Do Music Lessons Enhance IQ?
A Reanalysis of Schellenberg (2004)

Kenneth M. Steele
Appalachian State University

Schellenberg (2004) investigated whether music lessons improved IQ scores in young children in a pretest-posttest design. Six-year-old children were assigned to one of four treatment groups: keyboard instruction, Kodály vocal instruction, drama instruction, or no lessons for 36 weeks during the school year. All groups showed a significant increase in IQ scores over the year. A small, significant difference in gain of 2.7 IQ points was reported when the results of the keyboard and Kodály groups were combined and contrasted against the combined results for drama and no lessons. The combination of groups was not justified theoretically because the Kodály method was presented as being very different from standard musical instruction. Reanalysis of the original uncombined groups produced results that were statistically insignificant and had small effect size values. The hypothesized unique effect of music lessons on IQ scores is still in need of demonstration.

There has been current interest in the relationship between musical experience and intelligence. One reason was the proposal that simply listening to music would cause temporary increases in spatial IQ scores—the so-called Mozart effect (Rauscher, Shaw, & Ky, 1993, 1995). Subsequent work has shown that the original reports were difficult to verify (Chabris, 1999; Steele, Bass & Crook, 1999; Steele, Dalla Bella, et al., 1999) and that arousal or mood differences among testing groups could account for some results (Husain, Thompson, & Schellenberg, 2002; Steele, 2000; Thompson, Schellenberg, & Husain, 2001).

A second topic has been the effect of musical training on cognitive abilities. The argument is that music lessons may have side benefits that transfer to other academic or cognitive abilities and the existence of transfer effects would strengthen the case for the inclusion of music into the classroom curriculum. Costa-Giomi (1999, 2004) found that 3 years of music lessons produced no long-term gain in academic or cognitive performance as measured by standardized tests or by school performance. Other studies have reported positive effects. Rauscher et al. (1997) reported that preschool children provided with individual music instruction scored higher on tests of spatial ability than did children with computer keyboard training or those with no lessons. Rauscher and Zupan (2000) reported that kindergarten children with keyboard training scored higher on a spatial task than those with no training. Bilhartz, Bruhn, and Olson (2000) examined the effect of exposure to a Kindermusik program on subtests of the Stanford-Binet IQ measure. Unfortunately, none of the preceding positive reports used random assignment of subjects to treatment conditions. Lack of random assignment means that other variables may account for the positive effects. For example, Bilhartz et al. found that participation in the Kindermusik program was positively correlated (and thus confounded) with family income. In turn, family income is correlated with IQ (Ceci & Williams, 1997).

Schellenberg (2004) used random assignment of subjects to treatment conditions to examine whether music lessons produce changes in IQ scores in a pretest-posttest design. Six-year-old children were recruited in a

I thank Glenn Schellenberg for promptly and graciously providing the raw data for my additional analyses. I thank Stephen Black for comments on an earlier manuscript.

Correspondence concerning this article should be addressed to Kenneth M. Steele, Department of Psychology, Appalachian State University, Boone, NC 28608; e-mail: steelekm@appstate.edu.

local community newspaper with the offer of free weekly arts lessons. The 144 subjects were assigned to one of four treatment groups. Subjects were administered the WISC–III (Weschler, 1991) during the summer prior to the start of the school year, exposed to the assigned condition for 36 weeks during the school year, and were readministered the WISC–III the following summer. (Other measures were administered but the results of interest were the changes in full-scale IQ.)

The study had four treatment conditions: keyboard instruction, Kodály vocal music instruction, drama instruction, and no lessons. It is important to understand Schellenberg's rationale for inclusion of each group, as that rationale will affect the types of comparisons used in analysis. The keyboard instruction group was exposed to standard musical instruction. Standard musical instruction may involve long periods of focused attention, daily practice, memorization of passages, and the use of fine motor skills. Any or all of these factors may provide effects that plausibly could transfer to other cognitive or academic activities. The keyboard group is presumed to receive the most powerful and direct dose of music instruction.

The Kodály vocal method emphasizes singing, hand signs, clapping, and other sequenced activities in accompaniment with music. One reason for its inclusion in Schellenberg's study was a report by Gardiner, Fox, Knowles, and Jeffrey (1996) that children exposed to this method for 7 months showed improvements in reading and arithmetic. However, Schellenberg stated that "the source of the effect is unclear because Kodály pedagogy differs markedly from standard musical instruction" (Schellenberg, 2004, pp. 511–512). Later he pointed out that the use of both groups made it possible to assess the generality of music instruction effects and "whether nonmusical aspects of Kodály instruction accounted for the effect reported by Gardiner et al." (p. 512). Both statements indicate that the Kodály method was classified as an experience different from ordinary keyboard instruction. Drama instruction was chosen as a comparison activity because it is an artistic auditory activity, like music, and involves focused attention, memorization, and motor activities. Finally, the no-lessons group did not receive formally any of these additional experiences.

Table 1 shows descriptive results from Schellenberg (2004). Table 1 shows mean pretest and posttest WISC–III IQ scores (with standard deviations in parentheses). The results in Table 1 show that all groups show an improvement in IQ scores over the course of the year. Table 1 shows also the number of subjects on which the IQ scores were based. There are differences in attrition, with the most attrition occurring in the keyboard group and the least occurring in the no-lessons group. Subject attrition is an important issue because the pretest and posttest scores are computed from *different subjects*. For example, the posttest score for the keyboard group is missing almost 17% (6/36) of the original subjects. Therefore, the results in Table 1 do not indicate whether the IQ score increases were due to general improvement or subject attrition or both. Schellenberg chose to use "gain" scores (i.e., posttest score *minus* pretest score) in his analysis, which eliminated data from subjects that did not complete the study.

It is important at this point to consider what should be the proper analysis of the results in Table 1. One could argue that each group received a distinct treatment such that the Kodály group is not an intermediate condition between keyboard instruction and drama instruction. In this case, one would expect to see a one-way ANOVA on the gain scores, followed by post hoc contrasts. Another possibility would have been to compute an ANCOVA, using pretest scores as the covariate. The advantage of the analysis of gain scores is the conceptual simplicity of analyzing raw IQ score changes. The advantage of the ANCOVA is statistical power because sample characteristics are taken into account in the analysis.

Alternatively, one could take a continuum approach and argue that the Kodály group lies somewhere between standard music instruction and drama instruction in the amount of musical instruction. In this case, one would expect to see the use of a regression analysis where groups are ordered according to the dose amount of music instruction. Finally, one could perform a repeated-measures ANOVA and analyze for interaction effects

Table 1

Mean Sample Characteristics of Groups in Schellenberg (2004)

Group				
Sample Characteristic	Keyboard	Voice	Drama	No lessons
n before lessons	36	36	36	36
n after lessons	30	32	34	36
IQ before lessons	102.6 (8.8)	103.8 (10.9)	102.6 (13.6)	99.4 (9.7)
IQ after lessons	108.7 (12.5)	111.4 (12.6)	107.7 (13.8)	103.3 (9.9)

Note. Data taken from Table 1 of Schellenberg (2004). Standard deviations are in parentheses.

between time and treatment condition. Huck and McLean (1975) and Jennings (1988) have argued that the analysis of gain scores is preferred over a repeated-measures ANOVA in a pretest-posttest design.

The Schellenberg Analysis

Schellenberg (2004) did not analyze the treatment conditions as either independent experiences or experiences arranged along a continuum. Instead the keyboard and Kodály groups were combined to create a "music" group. The drama and no-lesson groups were combined to create a "control" group. IQ scores improved significantly for all groups, but the improvement for the combined scores of the two music groups (M = 7.0, SD = 8.6) was 2.7 IQ points higher than the combined scores for the other two groups (M = 4.3, SD = 7.3), $t(130)$ = 1.99, $p < .05$, $d = 0.35$.

Why were these groups combined in this fashion? "Similarities between the two music groups and the two control groups justified collapsing the data across groups in order to maximize power in tests of the central hypothesis" (Schellenberg, 2004, p. 512).

Schellenberg concluded that the "findings indicate that music lessons cause small increases in IQ" (Schellenberg, 2004, p. 513). I do not agree and will explain why.

Statistical Significance Was Achieved by Improper Combination of Separate Groups

I described the rationale for the four groups to make the point that there are theoretical questions about any combination of groups. Specifically, Schellenberg made the point that "Kodály pedagogy differs markedly from standard musical training" (Schellenberg, 2004, pp. 511–512) and posed the question of whether Kodály voice training would produce the same effect as keyboard training. The two music groups should have been analyzed separately. Similarly, 36 weeks of drama instruction differs markedly from exposure to nothing.

A conservative and neutral position is that the four groups were exposed to different and independent experiences. In this case, a proper analysis would have been a one-way ANOVA of the gain scores across the treatments, followed by pair-wise contrasts among the treatments. Performance of this analysis produced a statistically insignificant result, $F(3, 128) = 1.35$, $p = .26$, $\eta^2 = .03$. Importantly, no pair-wise contrast was close to producing statistical significance (all $p > .5$, Bonferroni adjustment). An ANCOVA on the posttest IQ scores, using pretest IQ scores as the covariate, produced a similar outcome, $F(3, 127) = 1.68$, $p = .17$, $\eta^2 = .02$.

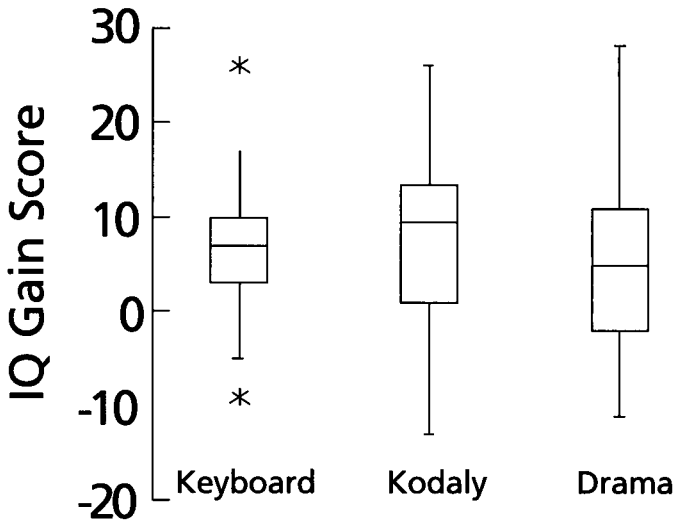
An alternative approach, noted before, was that one could take a continuum approach and argue that the Kodály group lies somewhere between standard music instruction and drama instruction in the amount of musical instruction. Figure 1 shows box plots of the gain scores for the four conditions. The results in Figure 1 do not suggest a dose-response relationship between amount of standard music instruction and IQ score gain. In this reanalysis, the keyboard, Kodály, drama, and no-lessons groups were assigned rank order values from 4 to 1 of music instruction dose and IQ gain scores were regressed on the dose variable. The results suggest there was no strong dose-response relationship between music instruction and IQ gain scores, $F(1, 130) = 3.37$, $p = .07$, $R^2 = .03$.

The ambiguous status of the nature of the Kodály method suggested one additional theory-based post hoc analysis. The Kodály group was removed and an ANOVA was performed on the keyboard, drama, and no-lessons conditions. The results of this analysis were not statistically significant, $F(2, 97) = 1.35$, $p = .26$, $\eta^2 = .03$.

The Central Problem of This and Similar Studies

My analyses suggest that Schellenberg (2004) did not demonstrate that "music lessons enhance IQ." Schellenberg obtained a significant difference by use of

Figure 1. Increase in IQ Score following 36-week exposure to a treatment condition. Treatment condition is listed under each box plot.



a dubious combination of groups. No significant differences were obtained when this specific grouping was eliminated and standard analyses were conducted.

However, for purposes of argument, assume that Schellenberg had obtained clear evidence that keyboard instruction did produce increased IQ scores. What will have been accomplished? I argue: very little. "Music lessons" is a poor choice as an independent variable because the crucial experience itself is in need of clarification. Schellenberg touched on the problem when speculating about the mechanisms by which music lessons could have effects on other cognitive activities. Schellenberg argued that transfer effects might be expected because music lessons involve long periods of focused attention, daily practice, memorization of extended passages, and refinement of fine motor skills. But many activities are likewise, such as gymnastics, playing video games, and attending school. What separates and makes playing music *uniquely different* from these other activities? What is the proper control condition in an experimental test of this hypothesis? If exposure to music lessons is supposed to involve more than a set of school-like experiences then clearly a no-lessons group is not a proper control condition. The issue of the appropriate control conditions is a central problem in these studies.

REFERENCES

Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (2000). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology, 20*, 615–636.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52*, 1051–1058.

Chabris, C. F. (1999). Prelude or requiem for the "Mozart effect"? *Nature, 400*, 826–827.

Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education, 47*, 198–212.

Costa-Giomi, E. (2004). Effects of three years of piano instruction on children's academic achievement, school performance, and self-esteem. *Psychology of Music, 32*, 139–152.

Gardiner, M. F., Fox, A., Knowles, F., & Jeffrey, D. (1996). Learning improved by arts training, *Nature, 381*, 284.

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze data from a pretest-posttest design: a potentially confusing task. *Psychological Bulletin, 82*, 511–518.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception, 20*, 151–171.

Jennings, E. (1988). Models for pretest-posttest data: Repeated measures ANOVA revisited. *Journal of Educational Statistics, 13*, 273–280.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365*, 611.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters, 185*, 44–47.

Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., & Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research, 19*, 2–8.

Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly, 15*, 215–228.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*, 511–514.

Steele, K. M. (2000). Arousal and mood factors in the "Mozart effect." *Perceptual and Motor Skills, 91*, 188–190.

Steele, K. M., Bass, K. E., & Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science, 10*, 366–369.

Steele, K. M., Dalla Bella, S., Peretz, I., Dunlop, T., Dawe, L. A., Humphrey, G. K., et al. (1999). Prelude or requiem for the "Mozart effect"? *Nature, 400*, 827.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science, 12*, 248–251.

Weschler, D. (1991). *Weschler intelligence scale for children* (3rd ed.). San Antonio, TX: Psychological Corp.