis:	Average Improvement per unit:
Analytical Task	
treatment 1 to treatment 2	1 scores lower than 2 by 1.24
treatment 1 to treatment 3	1 scores lower than 3 by 1.58

This is a rather remarkable disparity, considering there is only a range of four units total. It reveals that analysis was learned more readily with notational exercises (part writing and error detection) than with writing exercises. Vocalists did better with the part writing and error detection treatments than with the writing-to-learn treatment, and this may indicate that they simply need more time "in the music" than their counterparts in the instrumental world. Treatment was also a significant factor for all groups in this study when considering the post-test scores only (see Table 3).

Analysis of post-test scores for the writing task showed that students receiving the part writing treatment scored 0.53 units higher than students receiving the error detection treatment. In the analysis of total post-test scores, treatment was again a significant factor, showing that students that received the part writing treatment scored 1.11 units higher, on average, than did the students receiving the writing-to-learn treatment. For teaching harmonic constructs, part writing was clearly the most effective of the three treatments. This finding raises interesting issues in light of the fact that the writing-to-learn movement has shown promise in almost every field. Why is it that prose writing did not help students as readily as part writing did?

The writing-to-learn philosophy is based on the idea that "inner speech," the mechanism that translates innate knowledge into active vociferation, is activated most readily by the written word. The process of writing words slows the learner, so that communication can be clarified. Music, however, is basically non-linguistic,

Table 3: Treatment as a significant factor

Key to treatment numbers:
1 = writing
2 = error detection
3 = part writing

Post-Test Score Analysis:	Average Improvement per unit:
Analysis of Writing Task	3 scores higher than 2 by 0.53
Analysis of Total Post-Test Scores	3 scores higher than 1 by 1.11

even music that accompanies text. According to Serafine, music is cognition "in which the artform is not considered a clearly specified external object, but rather an internal, subjective entity springing from mental operations."[30] The most essential manner of learning music, then, may be purely through the medium itself. Part writing is, after all, a form of writing, and the many benefits of writing words may also be derived from "thinking" with notation. Perhaps writing-to-learn principles did prove valid, but in writing *notation*, rather than in prose.

Notation as Language

As a subject, music theory has unique characteristics that set it apart from other fields, even those such as mathematics that use symbolic language. The discipline focuses on notation, yet that notation is not music. Rather, it is a written representation, much the same as written prose can serve as a symbolic representation of spoken language. Many elements of writing-to-learn theory are applicable to music—but the writing involved concerns notation rather than language. The role of notation in learning music can be reexamined by returning to the theories of Vygotsky and the writing-to-learn proponents.

Vygotsky focused on language as a fomenter of learning with his paradigm of inner speech. His model of the interaction between language and thought allows for non-linguistic experiences in learning:

> Schematically, we may imagine thought and speech as two intersecting circles. In their overlapping parts, thought and speech coincide to produce what is called verbal thought. Verbal thought, however, does not by any means include all forms of thought or all forms of speech. *There is a vast area of thought that has no direct relation to speech.*[31] (Emphasis mine.)

---

[30]Serafine, M. L. (1988) *Music as Cognition: The Development of Thought in Sound*. New York: Columbia University Press, p. 233.

[31]Vygotsky, 88.

The inner speech that Vygotsky modeled forms a direct link between speech and thought. Might there be an area of pure thought not covered by inner speech that translates between thought and music—an "inner music"? Inner music would be condensed and abbreviated, just like inner speech. The act of transferring musical thought from the maximally compact inner music to the intricately detailed written notation might create what Vygotsky called "deliberate semantics"—the web of *musical* meaning. A musical web of meaning could describe how notation and thought interact continually to produce a network of ideas and feedback. Perhaps music is learned through an inner pool of thought that is intricately linked to the musical process, a process that creates musical meaning for each of us. Serafine describes this type of musical cognition as "the activity of thinking temporally with sounds, both simultaneous and successive."[32] Musical cognition does not just follow sensory input, but actively constructs musical events.[33]

"Inner music," like inner speech, would necessarily be dynamic, shifting, and unstable. Perhaps, as inner music becomes perceived music, a complex cognitive transformation occurs as the inner music is syntactically reinterpreted and organized. An inner music modeled on Vygotsky's inner speech would continuously and spontaneously engage with the music we hear and perform. Understanding harmonic functions, a goal of music theory, would then be facilitated by stimulating the interaction between thought and the musical phenomenon. Part writing, using the representational system of music, appears to help create a web of meaning needed to learn and understand the fundamentals of what is heard.

Vygotsky laid a foundation for exploring why notation, as the physical, symbolic representation of music, is an effective mediator for learning the concepts of music theory. The interaction between pure thought and musical action, as represented by an inner music, may explain why writing notation proved more effective than the other treatments. This effectiveness may also stem, however, from qualities that part writing shares with the writing-to-learn exercises.

---

[32]Serafine, 72.
[33]Serafine, 73.

Notation: The Written Sound

The writing-to-learn theorists claimed that writing prose influenced learning for a variety of reasons, most of which also seem applicable to writing musical notation. Writing notation, like writing prose, forces students to sharpen their focus to successfully resolve problems and to seek out and organize patterns. The factors that contribute to the positive role of prose writing in thinking may also apply to writing music: the permanence of the product, allowing revision and review; the explicitness required to maintain meaning; the facilitation of organization; and the active nature of the process.[34] The act of writing notation forces students to engage in the very activities that writing-to-learn proponents advocate. The benefits of writing prose can be directly applied to writing notation. Unless students are simply copying music, they are being forced to create new connections with what they have previously known. Part writing and prose writing share many of the same relationships to learning.

Recent cognitive research supports the idea that part writing is an important cognitive agitator in music. The cognitive theory of multiple intelligences[35] maintains that intelligence is multifaceted; rather than a single base intelligence, humans have a number of differing kinds of intelligences: mathematical, musical, linguistic, social, physical, etc. Diana Deutsch's research supports this hypothesis. In 1973 she investigated different types of interference patterns on her subject's musical memory, and found that non-musical interference patterns had less effect on musical memory than did musical interference patterns. Her research implies that musical memory is more dramatically influenced by musical events than by non-musical events.[36] David Marr suggested that the different in-

[34]Applebee, 577-596.

[35]See especially Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books, Inc.; and Gardner H. (1985). *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books, Inc.

[36]Deutsch, D. (1973). Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics* 13 (2), 271-275.

telligences, or "modules," operate according to their own principles, and have little interaction with the other intelligences.[37] Marr's theory implies that musical intelligence—in this case the recognition and construction of certain constructs—is aided through musical exercises that specifically activate the musical intellect.

Writing-to-Learn in Music Theory Instruction

The question remains: is prose writing a viable alternative in the music theory classroom? My answer is a cautious yes, but not in the role initially envisioned in this study. Music is different from other disciplines because it incorporates a language unique to itself. As Barrow states, "Just as there appears to be a universal genetic programming of humans that endows linguistic skills, so there might be a *universal musical grammar* that plays the same role for sound patterns."[38] (Emphasis mine.) Writing-to-learn theory holds that verbal language remains the most important mediator of concepts, even in symbolic fields. While for most music students this holds true, some students intuitively learn the grammar of some styles of music simply by listening to the music itself. There are too many examples of talented self-taught musicians who have a thorough aural grasp of advanced harmonic constructs to claim that verbal language is *necessary* for learning music. Musical genius aside, the normal music theory course can probably benefit from the use of writing in the classroom.

Borrowing from the studies in mathematics, this study examined writing as a primary pedagogical tool for learning specific music concepts. Writing was not as successful in this role as part writing. Prose writing may have a place, however, in helping students when the music itself is ambiguous, where context is ill-defined, and where students must think critically to make the appropriate decisions. As Rogers states:

---

[37]Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

[38]Barrow, J. (1995). *The Artful Universe*. Oxford: Clarendon Press, p. 194.

> The natural inclination to weight those aspects of musical experience that are the most "teachable" and "testable" should be carefully examined. Theory teachers, too often, tend to overstress topics or questions that permit only clear-cut right or wrong answers while avoiding those gray areas of ambiguity that can be so treacherous. Channeling of the thought process into black-and-white categories is important at a beginning stage so that basic concepts can be established, but eventually a theory program must move ahead to create tolerance and enthusiasm for the discovery, exploration, and comparison of a wide range of differing musical ideas and must promote the ability to back up decisions and judgments with logic, consistency, and imagination. In the long run, most questions that do have unequivocal answers turn out to be insignificant, whereas those that allow a variety of interpretation have the power to kindle real musical insight.[39]

Writing may be a way to find answers to interpretive situations, not by defining an augmented sixth chord, for example, but by helping a student decide if a particular sonority is functioning like an augmented sixth or something else. This study provided situations that were very clear for the students; little ambiguity was thrown at them in terms of context. If samples of real music had been used, however—samples that provided confusing situations—prose writing might have had a greater impact. Different learning goals may have also resulted in different outcomes for the study. If the investigation had measured how students grasped the concepts of dominant preparation or harmonic prolongation, or the idea of musical tension, then the writing-to-learn exercises may have had a greater impact. This is an area of research that should be explored in the near future.

Conclusion

This study examined the effect of three types of exercises on learning music theory. Of the three—part writing, error detection, and writing-to-learn—part writing was the most effective. The creative

[39]Rogers, 5.

openness of the exercise, and the musical literacy it required made it the most efficient way to help students learn the selected harmonic constructs. Prose writing may have a useful role in the music theory classroom, but not as a primary method of instruction for tasks such as these. Writing's role may be of more limited value in the music theory classroom than the writing-to-learn advocates anticipated.

The tools we bring to the pedagogy of music theory can include a variety of tasks, but exercises that engage the students in the notational system work best for the constructs examined in this investigation. Future curricular development should recognize this, and create for students an environment that facilitates their learning efficiently and well.

MusicCog/99
A Workshop Symposium in Music Cognition.

Reviewed by Edward Klonoski

MusicCog/99 was held on the campus of the Ohio State University from May 14-16, 1999. The program described the event as differing from more traditional conferences in that many of the MusicCog/99 presentations were general or review in nature—so that the workshop was an excellent opportunity to gain an overview of the state of the art in music cognition. Presentations covered a broad spectrum of topics including: infants' musical perception, ethnomusicological applications, melodic expectancy, key determination, music and emotion and rhythmic/metric perception. Overall, the papers were very informative and represented well the breadth of current music cognition research.

The conference featured a distinguished group of invited speakers, including Carol Krumhansl (Cornell University), Dirk-Jan Povel (Nijmegen Institute for Information and Cognition, the Netherlands), and Sandra Trehub (University of Toronto), as well as presentations by a number of leading music cognition researchers affiliated with Ohio State's Center for Cognitive Science. In addition, the annual Poland Lecture, which honors long-time, distinguished Ohio State music theory faculty member William Poland, was delivered by Nicholas Cook (University of Southhampton, England).

Given the expansive range of topics currently being explored in music cognition, one might reasonably expect that music instructional techniques and goals, especially those that fall under the broad umbrella of aural skills, would present fertile ground for the investigation of musically significant questions. However, the relationship of the research presented at MusicCog/99 to music theory pedagogy was, for the most part, indirect and of limited import. It should be noted here that pedagogy was not a central focus of the conference. In fact, with only a few exceptions, discussion of potential implications of perception research for developing and refining musical understanding and skills—the primary goal of music theory and aural skills instruction— was entirely avoided. This raises important questions about the relationship between music cognition and music theory pedagogy. But before examining more fully the issue of why a more direct and meaningful exchange between these two fields has failed to materialize, I will report some of the pedagogically relevant findings I was able to infer from the conference papers and discussions.

In her investigation of infants' musical perception, Sandra Trehub found that babies are better able to detect pitch changes in scales containing unequal steps than in scales composed of equal steps. She also found that adults are extremely biased by their overfamiliarity with the major scale, yet they detect pitch change no better with unfamiliar, unequal-step scales than with unfamiliar, equal-step scales. This reinforces the pedagogical tradition of using the major scale to introduce the general topic of scales, and then proceeding to other less familiar scale forms. Perhaps more importantly, it raises the possibility that further research might reveal a kind of optimal order for teaching various scale types that is based not on conceptual principles, but on perceptual grounds. Such potential points of convergence between cognition and music theory pedagogy need to be made explicit and further explored.

David Temperley, in his presentation "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," examined the performance of several key-finding algorithms and proposed a revised model that more accurately determines the key of a given passage when provided with input data. Temperley characterized his algorithm as a preference rule system "which considers many possible analyses of a piece or passage, evaluates them by certain criteria, and chooses the highest-scoring one." The principal improvement of Temperley's algorithm over other models is its ability to re-evaluate a key assignment in consideration of a larger context. His model continually reassesses key determinations in light of new pitch data. Temperley chose a computational, rather than an experimental, model because "seeking a model that performs a task successfully may shed light on how humans perform the task."

However, he acknowledges that a computer model that successfully performs a task may not reflect the cognitive processing that humans do to execute the same task. Until a direct relationship between key finding algorithms and human cognitive processes can be demonstrated, the pedagogical applications of such models are limited.

There were several papers devoted to melodic expectancy, a topic with considerable potential to inform aural skills pedagogy. Yet, the research presented showed more promise as a descriptive tool for distinguishing musical styles than as an instructional tool for refining perceptual skills. Melodic expectancy results from the interaction of numerous musical parameters, including rhythm, meter, timbre, harmony, and melody. Largely for reasons of experimental control, the studies presented at the conference focused almost exclusively on melody. This may well be a necessary first step in understanding the highly complex mental processes involved in melodic expectancy. Still, the research in this area does not yet adequately account for the influence of musical parameters other than melody to provide a basis for systematically teaching or refining melodic expectancy skills.

Because the rest of the papers presented at the workshop dealt with topics unrelated to music theory pedagogy, I turn now to the consideration of the tenuous relationship between music cognition and music theory pedagogy. One would expect that music theory and aural skills instruction would be the direct beneficiaries of cognition research. Based on the description of music cognition provided in the MusicCog/99 program, there would seem to be substantial overlap between the two fields: "Music cognition approaches the study of music as a manifestation of human minds/brains. The field involves music theorists, systematic musicologists, psychologists, cognitive scientists, and philosophers interested in comprehending human music-making and musicality." Music-making and musicality occupy central positions in music pedagogy as well. However, the considerable insights gained in recent years into how humans process auditory stimuli have had little if any meaningful influence on the methods used to develop and enhance musical skills and understanding. Similarly, the experiential/anecdotal evidence of human cognition upon which so many music instructional techniques are based finds its way into comparatively few cognition studies. To be sure, there is no single reason underlying the disconnect that exists between the two fields, and both disciplines have been slow to finds ways to bridge the gap.

That pedagogical issues were all but excluded from a conference with such an impressive representation of diverse disciplines is telling. Equally revealing was the conspicuous absence of music theorists and music educators from the workshop, save for a few Ohio State faculty and student participants. Whether this reflects a general sense of apathy toward music cognition on the part of educators and theorists, or is simply indicative of the failure of cognition research to reach the theory and aural skills classrooms, is open to debate. Most likely it is a combination of the two, with each feeding the other and resulting in an ever-widening communications gap. Since many cognition researchers have never taught a theory or aural skills course, it seems obvious that theory and aural skills instructors need to be more actively engaged in the field of cognition if substantive pedagogical questions are to posed and investigated.

I am not the first to put out such a call. David Butler, a leading cognition researcher at Ohio State, lamented the growing chasm between cognition and pedagogy in his article "Why the Gulf Between Music Perception Research and Aural Training?"[1] Butler acknowledged that few cognition experiments have much direct relevance to "real" music. This likely stems from the need for highly constrained experimental conditions that often restrict researchers to using contrived materials with only limited musical application—isolated intervals, melodic fragments in the absence of harmony or meter, etc. Butler urged music educators to become more involved in cognition research. He strongly suggested that it is music educators, rather than cognition specialists, who are best positioned to take the first step in the drive to develop effective teaching strategies founded on systematic scientific research. I echo his sentiments here and extend his exhortation to music theorists as well.

[1]David Butler, "Why the Gulf Between Music Perception Research and Aural Training?" *Bulletin of the Council for Research in Music Education* No.132, (Spring 1997): 38-48.

In fairness to theorists and music educators, it can be quite challenging to extract from cognition research information that can be easily applied in the classroom. At least part of the difficulty results from the abundance of scientific/psychological terminology that pervades most cognition books, articles, and conference presentations. The average music educator or music theorist is not well versed in scientific methodology and terminology. Yet given the current thrust of cognition research, we should not expect a decrease in the use of scientific language any time soon. Of course, jargon renders many subdisciplines within music—theory near the top of the list— inaccessible to the uninitiated. So, the problem is not new, nor is the solution: dialogue.

The language barrier is not one of mere semantics. Rather, it points to a fundamental difference between the goals of music cognition and those of music pedagogy. I suggested earlier that the highly constrained, scientifically framed experiments reported on at MusicCog/99 may be a necessary early stage in an evolving theory of music cognition that one day will be able to more fully accommodate the richness and complexity of real compositions. But it seems just as likely that the general direction of current cognition research is not a prelude to a theory aimed at better understanding music or teaching others to understand music. A discussion at the MusicCog/99 workshop succinctly captured the essence of this observation. In response to a question concerning the musical relevance of a particular cognition study, one of the conference participants responded that they [cognition researchers] "are studying cognition, not music." That is, the primary goal of music cognition is not to understand music, but to use music as a means to understand how the human mind processes auditory stimuli. Music provides an ordered, logical medium to investigate cognitive processes, but the vast majority of music cognition research focuses on understanding *existing* perceptual skills and not on refining or developing new perceptual skills.

This should not be taken as an indictment of the goals of music cognition research. It is intended as an honest assessment of the relationship between the research presented at MusicCog/99 and music theory pedagogy. The field of cognition is vibrant, exciting, and will undoubtedly continue to expand our understanding of numerous aspects of human cognition both directly and indirectly related to music. Its contribution to music theory pedagogy is at present unclear. The potential for a mutually beneficial partnership exists, but this potential can only be realized if participants from both fields begin to identify and actively pursue common goals.

Conference Report: Society for Music Perception and Congnition

Reviewed by James Mathes

The Society for Music Perception and Cognition held its annual conference August 14-17, 1999 on the beautiful campus of Northwestern University in Evanston, Illinois One of the attractions of the meetings of this society is the opportunity to listen to and interact with researchers within and outside of the discipline of music, including a large number of participants from outside of the USA. Institutions from Canada, the UK, Germany, France, the Netherlands, Sweden, Finland, Brazil, and Argentina were represented. Music addressed ranged from western classical music to jazz, popular and folk music, to Hindustani music. In all, twenty-seven paper sessions were held over a four-day period as the work of over 140 researchers was presented. These numbers are indicative of the wide range of topics that were addressed by the diverse group of participants. Although the field of music perception and cognition is traditionally associated with psychology, music theorists have taken more active research roles in recent decades, reflecting a growing interest across disciplines. The music theory departments from Northwestern and Ohio State Universities, particularly active in music perception and cognition research, were especially well represented on the programs. Though it was not possible to attend and to report on many intriguing sessions that were held—concurrent papers were continuously offered—I hope to give some idea of the types of research being conducted as evidenced by this conference.

The conference began on Sunday morning with three papers on atonal music as one of two concurrent sessions. All three papers reported on experiments concerning theoretical description and aural perception of similarity relations among atonal sonorities. Tuire Kuusi (Department of Composition and Music Theory, Sibelius Academy) reported on an experiment in which 58 subjects (unspecified) first rated similarities between pairs of pentachords, and then rated individual chords based on a number of separate verbal dimensions. She concluded that consonance or dissonance seemed to be the most important factor in rating similarities (chords associated with familiar sonorities, notably the dominant seventh, were rated highest), though whole-tone associations and chord settings (spacing, register, salient intervals) were noted as important factors. The results confirmed that interval-class content alone cannot explain aural perceptions.

Panayotis Mavromatis (Eastman School of Music) reported on research conducted with Virginia Williamson (School of Music, Penn State University) in which thirty novices (undergraduate psychology majors) and thirty-eight music majors participated. Their earlier work using expert listeners had identified the significance of interval content, common tones, and salient intervals in perception. Extensive handouts documented their more recent findings that expert listeners tend to correlate atonal sonorities with theoretical descriptions and that their perceptions are much different from non-experts. Perhaps most interesting, the consistency of responses among experts correlated significantly with the number of years of applied training though not with academic music credits completed or with absolute pitch.

Arthur Samplaski (School of Music, Indiana University) reported on an experiment taken by twenty-two "trained" musicians. Pairs of Z-related tetrachords [0146] and [0137], and [0157] were heard in varied voicings and timbres (clarinet and Shepard tones). The pitch height (6-14 semitones in his voicings) between outer voices was interpreted as the most important factor and the position of dissonant intervals as a secondary factor. The different timbres were apparently not a significant factor.

Each of these papers in different ways addressed the need for further research to clarify how different listeners prioritize various factors that shape perceptions of atonal pitch structures. Subsequent discussions emphasized that perception of atonal sonorities should be studied in context rather than in isolation, a recurring issue in many perception studies. Williamson and Mavromatis' work toward the long term goal of a perceptual definition of similarity and categorization of atonal sonorities seems especially promising.

A session on memory included two papers of particular interest. Christine Koh and Lola Cuddy (Department of Psychology, Queen's University at Kingston) conducted an experiment in which trained pianists were asked to recall unfamiliar folk melodies through reproduction at the piano. Their interest was in studying the ability to recall actual melodies rather than artificially generated sequences through reproduction (rather than notation). Subjects were given aural versions with no notation, notation with no aural version, or notation/sound together. They then had to play the tune back and MIDI data was collected. The researchers analyzed the mistakes to determine if serial (left-to-right, note-to-note) processing was used, or whether subjects processed in a hierarchical manner, i.e. subjects would forget embellishing but not structural pitches. An expert had rated the pitches of the melodies on the basis of structural importance. Their analyses of data showed that subjects were strongly inclined toward serial rather than hierarchic processing and that the conditions with visual presentation produced much better recall than the aural-only mode. One observer noted that in dictation exercises students often fill in the cadence, a hierarchic type of listening encouraged by aural skills. Thus, the results of this experiment seem to indicate that the mode of recall is a function of the task. For those involved in aural skills training, this research suggests the need to consider further the specific requirements of the different skills we teach. Comparison across different types of music training modality, suggested by the researchers, could be invaluable.

W. Jay Dowling (University of Texas-Dallas, Program in Cognitive Science) presented a paper on "The Flow of Music and Memory" in he which discussed how our memories of what we have just heard change as time passes. In a series of experiments, he observed that if we hear a melody and are asked to judge whether a comparison melody is exact or altered, we do better after a moderate-length delay than after a very short delay. His most interesting assertion was that brain processing of music just heard continues even as we hear new music. The implication is that a listener needs time and reflection to process structure and that the memories of a piece fluctuate in the course of listening.

A session on Sunday morning focused expressly on how various musical skills are acquired. Papers here tended to address specific performance problems—presentations on student practice habits, and on motor learning in advanced and novice performers were given—with one notable exception. In his paper "Auditory Imagery and Music Pedagogy: Reciprocity and Untapped Potential," Edward Klonoksi (School of Music, Northern Illinois University) offered a challenge to researchers in the field of auditory imagery as well as those involved in aural skills pedagogy. Though these fields in some ways follow parallel lines of inquiry, he argued, many perceptual studies focus on tasks that have limited application in real musical environments. Much of the research in auditory imagery, for example, demonstrates or examines basic imagery processes. In music pedagogy the capacity for basic imagery is assumed and the emphasis is on refining and expanding that capacity. On the other hand, the pedagogical ordering of aural skills, as was examined briefly by comparison of the tables of contents of two ear-training texts, is quite varied and largely based on anecdotal evidence; in his view, "no systematic, scientific backing" has been established to help shape current pedagogical strategies for aural skills. Auditory imagery researchers, he asserted, are well positioned to investigate how auditory imagery is developed and refined and thus have great potential for helping to shape music instruction. Conversely, aural skills pedagogues can provide a "virtually limitless stockpile of musically relevant questions" for investigation. For example, can imagery tasks be ordered according to difficulty? Most likely Klonoski's basic assertions and concerns would resonate with anyone who has faced the challenge of teaching aural skills, and his conclusion that only positive outcomes can result from increased interaction between the two fields rings true. What other types of musically relevant questions need to be asked, or a program of research for the problems and concerns raised was not addressed.

The question of innate musical capacities was raised to some extent during a session on the subject of tonality in music. In a paper directly related to aural training, Renè Van Egmond (Nijmegen Institute for Cognition and Information) reported on research conducted with Mila Boswijk on the perception of tonality in musical excerpts and chords. Listeners were asked to identify tonic by singing and by playing a keyboard 1) after each hearing of twenty musical excerpts from popular music, and 2) after hearing major, minor, and diminished triads, and dominant seventh chords in arpeggiated and block chord forms. Two different experiments were conducted, one using subjects with dissimilar training, the other using subjects with similar training. The subjects included music majors with limited ear training (one semester or less) and subjects who had completed a two-year ear training sequence. The results of the experiments as reported were interesting on several counts. In experiment 1, neither familiarity with the actual music nor the amount of ear training were significant factors in performance. In both experiments, listeners who scored 100 percent on the first task also indicated one tonic as uniquely belonging to the chords heard in the second task; the group who scored 75 percent or less produced many different tonics for the chords. Perhaps not surprisingly, the lower group most often heard the root of a dominant seventh as a tonic. These results suggest that the ability to identify a tonic pitch may be more innate than learned. One observer commented that there may not be as much difference between the skill level at the beginning and end of an ear training sequence, though others disputed this. The researchers emphasized their belief that discrepancies with other studies cited may reflect the difference in pools of subjects as well as the type of stimuli used: classical versus popular music, real music versus isochronous tone sequences.

Another study presented by Alexandra Lamont (Department of Psychology, Leicester University) examined the sensitivity of trained and untrained children (6-16 years of age) and adults to tonal content and hierarchy using probe-note methodology. To briefly summarize her findings, trained subjects had better sensitivity to tonal hierarchy though musically trained children aged 15-16 had more stable (consistent) results than musically trained adults. By extreme manipulation of materials, which minimized salient melodic features, she demonstrated the limits of probe-note methodology. Examples were played—a welcome part of the presentation too often omitted in many papers—that clearly illustrated the care one must take in considering the relation of the experimental model and the task at hand to real musical contexts when interpreting results.

A session on style and structure had presentations that largely focused on cognitive rules and perception. Eugene Narmour (Department of Music, University of Pennsylvania) gave a thought provoking presentation on the subject of whether or how listeners may or may not invoke cognitive rules that shape musical expectations. Though he argued persuasively that listeners do use rules, he raised questions of how we distinguish compositional rules from listening rules, and how stylistic and time-span constraints affect the application of rules.

David Temperley (School of Music, Ohio State Univeristy) reported on his efforts to model a listener's perceptions of contrapuntal music through a preference rule system and computational model. His research served to verify the importance of interval size, tempo, and rhythm in the perception of linear components, and identified some problems with the rigidity of certain well-formedness rules, e.g. the prohibition of crossing voices. Mila Boswik addressed a similar concern with preference rules as she applied Lerdahl and Jackendoff's well known methodolgy to the analysis of popular music. Her basic argument was that popular music requires different preference rules than classical music in the western tonal idiom. Nicola Dibben (Department of Music, University of Sheffield) reported on research conducted with Alexandra Lamont on the perceptions of similarity relations in tonal and atonal music. Listeners were asked to rate similarity between first and second themes of sonata form movements (Beethoven, Op. 10, No.1, I and Schoenberg Op. 33a) The subjects included musicians and non-musicians, with both predictable and surprising results. Non-musicians tended to be attentive to similar surface features (largely thematic) in both examples while musicians were more attentive to differences in pitch and tonality. This finding would be expected. What was puzzling was their finding that the first theme of the exposition often was not rated as very similar to the first theme in the recapitulation in either work. A problem was determining how the researchers or subjects were defining similarity and how factors were weighted, neither of which was entirely clear from the abstract or the presentation. (Unfortunately, no examples were played.) One explanation may be that multidimensional scaling was used, which groups the stimuli by listener ratings but does not interpret those groupings (the experimenter does). Further, there are several musical dimensions to be considered, and varied weighting of different dimensions may critically affect ratings and interpretation of results. These concerns aside, their ongoing research on perceptual versus functional notions of similarity in music, and on the perceptual relationship between musical surface and structure is addressing important questions concerning how expert and non-expert listeners hear music.

Both of the above topics were visited in other sessions. In a session on cross-modal cognition and music, Siu-Lin Tan gave an engaging report on research, conducted by a group of psychologists from Kalamazoo College, which examined the listening strategies of musically trained and untrained listeners. Subjects were told to write marks in a large rectangle as they listened to five, stylistically diverse orchestral excerpts, each one minute long and heard twice; no words, letters, or conventional musical symbols were permitted. They then were given five minutes to explain their markings through written descriptions. The thirty subjects who scored high on a musical literacy test tended to focus on elements of musical structure (theme/motif, tonality, form) and musical surface (i.e. tempo, volume, register), while thirty subjects low in musical literacy tended to focus on affect (emotive descriptions) and extra musical associations. These may be predictable results, but the types and consistency of the responses is noteworthy. Only 1 percent of the untrained listeners ever mentioned theme or motive; none mentioned key. The trained listeners generally leaned toward abstract though simplistic symbols but gave detailed musical descriptions, while the untrained listeners leaned toward pictorial symbols and often gave "silly narratives." Other observations pointed to the influence of a knowledge of musical notation on responses and generally corroborated that the formal training of musicians influences their listening perceptions.

In the same session, Candace Brower (School of Music, Northwestern University) presented her research and ideas on spatial imagery in music. After briefly discussing established models of musical space from historical treatises such as those by Heinichen, Mattheson, Vogler, and Schoenberg to current writings by theorists and psychologists including Krumhansl, Balzano, Lerdahl, and Cohn, she explored the metaphorical implications of such models. Her metaphorical notions of containers (e.g, space within a chord or key), pathways (ongoing motion), and goals (tonal arrivals) led to discussion of a model for what she termed "image-schematic analyses." Time did not permit an elaborated example of such analyses, but her discussion seemed to imply a greater emphasis on spatial motion and on aspects of the musical surface than are found in most other types of structural analysis. A question was raised concerning whether her work is developing a perceptual or an analytic model, and subsequent discussion addressed the sometimes elusive distinction between the two.

Two different sessions on expressive performance reflected the interest and value of studying expert performances as another way of examining the relation between musical surface and structure, and of reconsidering the basis and validity of that dichotomy. Of particular interest, based on the abstract and subsequent discussions, was the work of Van Egmond and Richard Ashley (School of Music, Northwestern University), which illuminated the importance of considering multi-dimensional aspects of musical structure in explaining contrasting performances of melodies in different contexts. Unfortunately, as in many cases, time and logistics prevented my attendance at the presentation.

I have highlighted specific sessions and papers that hopefully are of interest to readers of this journal, though a wealth of other topics were covered. Sessions on melody, timbre, rhythm, cognition and meter, music and language, music and media, emotion, and improvisation and expression, among many others, were presented. An evening session on music and evolution was especially interesting, with a particularly entertaining presentation by David Huron (School of Music, Ohio State University) on music-making as an evolutionary adaptation. Sessions such as this, as well as many of the papers and sessions mentioned, frequently involved topics, methodologies, or ideas normally outside the purview of traditional music disciplines or curriculums. Such research often challenges, in meaningful ways, our basic assumptions about music and about how we teach it, making attendance at an SMPC conference a valuable experience for anyone interested in music.

In closing, I must applaud the excellent organization and direction of the conference, overseen by Richard Ashley, and the wonderful hospitality shown by the faculty, staff, and students of Northwestern University. A memorable evening at Ravinia with great food and a gala concert by the Chicago Symphony Orchestra highlighted four stimulating days of activity. Attendance at an SMPC conference comes highly recommended. The next meeting will be in Toronto, a joint meeting of fifteen music societies to be held November 2-5, 2000.

CONTRIBUTORS

Bruce Kelley is Associate Professor of Music at Shepherd College in Shepherdstown, WV, and coordinates the Music Theory and Composition Area. His research is cross-disciplinary, integrating elements of music theory pedagogy, music cognition, and writing-to-learn.

Edward Klonoski is Associate Professor of Music at Northern Illinois University. He holds a Ph.D. from the Ohio State University, and his research interests include aural skills pedagogy and auditory imagery. He is currently a member of the Editorial Board for the *College Music Symposium*.

James Mathes is Associate Professor of Music Theory and Coordinator of Music Theory and Composition at the Florida State University School of Music.

Barbara Murphy is Associate Professor of Music Theory in the School of Music of the University of Tennessee-Knoxville.

Timothy A. Smith is Professor of Music Theory in the School of Music of Northern Arizona University. He has recently joined the JMTP Editorial Review Board.

Joseph Swain is Associate Professor of Music at Colgate University in Hamilton, NY. His most recent three books are: *Harmonic Rhythm: Analysis and Interpretation* (2002), *The Broadway Musical: A Critical and Musical Survey* (2002), and *Musical Languages* (1997).

GUIDELINES FOR CONTRIBUTORS

1. Articles on any aspect of teaching or learning music theory will be welcome. Contributions will be judged on originality, relevance, interest to a diverse audience, and clarity of writing.

2. Manuscripts should be typed double-spaced (including footnotes, references, and quotations) on 8 1/2 x 11" paper with at least one-inch margins. Please submit *five* clear copies. All footnotes, tables, figures, musical examples, and other material should be placed at the end on separate sheets with an indication of where they fit in the text. Long musical examples and complex diagrams or charts should be avoided when possible. Please also avoid musical symbols (i.e. notation) within the running text.

3. The author of an accepted manuscript will be asked to revise the article (if necessary) to fit the style guidelines of the journal and to submit it electronically in either Windows or Macintosh formats. Authors will be sent guidelines for style and guidelines for electronic manuscripts. Please do *not* send disks until requested to do so.

4. The author is responsible for obtaining permission to reproduce copyrighted material and for paying permission fees, if necessary. Permissions must accompany the electronic submission of accepted manuscripts.

5. The author is responsible for providing camera-ready copy for all non-text items (tables, figures, musical examples, and other material). This material should also be submitted electronically if possible.

6. Manuscripts are accepted on the condition that they are unpublished and are not presently being submitted for publication elsewhere. Since all submissions are reviewed anonymously, please include the author's name and address only in the cover letter and eliminate identifying references (such as names of schools) from the article.

7. Please address all correspondence to:

> *Journal of Music Theory Pedagogy*
> School of Music
> University of Oklahoma
> Norman, OK 73019

The *JMTP* staff will send guidelines for style and electronic manuscripts to anyone requesting this information. On most matters of form and style, *JMTP* follows *The Chicago Manual of Style.*