Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 2, 417–425

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000044

How Does Stroop Interference Change With Practice? A Reappraisal From the Musical Stroop Paradigm

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
University of Bourgogne

Most earlier studies investigating the evolution of the Stroop effect with the amount of reading practice have reported data consistent with an inverted U-shaped curve, whereby the Stroop effect appears early during reading acquisition, reaches a peak after 2 or 3 years of practice, and then continuously decreases until adulthood. The downward component of the curve suggests that skilled performers would be able to control their performance better than less-skilled performers. However, in these studies, the level of reading practice entirely coincides with age due to obvious practical and ethical constraints, and it is possible that the observed reduction in the Stroop interference is due to a growing ability of older children to inhibit nonrelevant information. In the present study, word reading, as source of interference, was replaced by note naming in musicians. The major advantage is that musical training can be easily decoupled from age. In 2 experiments exploiting the musical Stroop paradigm (Grégoire, Perruchet, & Poulin-Charronnat, 2013), we observed an early appearance of the interference effect, as reported for the color–word and picture–word Stroop tasks, but we did not replicate the inverted U-shaped curve. Experiment 2 revealed a linear and positive relation between the amplitude of the musical Stroop effect and the amount of musical practice across 5 years of musical training. These results suggest that reading practice in itself does not lead to increased control over reading and that the usual pattern of results is most likely due to the strong correlation between age and reading practice.

Keywords: Stroop interference, musical Stroop effect, automatisms, practice

Since the original study of Stroop (1935), in which an incongruent color word was shown to slow down the naming of the ink color in which it was printed, skilled reading has been considered as irrepressible (see MacLeod, 1991; MacLeod & MacDonald, 2000, for reviews). This irrepressibility is obviously dependent on learning literacy, but, surprisingly, only a small proportion of the wide body of literature dealing with the Stroop effect raises the question of how Stroop interference evolves with the amount of reading practice.

An Inverted U-Shaped Function

Comalli, Wapner, and Werner (1962) were the first to investigate this issue. Participants were divided into several age groups beginning with 7-year-old children. Stroop interference was measured as the difference between the time needed to name the color of the ink of incongruent color words (e.g., the word BLUE printed in *green* ink) and the time required to name the color of patches. The greatest Stroop interference was found for the younger children; then the amount of interference decreased up to 17–19 years and remained constant up to 44 years.

Using the same methodology as Comalli et al. (1962), Rand, Wapner, Werner, and McFarland (1963) tested children from 6 to 17 years old. Unlike in Comalli et al., Stroop interference (as assessed from their Table 10) increased from 6 to 9 years old. However, Rand et al. replicated the decrease in Stroop interference in older children, with a Stroop effect decreasing in amplitude from the age of 9. More recently, Peru, Faccioli, and Tassinari (2006) observed a similar pattern of results in a normative developmental study in Italian, in which children from 3 to 10 years old were tested. Stroop interference increased from 3 to 7 years old[1] and then decreased. Dash and Dash (1982), Schadler and Thissen (1981), and Schiller (1966) reported the very same inverted U-shaped evolution of Stroop interference, with a peak being located after 2 or 3 years of reading practice. It is worth noting that Schiller observed this pattern even though interference was measured as a ratio between response times (RTs) for incongruent and neutral stimuli in order to control for shorter RTs of older children.

Rosinski, Golinkoff, and Kukish (1975) exploited the picture–word variant of the Stroop task, in which interference on naming a picture is caused by an incongruent word superimposed on line drawing. They tested three groups of participants (Grade 2, Grade 4, and adults) and observed that picture–word interference decreased gradually with age.

This article was published Online First July 28, 2014.

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat, Laboratoire d'Etude de l'Apprentissage et du Développement–National Center for Scientific Research (LEAD-CNRS UMR5022), Université de Bourgogne.

This research was supported by Grant ANR-09-JCJC-0129 and the Grant Conseil Régional de Bourgogne No. 2012-9201AAO05S01353.

Correspondence concerning this article should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

[1] Some studies vary chronological age, while others vary school level. For the sake of comparison, 6-year-olds were in first grade and so on.

To sum up, all these studies concur in that they showed a linear decrease in Stroop interference from 9-year-olds onwards, but they differed in findings for younger children. Some studies observed a decrease from the outset, while others reported an inverted U-shaped curve, with an initial increase in interference. Most of this apparent discrepancy certainly reflects the fact that the authors began their observations more or less early during development. Presumably, some investigators have not observed a first rising component simply because the difficulty of carrying out the Stroop task in very young children has dissuaded them from including these children in their observations.[2] The upward component of the curve, whether actually measured or not, is a logical requirement, because insofar as the effect exists in adults, it must emerge somewhere during reading acquisition. Consequently, the fact that the relation between Stroop interference and the amount of reading practice follows an inverted U-shaped function can reasonably be taken as a robust phenomenon.

There remains a point of disagreement, nevertheless, regarding where interference begins to decrease. It is possible that, for a given age group, the level of reading proficiency differs as a function of several factors, as suggested by the findings of Armengol (2002). Armengol tested first to sixth graders in order to provide normative developmental data on the Stroop effect in Spanish, using a procedure very similar to Comalli et al. (1962), except that she added a group of younger children. For children coming from private school, Armengol observed results very close to those of Comalli et al., namely, a gradual decrease in Stroop interference from Grade 1 to Grade 6. For the children coming from public school, she replicated the inverted U-shaped curve obtained in many earlier studies, with an increase in Stroop interference from Grade 1 to Grade 2, followed by a gradual decrease. She explained the discrepancy between private and public schools for the first graders by a better reading proficiency at Grade 1 in private school.

Two Contrasted Interpretations

The downward component of the inverted U-shaped curve is somewhat surprising. If Stroop interference reflected the automaticity of reading, and if the level of automaticity were a growing function of the amount of practice, then the capacity of reading to interfere with another less-practiced activity should appear gradually and should increase as reading skills improve. The observed pattern of performance suggests that the irrepressibility of automatic behavior does not grow as a monotonic function of practice. Supporting this view, Logan (1985) wrote that “skilled performers are usually able to control their performance better than unskilled performers, even though their performance is likely to be more automatic,” (p. 379). Evidence for a better control has been observed by Logan (1982), who demonstrated that skilled typists were able to inhibit high-speed typing when detecting an error or an overt signal to stop. Tzelgov, Henik, and Leiser (1990), by testing bilinguals, observed that Stroop interference is controllable and that language proficiency is a precondition for such a control. Participants were able to reduce Stroop interference as a function of their expectancies in their native language but not in their second language. It is thus possible that the decrease in Stroop interference observed from childhood to adulthood reflects a genuine property of automatisms, whereby the possibility of cognitive control would increase with practice.

However, there is an alternative interpretation. In all the studies described thus far, the level of reading practice coincided entirely with age, and it is possible that at least a part of the observed changes in Stroop interference was due to general factors evolving with age. Comalli et al. (1962) suggested that the amount of interference could be a positive function of the amount of practice but that this effect would be overshadowed by the age-related variations in the ability to inhibit nonrelevant information. Age-related variations in cognitive control would explain the decrease in Stroop interference observed from childhood to adulthood, because it is commonly acknowledged that this ability grows during the relevant period of time (e.g., Bédard et al., 2002; Klenberg, Korkman, & Lahti Nuuttila, 2001).[3]

Controlling for the Influence of Age

In principle, the solution for disentangling the competing views is straightforward: Because age appears as a confounding variable, the effect of reading practice should be assessed while keeping age constant. This is not possible, due to obvious practical and ethical constraints. A second-best solution consists in exploiting the natural variation in reading-skill abilities for a given age group. Fast readers may be considered as having more extensive reading practice—or at least as having benefitted more from practice—than slow readers. If the amount of Stroop interference were a positive function of practice, then fast readers should present greater interference than slow readers. Overall, empirical data confirm this prediction. In Martin (1978), adult participants were divided into two groups of fast and slow readers according to their mean reading speed of random word passages. Martin observed that fast readers showed significantly larger Stroop interference than slow readers. Likewise, in Stanovich, Cunningham, and West (1981), first graders were ordered according to their reading ability by their teachers at the end of the school year. The first 12 children on this scale consistently showed greater interference than the last 12 children, although the differences did not reach significance. Moreover, correlational analyses showed a slight tendency for children who read the words faster and more accurately in several tests of reading to exhibit greater interference.

More recently, Braet, Noppe, Wagemans, and Op de Beeck (2011) investigated the relationship between Stroop interference and reading skills in the context of second-language acquisition in adults. First, they observed greater Stroop interference for native language (L1), for which reading is expected to be more automatic, than for second language (L2). Second, they found significant relations between Stroop interference and several measures of

[2] Or at least in their report—Comalli et al. (1962) noted in their footnote 4 that they also tested 5- and 6-year-old children, but they did not report the data.

[3] This interpretation fits well with the observation that the amount of Stroop interference increases from young adults to older people (e.g., Comalli et al., 1962). Indeed, cognitive control could be less efficient in older than in young adults (Diamond, 2013). However, the increase of the Stroop effect is seemingly dependent on the version of the Stroop task (e.g., Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010) and details of measurement of Stroop interference (e.g., Verhaeghen & De Meersman, 1998). This weakens any arguments based on the literature on aging.

L2-reading skills, with greater interference linked to higher skills in L2.

All these studies are consistent with the idea that, when the influence of age is removed, the amount of Stroop interference could be positively related to reading practice. However, the strength of the evidence is weakened by the fact that individual variations in reading skills for a given age group may depend on many cognitive and social factors in addition to the amount of reading practice. Moreover, the conclusions issued from studies comparing good and poor readers in the normal range are not confirmed by studies investigating reading disorders. In keeping with the data collected in the general population, people with reading disorders would be expected to exhibit a lower level of Stroop interference than skilled readers. Paradoxically, studies exploiting this approach have consistently observed the reverse (Everatt, Warner, Miles, & Thomson, 1997; Faccioli, Peru, Rubini, & Tassinari, 2008; Kapoula et al., 2010; Protopapas, Archonti, & Skaloumbakas, 2007). These studies do not rule out the idea that reading ability is positively related to Stroop interference when the influence of age is removed, because, as suggested by most authors of these studies, the stronger interference observed in dyslexic participants might not come from a potential impaired automaticity of reading in this population but rather from a reduced capacity to inhibit word reading in comparison with skilled readers. An unavoidable conclusion, however, is that the current data do not provide a clear and unambiguous picture on the relation between reading experience and Stroop interference.

The Present Study

Given that reading offers quite limited opportunity to investigate the effects of practice, we proposed to investigate the question of how Stroop interference evolves with the amount of training through a paradigm that no longer relies on reading as the interfering process. To our knowledge, the only similar attempt was reported by MacLeod and Dunbar (1988). The authors used a Stroop-like task in which arbitrary associations were created across repeated experimental sessions to ensure a better control on practice. Participants were trained to associate a color name to a shape during 20 days. In the experiment of concern here, participants were tested on different tasks on Day 1, Day 5, and Day 20. The authors observed virtually no interference of shape on color naming on Day 1, while the interference became significant on Day 5 and continued to increase on Day 20. To sum up, interference on color naming increased with shape-naming training. This result is consistent with the idea that Stroop interference would be positively related to training, but the question remains of whether a conclusion issued from a laboratory task, even extended along 20 sessions, can be generalized to the abilities resulting from the huge amount of practice common in everyday life.

In the present study, word reading, as source of interference, was replaced by note naming in musicians. In earlier works, we have shown the existence of a musical Stroop effect (MSE) in adults (Grégoire et al., 2013, 2014a). The stimuli comprised a staff with a note in various positions. A name of a note was printed inside the note. In the congruent condition, the note name was congruent with the note position, while in the incongruent condition, name and position were incongruent. Musically trained participants were asked to read aloud the printed name of notes. The *MSE* denotes the fact that reading note name was slowed down in the incongruent condition with regard to the congruent condition.

The musical Stroop paradigm shares most advantages of the classical color–word version to investigate automatisms. The most obvious is that note naming is an activity in which expert musicians are intensively engaged over years, attaining a level of practice that is out of reach of laboratory settings. Moreover, note naming shares with word reading the nice property of being a component of more complex activities, the automaticity of which is required to ensure the successful expression of the whole behavior. The automaticity of note naming is necessary to allocate musicians' attention to higher integrative processes devoted to analyzing chords and melodic lines of the musical work and to ensuring motor control for instrument playing,[4] as the automaticity of identification of single words is needed to free the mind to deal with higher order aspects of reading such as comprehension and metacognitive functions. However, as an experimental means of investigation, the musical Stroop paradigm avoids the shortcomings pointed out earlier. In particular, for most children, musical training begins after they start to learn to read, making it possible to capture the effects of the early phase of practice while avoiding problems linked to testing very young children. Moreover, the interindividual variations in the age to enter a music school make it possible to decouple age and amount of musical practice, provided there is an appropriate selection of participants.

In Experiment 1, the development of the MSE was addressed by testing children in their first, second, and third year of music training. In Experiment 2, the amount of training was extended to 5 years. In both experiments, the main issue was addressed through planned orthogonal contrasts testing for polynomial functions of different orders. If interference is a direct function of experience and if the downward component of the inverted U-shaped curve observed in standard Stroop tasks is due to the strong correlation between experience and age, then the evolution of the MSE with years of practice should follow a positive linear trend. If the inverted U-shaped curve is a genuine property of automatism formation, then the MSE should also follow a quadratic function.[5]

Experiment 1

Method

Participants. Forty-two children from two music schools (the Conservatoire à Rayonnement Régional of Dijon and the École Municipale de Musique of Longvic) took part in the experiment. These children were divided into three groups (for each, $n = 14$; Table 1) according to their year of musical training. An additional group was made up of musically untrained children ($N = 14$; Table 1). The children's vision was normal or corrected to normal, and written parental consent was obtained for each child.

[4] It might be argued that there is no need for note naming to achieve the automatization of the motor programs involved in the practice of a musical instrument. This and other concerns about the musical Stroop paradigm were addressed in Grégoire, Perruchet, and Poulin-Charronnat (2014b).

[5] As noted previously, a (decreasing) linear trend was observed in Stroop studies that did not involve young enough children. This outcome cannot occur in the present study, the sample of which comprised beginning musicians and even nonmusicians to ensure the initial emergence of the effect was not missed.

Table 1
Participant Characteristics and Error Percentages for the Note-Naming Ability Test and Experimental Conditions as a Function of Musical Training for Experiments 1 and 2

Variable	Nonmusicians	Year of musical training 1st	2nd	3rd	4th	5th
Experiment 1						
Age (years)	7.70 (0.46)	7.51 (0.27)	8.65 (0.47)	9.36 (0.72)		
n	6 females/8 males	6 females/8 males	7 females/7 males	10 females/4 males		
Ability test	None	5.97 (16.91)	3.59 (3.84)	0.63 (1.33)		
Congruent	0 (0)	0.87 (1.61)	0.33 (1.24)	0 (0)		
Incongruent	0 (0)	1.53 (1.68)	1.44 (4.24)	0.34 (0.88)		
Experiment 2						
Age (years)	9.76 (0.21)	9.56 (0.74)	9.09 (0.84)	10.32 (0.56)	10.58 (0.43)	10.97 (0.65)
n	6 females/6 males	5 females/7 males	8 females/4 males	9 females/3 males	9 females/3 males	8 females/4 males
Ability test	None	1.37 (2.29)	1.93 (5.36)	1.17 (2.12)	1.87 (2.74)	0 (0)
Congruent	0 (0)	0.40 (1.37)	0.79 (1.16)	0 (0)	0.39 (0.91)	0.25 (0.87)
Incongruent	0 (0)	1.21 (2.17)	1.64 (2.48)	0.59 (1.07)	1.55 (2.01)	1.38 (3.52)

Note. Standard deviations are in parentheses.

Material. The experimental material was composed of stimuli consisting of a treble staff with a note picture, which could appear on each of the eight possible positions going from C4 to C5. The name of a note was written inside the note picture. For the congruent condition, the note name was congruent with the note position on the staff (Figure 1a), whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names (e.g., when the note was DO, the written name was LA, SI, RE, MI, FA, or SOL; Figure 1b).

An additional test of note-naming abilities was designed to be closer to those involved in musical practice (Figure 1c). As for the congruent and the incongruent conditions, the note pictures from C4 to C5 were presented to the children, who had to name them.

To prevent the iconic memory of the staff from influencing the processing of the following note, we randomly displayed the stimuli at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. Stimuli were printed in black over a white background on a computer screen. Note names appeared in standard uppercase 14-point Arial font. The treble staff was 7.7 cm wide by 5.1 cm high. The stimuli delivery and response recording were controlled by PsyScope (Cohen, MacWhinney, Flatt, and Provost, 1993) running on an Apple iMac computer (Apple Inc., Cupertino, CA).

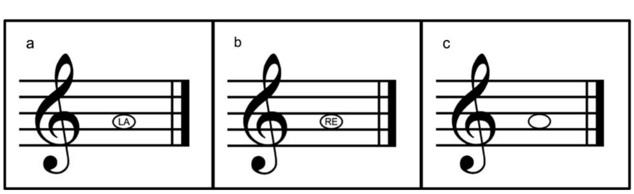

Figure 1. Examples of the different conditions used in Experiments 1 and 2: (a) congruent condition; (b) incongruent condition; and (c) note-naming ability test. Note that in musical notation in France (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Procedure. The musically trained children had to perform a word-reading task, in which they had to read aloud the printed word while ignoring the note picture. There were two mixed conditions: congruent (Figure 1a) and incongruent (Figure 1b). For each condition, the stimuli appeared six times on each of the eight locations, resulting in 48 trials per condition with a total of 96 trials (48×2). On each trial, a fixation cross was displayed for 1,000 ms at the center of the screen before the occurrence of the stimulus, which stayed on the screen until participant's response. The interval between the response and the next trial was 1,000 ms (Figure 2). The trials were pseudo-randomly ordered for each child, excluding immediate repetitions of note locations or note names. They were displayed as four blocks of 24 trials each with a self-paced break between blocks.

The experimental session was immediately followed by an additional note-naming ability test, in which children had to name notes (Figure 1c). The ability test included 48 trials, the notes occurring 6 times each. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of notes. They were displayed as two blocks of 24 trials each with a

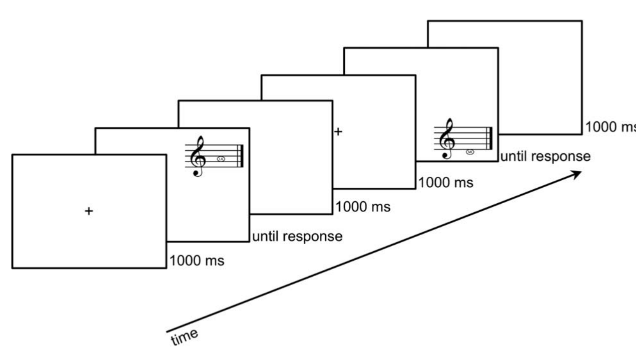

Figure 2. Sequence of events in the musical Stroop paradigm. Each trial began with a fixation cross of 1,000 ms, followed by an experimental stimulus, which could appear at one of the four quadrants of the screen. The stimulus stayed on the screen until the participant's response. The interval between the response and the next trial lasted 1,000 ms.

self-paced break between blocks. The control musically untrained children only performed the word-reading task.

Children were encouraged to respond as fast and as accurately as possible throughout the session. The RTs were recorded by a voice key. During the session, the experimenter noted error responses and voice-key dysfunctions. After the experiment, the children filled out a questionnaire about their musical training.

Results

Ability test (musically trained children only). Voice-key dysfunctions led to the exclusion of 13.89% of the data. The proportion of errors in the note-naming ability test was quite low ($M = 3.40\%$, $SD = 10.04$). Errors decreased with the year of musical training (Table 1). However, a one-way ANOVA on note-naming errors performed with year of musical training (first, second, and third year) as a between-subjects variable showed no significant effect, $F(2, 39) = 1.00$, $p = .378$.

RTs for correct responses beyond three standard deviations of the mean (1.24%) were removed. The remaining data are presented in Figure 3, top panel. The mean RTs decreased with the year of musical training. A one-way analysis of variance (ANOVA) revealed a significant main effect, $F(2, 39) = 5.61$, $p = .007$, $\eta_p^2 = .223$, with a significant linear trend, $p = .003$.

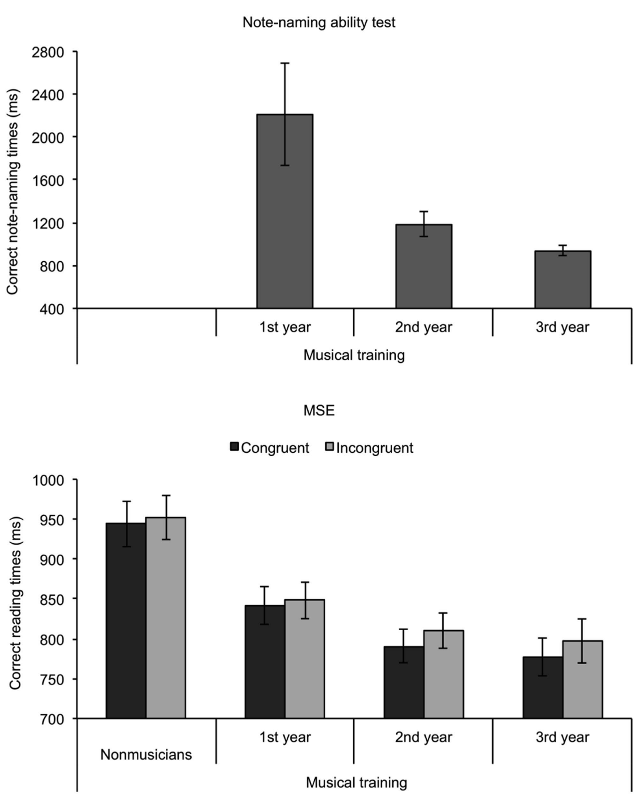

Figure 3. Correct note-naming times for the note-naming ability test (top panel), and correct reading times as a function of congruency (congruent vs. incongruent; bottom panel) in Experiment 1. Both note-naming times and reading times are presented as a function of the year of musical training (from nonmusicians to third year). Error bars indicate standard errors. MSE = musical Stroop effect.

Note that there was no evidence for a speed–accuracy trade-off in the note-naming ability test, as shown by the positive correlation between errors and RTs, $r(40) = .868$, $p < .001$.

Word reading (MSE). Voice-key dysfunctions led to the exclusion of 9.90% and 7.81% of the data for the musically trained and untrained children, respectively. A small proportion of errors was observed for the musically trained children (0.75%, $SD = 2.09$), while the nonmusicians made no errors of reading (Table 1). A repeated-measures mixed ANOVA on reading errors performed with congruency (congruent, incongruent) as a within-subject variable and year of musical training (nonmusicians, first, second, third) as a between-subjects variable showed a significant main effect of congruency, $F(1, 52) = 4.60$, $p = .037$, $\eta_p^2 = .081$, with more errors in the incongruent condition than in the congruent condition. There was no effect of the year of musical training, $F(3, 52) = 1.95$, $p = .134$, no Congruency × Year of Musical Training interaction, $F(3, 52) = 0.91$, $p = .440$, and neither the linear nor the quadratic components of this interaction reached significance, $ps > .10$.

RTs for correct responses beyond three standard deviations of the mean were removed, 0.67% and 1.29% for the musically trained and untrained children, respectively. The remaining data are shown in Figure 3, bottom panel. An ANOVA on RTs carried out as for reading errors gave a significant effect of year of musical training, $F(3, 52) = 8.63$, $p < .001$, $\eta_p^2 = .332$, with a decrease in mean RTs across years. The RTs were significantly longer for the incongruent condition than for the congruent condition, $F(1, 52) = 20.29$, $p < .001$, $\eta_p^2 = .281$. The Congruency × Year of Musical Training interaction was not significant, $F(3, 52) = 1.47$, $p = .234$. However, the difference between incongruent and congruent trials increased monotonically from the nonmusicians to the third year of musical training. Accordingly, planned contrasts[6] revealed that the linear component of the interaction approached the conventional cutoff for significance, $F(1, 52) = 3.36$, $p = .073$, $\eta_p^2 = .061$, whereas there was no statistical evidence for a quadratic trend, $F(1, 52) = 0.04$, $p = .838$.

Correlational analyses. The amplitude of the MSE was related to age, although the correlation was not significant, $r(40) = .211$, $p = .182$.[7] It is worth stressing that this correlation was positive, whereas the downward component of the inverted U-shaped relation described in the introduction suggest a negative correlation between the amount of Stroop interference and age. A hypothesis is that the observed correlation could be mediated by the level of musical practice. Indeed, on average, older children had more musical practice. Partial correlations were carried out to address this issue. When controlling for year of musical training, the partial correlation between MSE and age dropped to zero, $r(39) = .020$, $p = .902$, suggesting that performance was independent from age.

The amplitude of the MSE was also linked to the naming times in the note-naming ability test, $r(40) = -.300$, a correlation that just failed to reach significance, $p = .053$. When the level of

[6] Planned comparisons are sometimes claimed to be inappropriate when the overall F is not significant. However, a number of experts in statistics have convincingly argued that this general advice was unwarranted (e.g., Howell, 2010; Ryan, 1959; Wilcox, 1987), especially when planned contrasts are orthogonal as here.

[7] All the correlations were calculated only on the groups of musicians because certain correlations involved the note-naming ability test, which nonmusicians did not undergo.

practice was controlled, this correlation dropped but remained substantial, $r(39) = -.222$, $p = .163$. This suggests that the test of note-naming ability could capture some relevant individual differences, which operate in addition to the level of musical education. Note that, unsurprisingly, the naming times in the note-naming ability test significantly correlated with the year of musical training but to an extent that left room to independent sources of influence, $r(40) = -.446$, $p = .003$.

Discussion

The results of Experiment 1 confirm in children the results previously obtained in adults (Grégoire et al., 2013) regarding the existence of an MSE: On average, incongruent trials generated more errors and longer reading times than congruent trials. More important, our study provides information about the evolution of the MSE with musical practice. Although the linear trend was only marginally significant, there was a numerical trend for a continuous increase in the size of the MSE across practice without any evidence for a downward component.

A conservative conclusion is that this experiment does not replicate the inverted U-shaped curve that is commonly observed with the color–word Stroop task. This suggests that the downward part of the curve could be due to the fact that reading acquisition occurs during a life period where cognitive control increases. However, another possibility is that musical practice was not extensive enough to allow the observation of a decrement in the amplitude of the MSE. As outlined in the introduction, studies carried out with the standard Stroop paradigm showed that the exact location of the peak of the inverted U-shaped curve varied from one study to another, and even if this location was precisely defined, generalization to the musical Stroop paradigm would be hazardous.

In Experiment 2, we sought to address this concern, by exploring the effect of practice over 5 years of musical training instead of 3. Even though the effect of age was found to be negligible in Experiment 1, a special effort was made in Experiment 2 to kept age as constant as possible across the five levels of musical training.

Experiment 2

Method

Participants. Sixty children from several music schools of Dijon and its suburbs took part in the experiment. The children were divided into five groups (in each, $n = 12$; Table 1) according to their year of musical training. A group made up of musically untrained children was added ($N = 12$; Table 1). None of the children participated in Experiment 1. The children's vision was normal or corrected to normal, and written parental consent was obtained for each child.

Material and procedure. The material and the procedure were exactly the same as those in Experiment 1.

Results

Ability test (musically trained children only). Voice-key dysfunctions led to the exclusion of 14.06% of the data. The proportion of errors in the note-naming ability test was quite low ($M = 1.27\%$, $SD = 3.01$; see Table 1). A one-way ANOVA on note-naming errors performed with year of musical training (from the first to the fifth) as a between-subjects variable showed no significant effect, $F(4, 55) = 0.79$, $p = .534$.

RTs for correct responses beyond three standard deviations of the mean (1.06%) were removed. The remaining data are shown in Figure 4, top panel. The mean RTs decreased with year of musical training. A one-way ANOVA revealed a significant main effect, $F(4, 55) = 10.09$, $p < .001$, $\eta_p^2 = .423$, with a significant linear trend, $p < .001$.

Note that there was no evidence for a speed–accuracy trade-off in the note-naming ability test, as shown by the positive correlation between errors and RTs, $r(58) = .229$, $p = .078$.

Word reading (MSE). Voice-key dysfunctions led to the exclusion of 9.98% and 6.34% of the data for the musically trained and untrained children, respectively. A small proportion of errors was observed for the musically trained children (0.82%, $SD = 1.84$), while the nonmusicians made no errors of reading (Table 1). A repeated-measures mixed ANOVA on reading errors performed with congruency (congruent, incongruent) as a within-subject variable and year of musical training (nonmusicians, first, second, third, fourth, and fifth) as a between-subjects variable showed a main effect of congruency, $F(1, 55) = 7.85$, $p = .007$, $\eta_p^2 = .106$, with more errors in the incongruent condition than in the congruent condition. There was no effect of year of musical training, $F(5, 66) = 1.67$, $p = .153$, no Congruency $\times$ Year of Musical Training interaction, $F(5, 66) = 0.42$, $p = .836$, and neither the linear nor the quadratic components of this interaction reached significance, $ps > .10$.

RTs for correct responses beyond three standard deviations of the mean were removed, 0.66% and 0.56% for the musically trained and untrained children, respectively. The remaining data are shown in Figure 4, bottom panel. An ANOVA on RTs carried out as for errors revealed a main effect of the year of musical training, $F(5, 66) = 2.70$, $p = .028$, $\eta_p^2 = .170$, with a significant decreasing linear trend, $p = .001$. There was also an effect of congruency, $F(1, 66) = 13.97$, $p < .001$, $\eta_p^2 = .175$, which was qualified by a significant Congruency $\times$ Year of Musical Training interaction, $F(5, 66) = 3.70$, $p = .005$, $\eta_p^2 = .218$. The decomposition of this interaction revealed a linear increase of the MSE with years of musical training, $F(1, 66) = 11.57$, $p = .001$, $\eta_p^2 = .149$. By contrast, the quadratic component of the interaction was not significant, $F(1, 66) = 0.072$, $p = .789$.

Correlational analyses. Although the age span of the children (1.5-year between the most extreme groups) was smaller than in Experiment 1, the amplitude of the MSE was again positively related to age, $r(58) = .249$, $p = .055$.[7] However, as in Experiment 1, when controlling for year of musical training, the partial correlation dropped to zero, $r(57) = -.054$, $p = .686$, confirming that the MSE was independent from age.

The amplitude of the MSE was also significantly related to the naming times in the note-naming ability test, $r(58) = -.348$, $p = .006$. When the level of practice was controlled, this correlation dropped substantially, $r(57) = -.105$, $p = .427$. This suggests that the performance in the test of note-naming ability is more related to the level of musical training than in Experiment 1. Indeed, the naming times in the note-naming ability test correlated more strongly with the level of musical practice in Experiment 2, $r(58) = -.630$, $p < .001$, than in Experiment 1.

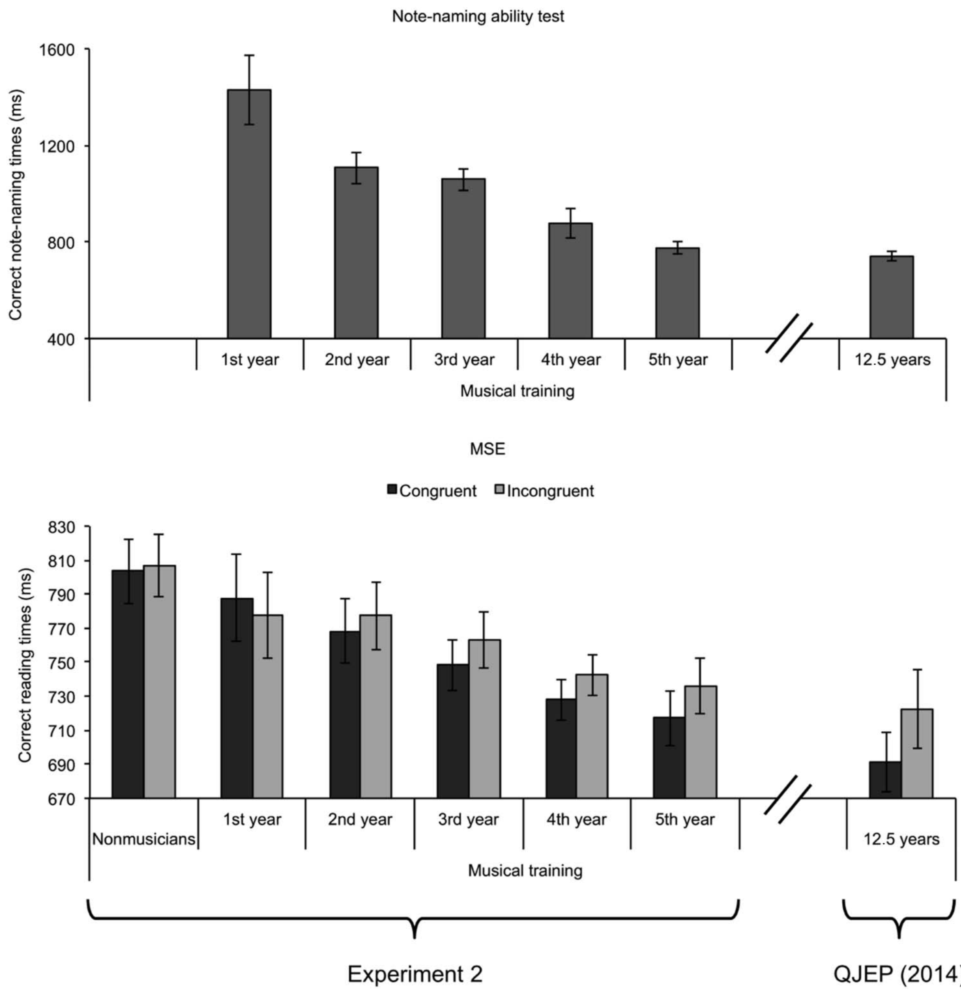

Figure 4. Correct note-naming times for the note-naming ability test (top panel), and correct reading times as a function of congruency (congruent vs. incongruent; bottom panel) in Experiment 2. Both note-naming times and reading times are presented as a function of the year of musical training (from nonmusicians to fifth year). The data collected by Grégoire et al. (2014a; *Quarterly Journal of Experimental Psychology* [QJEP]) from young adult musicians (12.5 years of musical training) have been added for comparison (see General Discussion section). Error bars indicate standard errors. MSE = musical Stroop effect.

Discussion

The results of Experiment 2 confirmed and strengthened the main conclusions of Experiment 1, by showing a significant increase of the MSE over 5 years of musical training. However, a significant MSE appeared as soon as the second year of musical training in Experiment 1, whereas the effect became significant only from the third year in Experiment 2 (see Figures 3 and 4). This difference may be related to the sampling of participants. In Experiment 1, most musically trained children came from the music conservatory of Dijon, whose objective is to train professional musicians. In Experiment 2, the objective of reducing age differences between the five groups of musically trained children led to the recruitment being extended to a larger number of music schools from the suburban area of Dijon, making musical achievement more heterogeneous and certainly of a lower standard. This difference and its consequence are reminiscent of the study by Armengol (2002), who observed a 1-year difference in the appearance of the classical Stroop effect between students from private and public schools.

General Discussion

Summary of the Main Results

In the present study, word reading, as the usual source of interference for the Stroop effect, was replaced by note naming in musicians. Grégoire et al. (2013, 2014a) demonstrated the existence of an MSE in adults, whereby reading a note name written inside a note on a staff was slower when note name and note location were incongruent than when note name and note location were congruent. Like word reading, note naming is an activity that is intensively practiced over years by expert musicians but, in contrast to word reading, the level of practice on note naming does not necessarily coincide with the different grades of elementary school. In particular, the two reported experiments did not involve the very young children who are necessarily included in the study population whenever one intends to track the early emergence of the standard Stroop effect. Moreover, in Experiment 2, in which all children were over 9 years of age, the age difference between the extreme groups (first year vs. fifth year of musical training) was reduced to 1.5 year, which is obviously impossible when reading is the interfering process.

In these conditions, we also observed an early appearance of the interference effect, as reported for the color–word Stroop task, but we did not replicate the inverted U-shaped curve. Both Experiments 1 and 2 revealed a linear and positive relation between the amplitude of the MSE as assessed by RTs and the amount of musical practice, which reached significance when 5 years of musical training were considered in Experiment 2. In both experiments, the MSE as assessed by the proportion of errors also increased with practice (the slopes of the regression lines, although not significantly different from zero, were positive), hence running against the possibility of a speed accuracy trade-off. Note that this gradual increase of the MSE is all the more noticeable as it occurred within a context where the overall reading times (and error proportions) substantially decreased. This means that the increase of the MSE cannot be a mechanical consequence of scale effects whereby larger values would offer more space for an effect to emerge.

Could Additional Training Reveal an Inverted U-Shaped Relation?

Are 5 years of training sufficient to definitely rule out the existence of a downward component to the relation between the MSE and musical practice? It is clear that all Stroop studies described previously reported a decrement of the Stroop interference long before 5 years of reading practice. Moreover, the level of automaticity resulting from a same amount of practice obviously depends on the processes involved in the tasks. For instance, Logan and Klapp (1991) showed that genuine properties of automaticity could emerge after less than 15 min of training in an alphabet–arithmetic task when there were just six problems, whereas in a more complex situation, in which 40 alphabet–arithmetic facts were used, 12 sessions were necessary to produce the same level of automaticity. For our concern, the musical notation looks much simpler than the writing system, providing additional evidence against the possibility that a decrement of the MSE could appear after 5 years of musical training. On the other hand, it could be argued that using the number of years of practice as a benchmark is misguiding, because even for musicians, note naming is certainly a less pervasive activity than reading. Undoubtedly, empirical evidence would be necessary to definitely resolve the issue.

Exploring the evolution of the MSE over a time period much longer than 5 years, while keeping age constant as in the present study, raises a practical difficulty, simply because the number of people entering a music school well after age 10 or 11 (the age range of participants in Experiment 2) is limited. However, keeping age constant may have far less importance at this stage. Indeed, the capacity of cognitive control of 10- to 11-year-old children is certainly close to their adult achievement, even though further improvement cannot be excluded (e.g., Diamond, 2013). For exploratory purposes, we compared the data from Experiment 2 to data collected in young adult musicians whose the mean amount of musical practice was 12.5 years (*SD* = 4.51; Grégoire et al., 2014a).[8] The results are shown in Figure 3. It appears that adult performances provide a nearly perfect continuation of children curves: The reading times for both congruent and incongruent stimuli still decrease, and crucially, the difference between the two values still increases. The MSE reached 31.36 ms in adults, whereas the corresponding value for children after 5 years of musical practice was 18.91 ms. Note that this pattern cannot be ascribed to residual differences in cognitive control between 10- to 11-year-old children and adults, given that the (potentially) better control of adults should produce a *decrease* of the effect. These data make it unlikely that a too short amount of practice would be responsible for the observation of a monotonic increase of the MSE with musical practice.

Implications for Reading

In introducing this article, we posited that the decrease in interference from childhood to adulthood observed in Stroop tasks involving reading attests to an increase in cognitive control that could be due either to general factors linked to age, to the strong correlation between age and reading practice, or to the prolonged practice of reading (e.g., Logan, 1985; Tzelgov et al., 1990). Demonstrating that the MSE increases with practice when the effect of age is controlled strongly suggests that the decrease in interference observed in conventional Stroop tasks is due to the former cause. Reading practice in itself would not lead to increased control over reading. Some final words of caution are in order, nevertheless.

Our data indisputably run against the view that increased practice leads to improved cognitive control in general. The counterevidence we provided on note naming in musicians is compelling. But there is a subtle difference in conditions where note naming and word reading are practiced, which could raise a problem, which could prevent a direct generalization of our conclusion to reading.[9] Musicians are exposed to printed music mainly, if not exclusively, on occasions where pronouncing the names of the notes vocally (e.g., at the beginning of practice) or subvocally (e.g., when playing an instrument later in practice) is well suited. By contrast, printed language is so ubiquitous in the environment of educated people that reading anything around us would be counterproductive. Optimal adaptation requires the skill of not attending to, or disengaging attention from, printed language when necessary. As a consequence, it remains possible that a part of the decrease in interference in Stroop tasks involving reading reflects the learned ability to ignore printed matters when the context makes reading inappropriate in real-world conditions. In this framework, improved cognitive control could indeed emerge as a function of practice, but as a domain-specific consequence of learning not to respond to stimuli whenever responding would distract from other activities. Further studies are needed to exam-

[8] We choose these data as a point of comparison, because experimental conditions were closer to the present ones than in Grégoire et al. (2013). The conditions were not exactly identical, but the differences went in a direction that would tend to reduce the magnitude of the MSE in adults (in particular, children were tested with the most familiar octave, C3–C4, whereas adults were tested with two octaves).

[9] We thank an anonymous reviewer of an earlier version for bringing this interpretation to our attention.

ine what would be the respective weight of age and context specific training to account for the inverted U-shaped curve relating interference and practice in standard Stroop tasks.

References

Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *Clinical Neuropsychologist, 16,* 67–80. doi:10.1076/clin.16.1.67.8337

Bédard, A. C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology, 21,* 93–111. doi:10.1207/S15326942DN2101_5

Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. (2011). Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology, 64,* 596–607. doi:10.1080/17470218.2010.513735

Cohen, J. D., MacWhinney, B., Flatt M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color–word test in childhood, adulthood, and aging. *Journal of Genetic Psychology, 100,* 47–53. doi:10.1080/00221325.1962.10533572

Dash, J., & Dash, S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychologist, 1,* 24–33.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. doi:10.1146/annurev-psych-113011-143750

Everatt, J., Warner, J., Miles, T. R., & Thomson, M. E. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3,* 222–228. doi:10.1002/(SICI)1099-0909(199712)3:4<222::AID-DYS12>3.0.CO;2-P

Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental dyslexia. *Child Neuropsychology, 14,* 277–283. doi:10.1080/09297040701290040

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60,* 269–278. doi:10.1027/1618-3169/a000197

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014a). About the unidirectionality of interference: Insight from the musical Stroop effect. *Quarterly Journal of Experimental Psychology*. Advance online publication. doi:10.1080/17470218.2014.896932

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014b). Is the musical Stroop effect able to keep its promises? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61,* 80–83. doi:10.1027/1618-3169/a000222

Howell, D. C. (2010). *Statistical methods for psychology* (8th ed.). Belmont, CA: Cengage Wadsworth.

Kapoula, Z., Le, T. T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., . . . Yang, Q. (2010). Poor Stroop performances in 15-year-old dyslexic teenagers. *Experimental Brain Research, 203,* 419–425. doi:10.1007/s00221-010-2247-x

Klenberg, L., Korkman, M., & Lahti Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20,* 407–428. doi:10.1207/S15326942DN2001_6

Logan, G. D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 778–792. doi:10.1037/0096-1523.8.6.778

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 39,* 367–386. doi:10.1037/h0080066

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195. doi:10.1037/0278-7393.17.2.179

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the color Stroop test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51,* 135–142. doi:10.1016/j.archger.2009.09.040

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. doi:10.1037/0033-2909.109.2.163

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 126–135. doi:10.1037/0278-7393.14.1.126

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4,* 383–391. doi:10.1016/S1364-6613(00)01530-8

Martin, M. (1978). Speech recoding in silent reading. *Memory & Cognition, 6,* 108–114. doi:10.3758/BF03197435

Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie, 144,* 45–62.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54,* 251–282. doi:10.1016/j.cogpsych.2006.07.003

Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color–word test. *Journal of Personality, 31,* 534–558. doi:10.1111/j.1467-6494.1963.tb01318.x

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture–word interference task. *Child Development, 46,* 247–253. doi:10.2307/1128859

Ryan, T. A. (1959). Comments on orthogonal components. *Psychological Bulletin, 56,* 394–396. doi:10.1037/h0041280

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9,* 132–141. doi:10.3758/BF03202327

Schiller, P. H. (1966). Developmental study of color–word interference. *Journal of Experimental Psychology, 72,* 105–108. doi:10.1037/h0023358

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13,* 57–74.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. doi:10.1037/h0054651

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 760–771. doi:10.1037/0278-7393.16.5.760

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging, 13,* 120–126. doi:10.1037/0882-7974.13.1.120

Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology, 38,* 29–60. doi:10.1146/annurev.ps.38.020187.000333

Received January 24, 2014
Revision received May 26, 2014
Accepted May 28, 2014 ■