The purpose of this study was to evaluate instructional activities and practice techniques of musicians attempting to improve their accuracy in detecting errors in music examples heard. In this study, a commonly recommended practice procedure, keyboard sight-reading, was compared with listening to recorded examples of simple piano works characteristic of those used in college-level class piano courses. The authors randomly assigned 59 college music majors to two groups. One group (Group R) studied examples by sight-reading music excerpts at the keyboard; the other group (Group L) studied the same excerpts by listening to recordings. Both groups were tested using taped examples of the excerpts with harmonic alterations. When data were collected for harmonic alterations not detected (misses) and for errors indicated where none were performed (false alarms), Group L was significantly more accurate (p = .0001) in detecting harmonic alterations than was Group R. The difference between the two groups was the same regarding false alarms (p = .0001). A repeated measures design was employed 2 weeks later with similar results. The data also indicated a possible effect of treatment order (listening first or sight-reading first). Implications are drawn for classroom application and for further study.

Gail de Stwolinski, James Faulconer, and A. B. Schwarzkopf, *University of Oklahoma, Norman*

A Comparison of Two Approaches to Learning to Detect Harmonic Alterations

This study was based on research we are now conducting to evaluate instructional and practice activities of musicians who are attempting to improve their accuracy in detecting errors in music examples. We compared the techniques of keyboard sight-reading and listening to recorded examples of simple piano works.

Authors of many widely adopted textbooks in music theory have suggested that keyboard reading is a valuable way to acquire familiarity

with excerpts from music literature. Bruce Benward (1985) suggested that examples throughout the text should be studied and ". . . if possible, [the students should] play them on the piano" (p. x). Kosta and Payne (1984) suggested that "if you cannot arrange to be at a piano while reading this book, try to play through the examples just before or right after reading a particular section or chapter" (p. ix). Many textbooks include tape-recorded examples of the printed excerpts or exercises. These tapes are variously considered as "integral parts of the text" (Trubitt & Hines, 1979, p. ix) or as supplementary practice materials (Benward, 1985, p. x). Although these suggestions are valuable for the ear-training student, little, if any, research has been conducted to evaluate their effectiveness when used as practice for detecting harmonic errors.

In the broader area of music memory, most of the research has focused on memory of melody and rhythm. Consequently, the literature is limited and findings are contradictory regarding the variables that influence accuracy in detecting harmonic alterations in music excerpts. Generally, there is a paucity of published research on the learning processes through which young musicians acquire the ability to detect harmonic alterations and on any of the pedagogically feasible methods of evaluating this ability.

The authors of the following articles, however, have reported research that relates directly to our research efforts: In each study, subjects used detection of errors to identify harmonic alterations. Hansen (1955) found that piano performance experience, piano sight-reading skill, and listening to chord quality were significantly related to successful detection of harmonic errors. Hansen also found that grades in ear-training courses were closely correlated with achievement on the detection test. Sidnell (1971), however, reported no correlation between aural achievement and the ability of instrumental music education majors to detect harmonic errors in tape-recorded instrumental excerpts. Brand and Burnsed (1981) found no statistically significant correlation between skill in detecting musical errors in instrumental examples and each of five variables (number of instruments played, ensemble experience, ability in music theory, skill in sight singing and ear training, and number of years of private instrumental instruction before entering college).

METHOD

Subjects and Materials

We recruited 59 college music majors who had reached at least Level III of class piano from two universities, the University of Oklahoma at Norman and Oklahoma Baptist University at Shawnee. (Level III includes students who have demonstrated minimum keyboard proficiency in the third semester of college-level class piano. None of the subjects had declared piano performance as a major.) We randomly assigned the subjects to two groups, Group L (the listening group) and Group R (the sight-reading group).

The items for study and testing were four short keyboard excerpts by four different composers (see Appendix 1). Class piano instructors at both institutions determined the difficulty of the excerpts to be Level III. It was important to have a variety of styles represented in the excerpts so that the results would not be influenced by a familiarity with a single composer's style. Each excerpt was taken from the beginning of a composition since this portion generally exhibits many of the stylistic characteristics of the total work. Two additional criteria for the selection of the four musical excerpts were lengths of 6 to 12 measures (approximately equal in time span although different tempi) and closure with an appropriate cadence. Excerpts were deliberately brief to avoid the confounding effect that saturation might have on the subjects.

Four to five harmonic alterations were inserted at logical points. Usually only one note of the chord was altered, which changed the chord quality, the chord function, or both. Some examples of these alterations were changes from a dominant to a leading tone chord, from a leading tone to a supertonic chord, or from a diatonic seventh chord to an augmented sixth chord.

A correct version of each excerpt, professionally performed on an electronic piano, was recorded and copied twice as preparation material for Group L. The three correct versions of each excerpt were followed on the tape by one incorrect version for collective testing, also professionally performed on an electronic piano.

Each subject in Group R heard a separate tape of the same incorrect version. Preparation in this group involved piano sight-reading of each excerpt three times in succession and testing through headphones. Several members of the piano faculty indicated that it was extremely important for the subjects involved in piano sight-reading to achieve a specified number of complete readings. If, instead, the sight-reading group had been allowed only a length of preparation time that was equal to the three correct performances of the professional performer, slower readers might have been stopped with a partial and frustrating aural experience immediately before testing. We therefore decided that equal preparation was better achieved by an identical number of excerpt replications than by identical time spans.

Procedures

The 59 college music majors who had reached at least Level III of class piano were randomly assigned to one of two groups of 29 and 30 students each. To assess the equality of piano background between groups, we gave a questionnaire to each subject to determine the number of years of piano study before college and the number of semesters of college piano study. A t test was applied to the results, and the difference between groups was not significant ($p = .87$).

Part 1

The subjects in both groups heard the four music excerpts during a 40-minute period. We decided on completion of all four excerpts within

one hour to avoid losing volunteer subjects who needed to attend other classes. Members of Group R (the sight-reading group) used the electronic piano to study each excerpt. The subjects were encouraged to sight-read the excerpts at a reasonable performance tempo; the proctor provided a metronomic tempo for reference before the subjects heard the presentation of each excerpt with harmonic alterations. We estimate that by starting each of the four excerpts together, those in Group R used a maximum of 60 seconds more time per excerpt than did those in Group L. Subjects in Group L studied each excerpt by listening to a correct tape-recorded performance. The recordings were performed on the same model of electronic piano as those used by Group R. After hearing the correct version of each excerpt three times with a 5-second pause between, subjects in Group L listened to one recorded perform-ance of the excerpt with harmonic alterations.

Both groups responded to the test playing of the altered version by marking the printed score on the beats that they perceived to be different from the preparation excerpt. Results were tabulated for each subject in two categories: (a) "misses," for those beats where a harmonic alteration was performed but none was marked on the score; and (b) "false alarms," for those beats where a mark was made by the subject although no alteration had been performed.

Part 2

For the second part of the study, we reversed the study procedures of the groups to create a repeated measures design. This occurred for each subject at least 2 weeks after, but within 4 weeks of completion of Part 1. The same excerpts, equipment, and recordings were used. The subjects who had sight-read at the keyboard during Part 1 listened to the examples, and those who listened during Part 1 used the electronic piano as their study tool. Results were tabulated for misses and false alarms.

RESULTS

Part 1 of this study was a straightforward two-sample comparison between the Trial 1 responses of two groups of students, who applied alternative methods of study to the same four excerpts. Figures 1 and 2 show the average number of misses on each of the excerpts by students who used sight-reading or listening as their method of study. These figures show that students who used listening to study the excerpts missed fewer harmonic alterations (see Figure 1) and identified fewer false errors than did those who used sight-reading on all of the four works. Excerpt 4 (by Reinecke) produced considerably more errors than did the other works.

We used the analysis of variance (ANOVA) procedure, of which the results are shown in Table 1, to test the significance of the differences in number of misses and the number of false alarms on Trial 1. The different study methods (Method) and individual excerpts (Excerpt) were the main effects of the analysis. Interaction between these effects
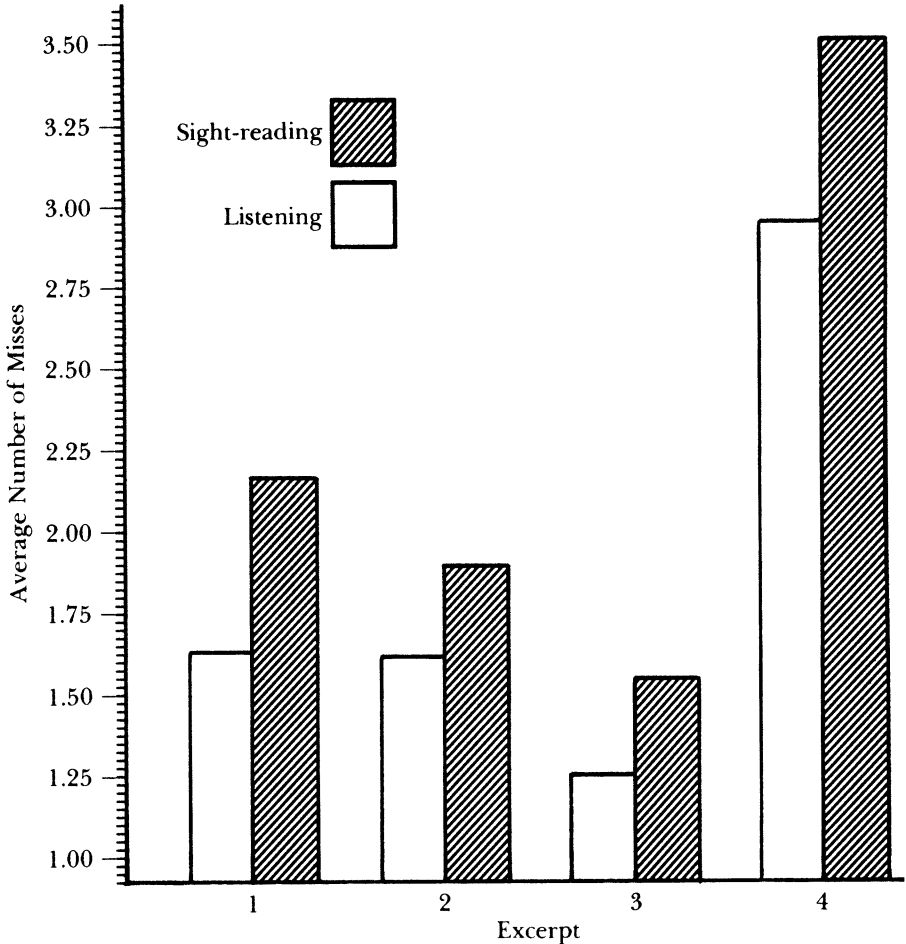
Figure 1. Listening versus sight-reading in music error detection: Average misses on each excerpt—Trial 1

(Method × Excerpt) was also included in the model to determine whether the effectiveness of the study method depended on which excerpt was being reviewed.

Technically the design was a three-way ANOVA with student identification number (ID) nested in Method and with Excerpt as the remaining main effect. We chose this design to obtain maximum power of the test to identify differences among study methods, taking into account the differences in test excerpts.

Review of the p values for Method in Table 1 shows that there was a very significant difference between the effectiveness of sight-reading and listening as study methods for detecting harmonic alterations ($p <$ .001). Differences between excerpts were also significant. More important, the interaction effects were not significant ($p <$ .5051 for misses and
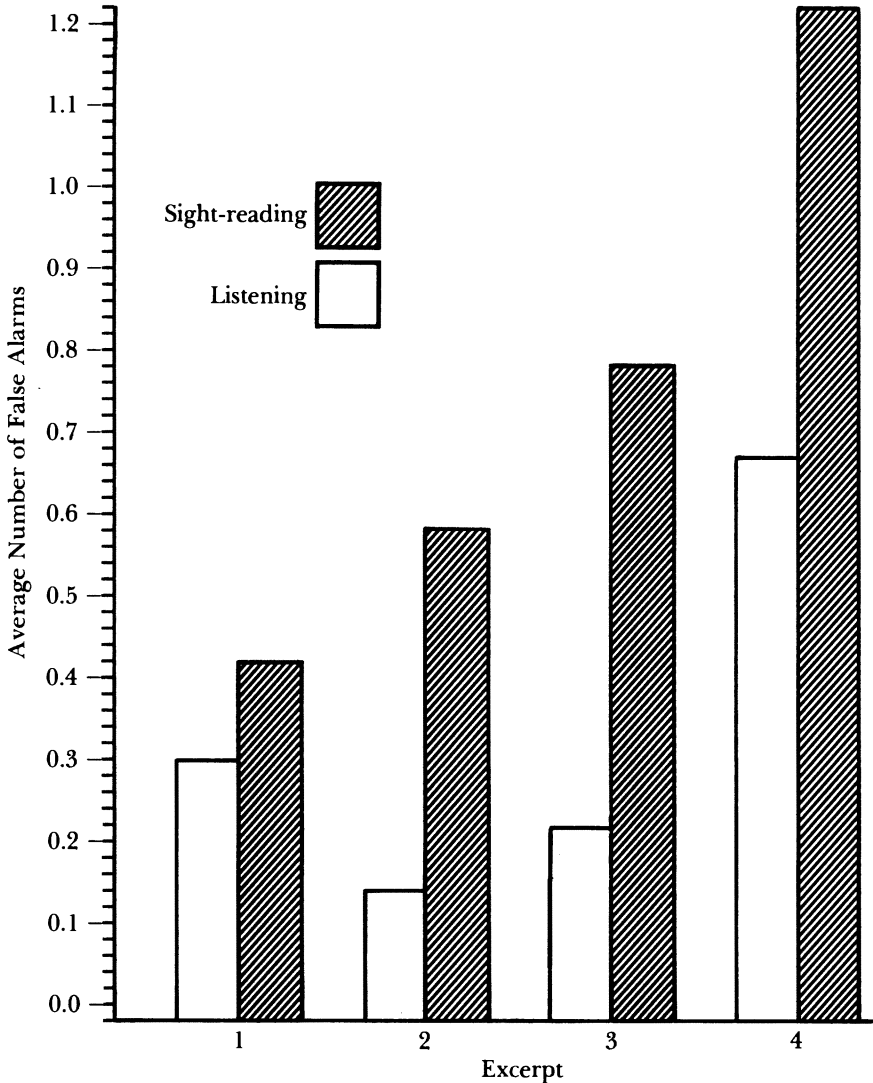
Figure 2. Listening versus sight-reading in music error detection: Average false alarms on each excerpt—Trial 1

$p < .2313$ for false alarms). This means that listening was generally more effective than sight-reading over all the excerpts used in this study.

Part 2 of the study consisted of a second session with the same four excerpts but with study methods exchanged between groups. The original expectation was that learning from session to session would be minimal and that what learning did occur would be similar from group to group due to exposure to the experiment. This similarity would have produced a controlled experiment in which both groups received both treatments.

Table 1
Comparison of Sight-Reading and Listening as Study Methods for Learning to Detect Harmonic Alterations: Trial 1

Source of variation	df	Sum of squares	F	p
Misses [a]				
Method (sight-reading or listening)	1	9.00	6.86	.0010
ID[b]: error for Method	56	73.46		
Excerpt (Excerpt 1, 2, 3, or 4)	3	107.07	52.48	.0001
Method × Excerpt (interaction)	3	1.61	0.79	.5051
Residual: error for Excerpt and Method × Excerpt	167	113.57		
False Alarms [c]				
Method (sight-reading or listening)	1	9.50	12.54	.0010
ID[b]: error for Method	56	42.42		
Excerpt (Excerpt 1, 2, 3, or 4)	3	12.42	10.94	.0001
Method × Excerpt (interaction)	3	1.64	1.44	.2313
Residual: error for Excerpt and Method × Excerpt	167	63.19		

[a]Alterations not detected by subjects
[b]Student identification number
[c]Alterations marked by subjects where none were present

The design of the Part 2 analysis consisted of an ANOVA with two main effect components: Group ("read then listen" or "listen then read"), Method (sight-reading or listening), and with subject ID nested in the group. The dependent variable for this analysis was the average number of misses for each student over all four excerpts. We also used a more complex design, treating excerpt as an additional effect, but the results were not different from the other analysis, and they were much harder to interpret.

The results of this analysis appear in Table 2. The Group × Method interaction measures whether the anticipated learning effect is the same across both groups. Since this interaction was significant ($p < .002$ for misses and $p < .0001$ for false alarms), we concluded that the learning from trial to trial was different in the two groups and that the results of the main effects tests for differences due to Group or Method alone had to be examined very carefully.

Table 3 shows a more detailed analysis of cell differences for both misses and false alarms with contrasts based on the ANOVA in Table 2. The Group L contrast of misses in Table 3 is a calculation of the difference between number of misses sight-reading and listening by

Table 2

Comparison of Sight-Reading and Listening as Study Methods for Learning to Detect Harmonic Alterations: Trials 1 and 2 Combined

Source of variation	df	Sum of squares	F	p
Misses [a]				
Group ("sight-read then listen" or "listen then sight-read")	1	0.06	0.15	.7046
ID[b]: error for Group	57	24.02		
Method (sight-read or listen)	1	3.89	29.24	.0001
Group × Method (interaction)	1	1.33	10.03	.0026
Residual: error for Method and Group	52	6.91		
False Alarms [c]				
Group ("sight-read then listen" or "listen then sight-read")	1	0.52	2.35	.1309
ID[b]: error for Group	57	12.72		
Method (sight-read or listen)	1	2.38	27.63	.0001
Group × Method (interaction)	1	2.20	25.52	.0001
Residual: error for Method and Group	52	4.48		

[a]Alterations not detected by subjects
[b]Student identification number
[c]Alterations marked by subjects where none were present

subjects who sight-read on the first trial and listened on the second. For Group R, we also calculated the difference between the mean number of misses sight-reading and listening, but for the group who sight-read first. The T and p values in the table are measures of the statistical significance of these differences based on the errors shown in Table 2.

Table 3 shows that all contrasts were strongly significant except those for Group L. This finding means that listening produced significantly fewer errors than sight-reading except when comparing the sight-reading errors on Trial 1 to the listening errors on Trial 2.

A graphic summary of this information is shown in Figures 3 and 4, which show the mean number of misses and false alarms by each group in both Trial 1 and Trial 2. These figures also show individual excerpt results that reinforce the pattern from Table 3 and confirm the conclusion that listening is better study preparation than is sight-reading.

The two separate comparisons for Trial 1 and Trial 2 show that the difference in the average number of misses was essentially the same at

Table 3

Comparison of Sight-Reading and Listening as Study Methods for Learning to Detect Harmonic Alterations: Contrasts of Improvement[a] in Performance

Contrast	Improvement[a]	T	p
Misses [b]			
Group L (listen on Trial 1, sight-read on Trial 2)	0.144	1.43	.1598
Group R (sight-read on Trial 1, listen on Trial 2)	0.607	6.23	.0001
Trial 1 (first time through the excerpts)	0.380	3.93	.0003
Trial 2 (second time through the excerpts)	0.371	3.66	.0006
False Alarms [c]			
Group L (listen on Trial 1, sight-read on Trial 2)	−0.010	−0.12	.9064
Group R (sight-read on Trial 1, listen on Trial 2)	0.589	7.51	.0001
Trial 1 (first time through the excerpts)	0.396	5.09	.0001
Trial 2 (second time through the excerpts)	0.184	2.25	.0290

[a]Improvement = errors reading − errors listening.
[b]Alterations not detected by subjects
[c]Alterations marked by subjects where none were present

both times and was significant both times. A similar pattern existed for false alarms, although the difference was not significant ($p = .03$) at the .01 level. The average number of false alarms per excerpt on Trial 2 was only 0.14, however. Overall, these results show that listening was a better short-term preparation method for recognizing performance mistakes than was sight-reading.

An alternative view of these data can be obtained by examining individual group performances. These separate analyses show that improvement in performance was strongly dependent on the group. Group L subjects, who listened on Trial 1 and sight-read on Trial 2, showed no significant difference between sight-reading and listening on either misses or false alarms. In fact, the number of false alarms recorded after listening (Trial 1) was slightly larger than was the number of false alarms recorded after sight-reading (Trial 2). Group R, in contrast, recorded considerable improvement in performance after listening (Trial 2) compared with that after sight-reading (Trial 1).

These overall results are displayed in Figures 3 and 4 and confirm that there was considerable learning between Trial 1 and Trial 2. It is not clear from these data whether there was general learning from Trial 1 to Trial 2, as separate time analyses would suggest, or whether the order of tasks itself led directly to the improvement observed in Group R. In any
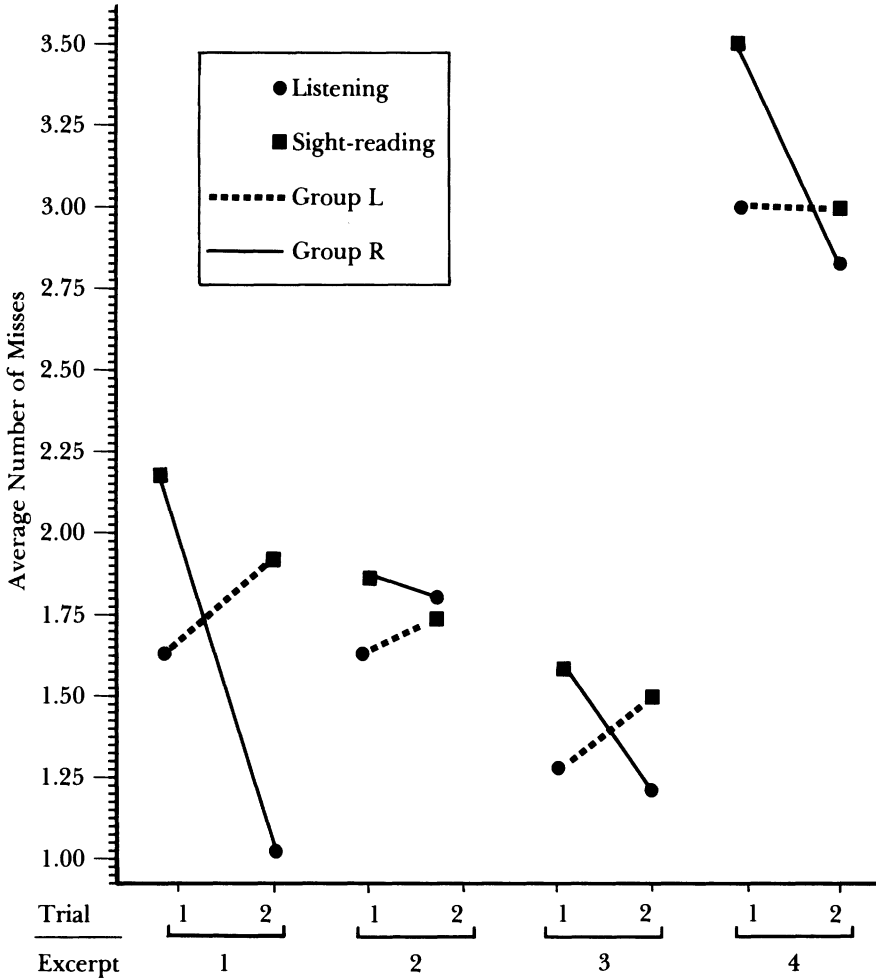
Figure 3. Listening versus sight-reading in music error detection: Average misses on each excerpt—Trials 1 and 2

case, listening to an excerpt immediately before evaluating it for errors produced better detection of errors than did studying it by sight-reading.

The results point to two important inferences for aural perception pedagogy: (a) Since listening to an excerpt is demonstrated as a more effective method of study than is keyboard sight-reading for students at this level of keyboard proficiency, students in aural perception classes should be strongly encouraged to listen to performance-tempo recordings of excerpts as a study method for class preparation; and (b) the difference in performance level of the groups based on the order of study needs further research. It is possible that using a combination of listening and sight-reading at the keyboard would be more efficient than using either method separately.
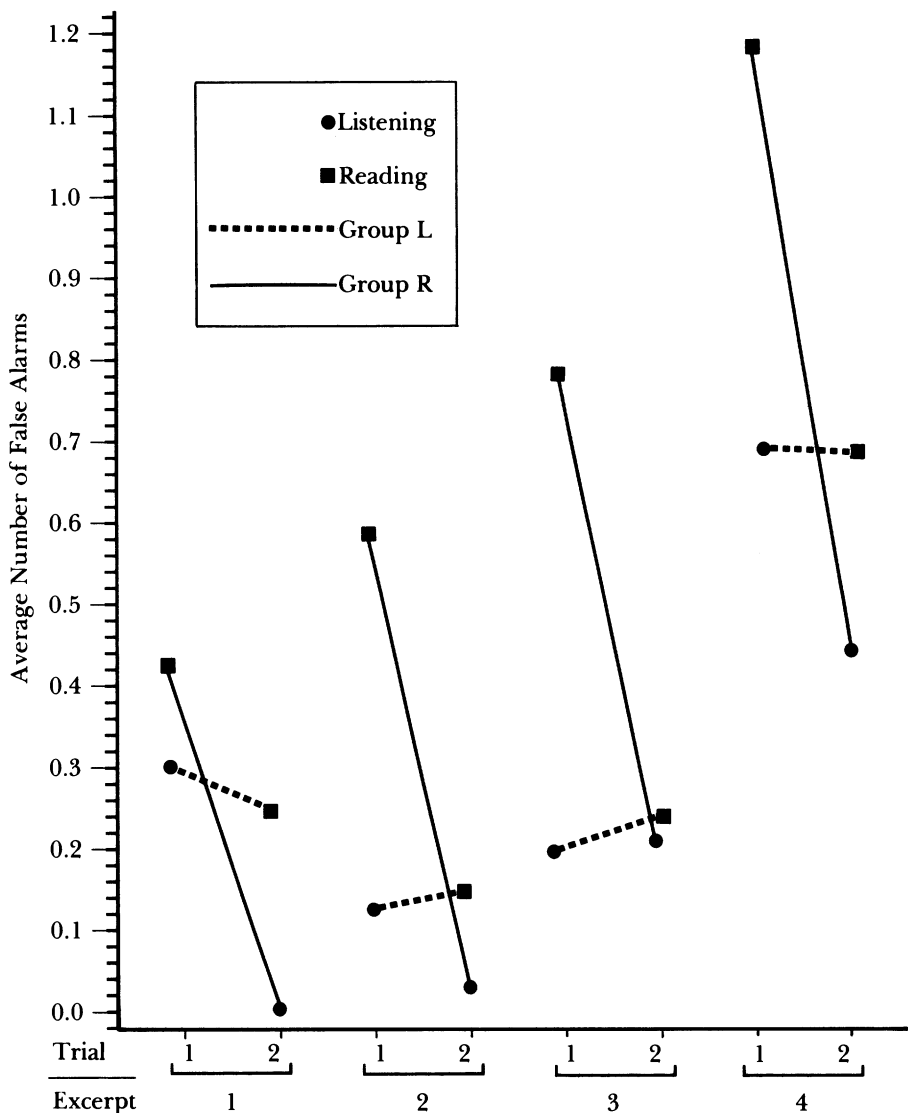
Figure 4. Listening versus sight-reading in music error detection: Average false alarms alarms on each excerpt—Trials 1 and 2

REFERENCES

Benward, B. (1985). *Music in theory and practice, Vol. 1* (3rd ed). Dubuque, IA: Wm. C. Brown.

Brand, M., & Burnsed, F. (1981). Music abilities and experiences as predictors of error detection skill. *Journal of Research in Music Education, 29,* 91–96.

Hansen, L. A. (1955). *A study of the ability of musicians to detect melodic and harmonic errors in the performance of choral music while inspecting the score.* Unpublished doctoral dissertation, University of Kansas, Lawrence.

Kostka, S., & Payne, D. (1984). *Tonal harmony with an introduction to twentieth-century music.* New York: Knopf.

Sidnell, R. G. (1971). Self-instructional drill materials for student conductors. *Journal of Research in Music Education, 19,* 85–91.

Trubitt, A. R., & Hines, R. S. (1979). *Ear training and sight-singing: An integrated approach.* New York: Schirmer Books.

Appendix 1

Excerpts Used to Test Ability to Detect Harmonic Errors

Excerpt	Composer	Title	Key	Meter	Measures
1	Peter Ilyitch Tchaikovsky[a]	"The Sick Doll"	G	2/4	5–16
2	Johann F. Burgmüller[b]	"Ave Maria"	A	3/4	1–8
3	Aram Khachaturian[c]	"Andantino"	C	4/4	2–9
4	Carl Reinecke[d]	"Little Song"	F	4/4	4–10

[a]Zeitlin, P., & Goldberger, D. (Eds.). (1967). *Russian music for the young pianist.* New York: MCA Music.

[b]Burgmüller, J. F. (1903). *Twenty-five easy and progressive studies for the piano, Op. 100.* New York: G. Schirmer.

[c]Robinson, H. (1964). *Basic piano for adults.* Belmont, CA: Wadsworth.

[d]Elson, L. C. (Ed.). (1918). *Modern music and musicians, Vol. 5.* New York: The University Society.

October 10, 1986