CONTEXTUAL INTERFERENCE: A META-ANALYTIC STUDY

FRANK BRADY

Department of Health, Physical Education, and Movement Science
C.W. Post Campus, Long Island University

Summary.—A meta-analysis of the contextual interference effect produced 139 estimates of effect sizes from 61 studies. The average overall effect size was .38. The effect size for basic research (.57) was significantly different from applied research (.19). Significant differences were also obtained between the effect sizes for adults (.50) and those for younger learners (.10). Power for retention and transfer scores was not significantly different. The overall mean power of the studies reviewed was .43.

The contextual interference effect has generated much research and debate in motor-learning literature in recent years (Magill & Hall, 1990; Brady, 1998). Battig (1966) first identified this phenomenon or practice peculiarity in verbal learning studies. The contextual interference effect refers to the relatively consistent finding that practicing several related skills in a randomized order, defined as high contextual interference, hinders performance during acquisition but enhances learning in retention and transfer tests, relative to a blocked practice schedule. However, when the skills are practiced in a blocked or repeating schedule, defined as low contextual interference, acquisition is enhanced while retention and transfer performances are impaired relative to a random practice schedule. Shea and Morgan (1979) introduced the concept to motor learning and the results of their pioneering study strongly supported Battig’s hypothesis. These researchers consequently urged practitioners to teach several skills in each practice session so as to maximize the retention and transfer benefits. Similarly, Schmidt (1988) supported the application of contextual interference procedures to the practice of motor skills, claiming that the effect was a stable and dependable principle of motor learning. Magill (1992) noted that, while motor learning researchers sought to bridge the gap between theory and practice, few findings were as applicable to the practitioner as contextual interference effects.

However, some researchers have questioned or cautioned against overgeneralization by extrapolating concepts and methods from one domain to another (Adams, 1983; Newell & McDonald, 1992). Greenwald, Pratkanis, Leippe, and Baumgardner (1986) claimed that theory development inevitably entailed risks of overgeneralization, mainly due to the researchers’ incomplete control over the relevant or moderating variables. Shewokis (1997)

1Address correspondence to Frank Brady, Pratt Recreation Center, 720 Northern Blvd., Brookville, NY 11548.
noted that it was difficult to assess the generalizability of the contextual interference phenomenon given the influence of the participants, tasks, and the myriad of tests used. Brady (1998) concluded from a literature review that the contextual interference effects were generalizable to motor skills; however, the effects appeared to be mediated by boundary conditions such as age, skill, tasks, personality, and amounts of contextual interference. The effect seemed to be more robust in basic research and more appropriate for adult participants.

The literature on contextual interference contains many studies performed with different groups, different tasks, small sample sizes, and low power, thus rendering generalizations based solely on probability misleading. Cohen (1988) wrote that small and trivial differences may be declared statistically significant regardless of their magnitude or meaningfulness. Completing a meta-analysis would help to resolve some of the disparate findings and identify more clearly some of the boundary or mediating conditions of the contextual interference effect. This procedure, popularized by Glass (1982), combines the results of independent studies by calculating an effect size, which represents the magnitude in standard deviation units of a treatment. The effect size is an objective measure of the meaningfulness which, when combined with statistical significance, results in a more accurate picture of treatment effects. The purpose of this research was to examine systematically studies that used blocked, random, and mixed practice schedules of select groups to determine their effect on retention and transfer tests. A secondary purpose was to conduct a power analysis to determine the mean power of the studies on contextual interference.

**Method**

Using computer databases (PsycINFO, FIRSTSEARCH, PROQUEST, INFOTRACK, SPORTDISCUS, and ERIC), data were collected on available studies on contextual interference. One hundred and forty studies were located; however, 63 were selected based on the following criteria: First, studies had to provide means, standard deviations, and sample sizes. If these statistics were not reported, F and t ratios plus degrees of freedom had to be available to estimate the effect size. Second, studies had to include a blocked practice group, a random practice group, and/or a mixed practice group. Third, the studies had to yield a measure of retention or transfer. Fourth, studies had to be reported in refereed journals. Fifth, studies had to be from the discipline of motor learning.

Based upon these criteria, the 63 studies selected (marked with asterisks in the Reference section) yielded 139 effect sizes. Some studies included more than one dependent measure, i.e., absolute and variable error. These were averaged, resulting in one effect size per task or skill, while constant error
scores were not included as they are not regarded as measures sensitive to movement accuracy (Yan, Thomas, & Thomas, 1998). In studies reporting immediate and delayed scores, the latter were included, as learning effects may be obscured in the more immediate measures (Giuffrida, Shea, & Fairbrother, 2002).

The selected studies were coded for variables that might potentially moderate effect sizes: (1) Nature of research—basic versus applied: To be classified as applied, the studies had to be conducted in a field setting using typical sports skills, while basic research had to be conducted in a controlled laboratory environment. (2) Amounts of contextual interference: Studies using random practice schedules were classified as High Contextual Interference; those with blocked practice were labeled as Low Contextual Interference while studies incorporating a mixture of random blocked practice were classified as Mixed Contextual Interference. (3) Age of learner: Participants at or below the eighth grade were classified as children, those in the ninth to the twelfth grades inclusive were classified as high school youth, while those in college and beyond were termed adults in this analysis. (4) Skill level: Participants who were labeled as novices or beginners were classified as low skilled, while those labeled as experienced or ranking above the mean of the group were regarded as skilled. (5) A minimum of five effect sizes were required for each variable to be examined statistically.

Effect sizes were calculated using Cohen's $d$ as an index, expressed by the standardized mean differences divided by the pooled standard deviations of the groups. When these statistics were not reported, the $F$ and $t$ ratios were used when available. The effect sizes were then corrected for bias and sample size according to the procedures specified by Hedges and Olkin (1985). Tests for homogeneity were conducted along with analysis of variance for the identified variables. Rosenthal's formula (1979) for the “file drawer problem” was applied to examine the tolerance of the overall effect size, to negative experimental results. This formula estimates how many new, filed, or unretrieved studies would be needed to increase the effect size to a nonsignificant level. Based on an estimated effect size of .4 and a probability of .05, a power analysis of the selected studies was conducted as specified by Thomas, Lochbaum, Landers, and He (1997).

Results

The means and standard deviations are included in Table 1. A total of 63 of the retrieved studies met the criteria for inclusion. This resulted in 139 effect sizes; however, two were omitted as they were more than three standard deviations removed from the mean. The overall mean of the remaining 137 effect sizes was .38 ($SD = .52$). The $H$ statistic, a test for homogeneity to assess whether all the effect sizes were similar, was significant [$Q_t(136) = $]
104.6, \( p < .01 \)]. The significant \( H \) statistic warranted the investigation of the potentially moderating variables that resulted in effect sizes of different magnitudes. The nature of the research was significant. Basic or lab-oriented research had an average effect size of .57 (SD = .40) that was significantly greater than that of applied or field-based research (.19; SD = .57; \( F_{1,136} = 24.57, p < .01 \)). Age was also a significant variable (\( F_{2,135} = 14.3, p < .01 \)). Adults averaged an effect size of .50 (SD = .45) that was significantly greater than those for either high school youth (.10; SD = .12) or children (.09; SD = .52). However, an analysis of the effect sizes of the latter two groups indicated that they were derived from field-based settings. A further analysis of the field-based effects of the three age groups indicated that the adults average effect size (.35; SD = .46) was still significantly greater than that of high school youth (.10; SD = .12) and children (.09; SD = .52; \( F_{2,65} = 3.24, p < .05 \)). The

<table>
<thead>
<tr>
<th>Variable</th>
<th>( N )</th>
<th>( ES )</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>137</td>
<td>.38</td>
<td>.38</td>
</tr>
<tr>
<td>Basic research</td>
<td>66</td>
<td>.57</td>
<td>.40</td>
</tr>
<tr>
<td>Adult high contextual interference retention</td>
<td>47</td>
<td>.61</td>
<td>.43</td>
</tr>
<tr>
<td>Adult high contextual interference transfer</td>
<td>19</td>
<td>.47</td>
<td>.31</td>
</tr>
<tr>
<td>Applied research</td>
<td>71</td>
<td>.19</td>
<td>.57</td>
</tr>
<tr>
<td>Adult high contextual interference retention</td>
<td>16</td>
<td>.40</td>
<td>.53</td>
</tr>
<tr>
<td>Adult mixed contextual interference retention</td>
<td>5</td>
<td>.10</td>
<td>.22</td>
</tr>
<tr>
<td>Adult high contextual interference transfer</td>
<td>7</td>
<td>.43</td>
<td>.63</td>
</tr>
<tr>
<td>High school high contextual interference retention</td>
<td>6</td>
<td>.15</td>
<td>.10</td>
</tr>
<tr>
<td>High school mixed contextual interference retention</td>
<td>6</td>
<td>.05</td>
<td>.13</td>
</tr>
<tr>
<td>Children high contextual interference retention</td>
<td>13</td>
<td>.08</td>
<td>.57</td>
</tr>
<tr>
<td>Children mixed contextual interference retention</td>
<td>5</td>
<td>.30</td>
<td>.28</td>
</tr>
<tr>
<td>Children high contextual interference transfer</td>
<td>6</td>
<td>.10</td>
<td>.54</td>
</tr>
<tr>
<td>Children mixed contextual interference transfer</td>
<td>6</td>
<td>.08</td>
<td>.52</td>
</tr>
</tbody>
</table>

Note.—Cohen's \( d \) is an effect size index expressed as \( M_1 - M_2 / SD \) pooled, \( N \) = number of effect sizes, \( ES \) = effect size.

The comparison of effect sizes for retention (.40; SD = .51) and transfer (.31, SD = .50) was not significant (\( F_{1,136} = 1.11, p > .05 \)). The amount of contextual interference was significant for retention scores. High amounts of contextual interference (.45, SD = .44) was significantly greater than mixed amounts (.23, SD = .25; \( F_{1,97} = 19.36, p < .01 \)). However, when high contextual interference and mixed amounts were compared for applied tasks, the effect was not significant. It was not possible to conduct a meaningful analysis on the effects of magnitude of skill as only two studies provided data on this variable. The overall mean power was calculated to be .43. The mean power for adult field-based studies was .44, while that of children and high school youth was
.42. The tolerance of these results to null experimental outcomes was calculated to be 204. Thus, 204 effects which were not significant would have to occur to bring the present overall effect size to no significance.

**Discussion**

The results of this meta-analysis were generally in accord with a previous experimental review (Brady, 1978). He concluded that, although there was widespread support for the contextual interference effect in the literature, it was mediated or constrained by a number of boundary conditions, specifically the research setting and the characteristics of the learner. This study reinforced and amplified the nature of some of the boundary conditions.

Cohen, (1988) stated that in the behavioral sciences an effect size of .20 represented small differences, .50 moderate differences, and .80 large differences. This meta-analysis produced an overall effect size of .38. In basic research studies, the average effect size was .57, while the applied research generated an average effect size of .19. The most important finding was the considerable disparity between effect sizes obtained in basic research versus applied research. Obviously the relatively small effect sizes obtained by children and high school youth in applied settings compounded this difference. However, when the comparison is limited to adults, significant differences exist, although of less magnitude.

A number of researchers have questioned the utility and relevance of basic research in motor learning, claiming it lacked validity or fidelity to real world environments (Hoffman, 1990; Newell & Rovegno, 1990; Shebilske, & Worchel, 1993). Lee and White (1990) suggested that the contextual interference may be more evident in the laboratory and more amenable to subjects performing simple and less intrinsically interesting tasks. Hebert, Landin, and Solmon (1996) noted that laboratory-based research produced moderate to robust findings; in applied settings the findings tended to be inconsistent and equivocal. According to Lee and White, contextual interference effects were more readily elicited in laboratory settings as tasks posed few motor demands, were cognitively loaded, lacked intrinsic interest, and quickly reached an asymptote. By contrast, they argued that the more challenging and inherently more intrinsic nature of sports skills would preclude producing contextual interference effects. Landin and Hebert (1997) and Shea, et al. (1993) stated that laboratories were rigorously controlled environments, wherein confounding variables were repressed. By contrast, a myriad of factors that could influence performers occurred freely and interacted differentially from task to task in applied settings. Thomas and Nelson (2001) noted that basic research ranked high in internal validity; this was obtained at the expense of external or ecological validity. Goode and Magill (1986)
attributed the lack of significant findings in field-based research to the less sensitive or discriminative measurement units employed. For example, instruments calibrated in milliseconds lead to more discriminative measures of performance in basic research in comparison to more gross measures in applied settings such as the number of putts, hits, baskets, etc.

Another noteworthy and significant finding was the different effect sizes for the adults and the two younger groups. The average effect size for adults in applied settings was roughly moderate (.35), while it was almost trivial, approximately .10, for the young groups. This finding is hardly surprising and generally in accord with the developmental status of younger and inexperienced learners, given their more limited information-processing capacities. Newell and McDonald (1992) claimed that the influence of random practice occurred in the later stages of skill acquisition and that practice schedules do not influence the processing operations involved in establishing a basic movement pattern (Gentile, 1972). It appears that the tasks in this review presented sufficient contextual interference for the learners without the need for random practice. Wulf and Schmidt (1994) suggested that random practice for novices may result in excessive response variability, thereby inhibiting the development of a stable motor pattern. Invoking a similar concept, Guadagnoli, Holcomb, and Weber (1999) proposed that the performer’s stage of learning and the practice schedule interact with efficient learning. Although the interaction of stage of learning and practice certainly applies to adults, the interactive effect would be expected to be more acute for younger and more inexperienced learners. From a practical and pedagogical standpoint, it could be reasonably argued that the complexity of sports skills coupled with the more limited information-processing capabilities may render random practice overwhelming for children. Before progressing to random practice, beginners need sufficient time to explore and establish the basic coordinated movement pattern (Gentile, 1972), break the proficiency barrier (Haubenstricker & Seefeldt, 1986), or experience the transformation of power (Siedentop, 1983). Hebert, et al. (1996) stated that random practice schedules were counterproductive during the early stages of skill learning and that novices should adopt a blocked practice schedule. In addition to the magnitude of skill, the number of skills being taught needs to be considered. Too many skills being taught in a single practice session may overload or overwhelm the young learner. Researchers might incorporate this as another moderating variable.

The effect sizes for retention and transfer were not significantly different although the retention scores were marginally higher. Shewokis and Snow (1997) stated that transfer tests were more reliable indicators of the contextual interference effect. Magill (2004) noted that transfer tests were more reflective of adaptability, while retention tests were specific to mea-
sures of learning. Thus, while overlap between retention and transfer is to be expected, they measure related but different constructs.

Landin and Hebert (1997) proposed that mixed amounts of contextual interference would be the most beneficial for the learning process. This review did not support that proposition. Higher contextual interference, as measured by retention, was significantly greater than mixed levels. A number of factors confluened to produce this result, specifically the relatively moderate to large effect sizes of adults in basic research, coupled with the much smaller effect sizes of the younger subjects in field settings. However, the comparison between mixed amounts and high contextual interference in applied settings was not significant. This may be attributed to the various uncontrolled factors that interact differentially in applied settings (French, Rink, & Werner, 1990; Landin & Hebert, 1997).

Although skill was coded as a potentially moderating variable, there were insufficient effect sizes to conduct a meaningful analysis.

Cohen (1988) and Thomas, et al. (1997) advocated selecting sample sizes that guarantee a power of .80 or more to enhance the chances of making a correct decision concerning real differences among treatment groups. The overall mean power was calculated to be .43, while the mean power for adult field-based studies was .44 and that of children and high school youth was .42. The power analysis was based upon an estimated effect size of .4, a probability level of .05, and the sample sizes in the review of the studies on the interference effect. Cohen reported that power of less than .80 incurred greater risk of a Type II error as statistical significance is heavily affected by large variances and small sample sizes. Obviously, the lack of adequate power is a critical issue in studies on contextual interference, particularly in field-based studies, with the problem being more acute in the younger populations where the effect size is small.

It may be concluded from this meta-analysis that the overall effect size fell between small and moderate (.38), while in applied research, it was small (.19). The effect sizes for adults is approximately moderate (.50) while it is relatively small (.10) for the younger populations. The meta-analysis suggests a number of implications for research, especially larger samples in applied research for all populations to increase the power. There is also a much greater need for research on manipulating different amounts of contextual interference in applied settings. Perhaps a greater focus on ecologically based research with appropriate control of moderating or mediating factors would help answer the question of whether "contextual interference is a laboratory artifact or sport-skill related" (Al-Mustafa, 1989).

REFERENCES


*References marked with an asterisk indicate the studies are included in the meta-analysis.*


Accepted May 17, 2004.
This article has been cited by:


5. R. MENAYO, F. J. MORENO, R. SABIDO, J. P. FUENTES, J. A. GARCÍA. 2010. SIMULTANEOUS TREATMENT EFFECTS IN LEARNING FOUR TENNIS SHOTS IN CONTEXTUAL INTERFERENCE CONDITIONS. *Perceptual and Motor Skills* **110**:2, 661–673. [Abstract] [PDF] [PDF Plus]

6. ANTONIOS K. TRAVLOS. 2010. SPECIFICITY AND VARIABILITY OF PRACTICE, AND CONTEXTUAL INTERFERENCE IN ACQUISITION AND TRANSFER OF AN UNDERHAND VOLLEYBALL SERVE. *Perceptual and Motor Skills* **110**:1, 298–312. [Abstract] [PDF] [PDF Plus]

7. MAURIZIO BERTOLLO, MARIKA BERCHICCI, ATtilio CARRARO, SILVIA COMANI, CLAUDIO ROBAZZA. 2010. BLOCKED AND RANDOM PRACTICE ORGANIZATION IN THE LEARNING OF RHYTHMIC DANCE STEP SEQUENCES. *Perceptual and Motor Skills* **110**:1, 77–84. [Abstract] [PDF] [PDF Plus]


10. HIROSHI SEKIYA. 2006. CONTEXTUAL INTERFERENCE IN IMPLICIT AND EXPLICIT MOTOR LEARNING. *Perceptual and Motor Skills* **103**:2, 333–343. [Abstract] [PDF] [PDF Plus]

11. GAYE J. KELLER, YUHUA LI, LAWRENCE W. WEISS, GEORGE E. RELYEA. 2006. CONTEXTUAL INTERFERENCE EFFECT ON ACQUISITION AND RETENTION OF PISTOL-SHOOTING SKILLS. *Perceptual and Motor Skills* **103**:1, 241–252. [Abstract] [PDF] [PDF Plus]