Enhancement of Executive Control through Short-term Cognitive Training: Far-transfer Effects on General Fluid Intelligence

Abstract: We predicted that short-term training of executive control would improve both cognitive control itself (Hypothesis 1) and general fluid intelligence (Hypothesis 2). We randomly assigned 120 high school students to the experimental and control groups. The former underwent a 14-day training of four executive functions: interference resolution, response inhibition, task switching, and goal monitoring. The latter did not train anything. The training significantly improved cognitive control and IQ. The control group also improved their IQ scores but gained less than the experimental one. However, the amount of IQ scores enhancement did not correlate with the majority of indices of the training effectiveness, thus justifying the supposition that such enhancement resulted from the general effect of participation in the training rather than from the improvement of executive functions underlying Gf.

Key words: cognitive training, executive control, intelligence
amount to knowledge acquisition nor to teaching people how to take intelligence tests. Therefore, their outcomes cannot be accounted for in terms of enhancement of crystallized intelligence through knowledge acquisition. Metaphorically speaking, cognitive trainings might be compared to the tuning of car’s engine in order to make it more powerful. In this study, we aimed at “tuning” just one aspect of mind’s “engine” – executive control.

Attempts to improve Gf through working memory (WM) training became popular about 15 years ago (e.g., Buschkuehl & Jaeggi, 2010; Klingberg, Forssberg, & Westerberg, 2002; Klingberg et al., 2005), although meta-analyses suggest that their effects may be small and short-lived (e.g., Au et al., 2014; Melby-Lervåg & Hulme, 2013). Similar efforts concerning executive functions are less popular. Results of several studies suggest that systematic trainings of executive control may be effective in children (Banachewsky & Rothenberger, 2006; Klingberg et al., 2002, 2006; Thorell et al., 2009) and adults (Cepeda et al., 2001; Bherer et al., 2005; Karbach and Kray, 2009). For instance, preschool children who trained cognitive control improved both executive attention and IQ test scores (Rueda, Posner, & Rothbart, 2004; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005). However, a meta-analysis of 12 works published between 1986 and 2011 (Karch et al., 2013) revealed that the effectiveness of executive control training is questionable. Furthermore, in a study with 11,340 volunteers engaged in an on-line training of cognitive functions (Owen et al. 2010), the authors found modest improvements concerning the trained functions but no far-transfer effects. Maybe cognitive trainings work for particular samples, like small children (Rueda et al., 2004, 2005), or elderly persons (Bherer et al., 2005; Karbach & Kray, 2009), being less efficient for young healthy volunteers.

An important issue must be raised concerning the function of updating. Updating is a process involved in WM functioning and at the same time it is commonly defined as one of executive functions (Miyake et al., 2000). Importantly, it seems to be the only executive function significantly correlated with Gf (e.g., Friedman et al., 2006). Many studies on WM training may be also interpreted as referring to executive control, if the updating function is involved. Therefore, we decided to exclude updating from our procedure and choose two other functions: interference resolution, which consists in overcoming cognitive conflict inflicted by inconsistent stimuli (e.g., Eriksen & Eriksen, 1974; Stroop, 1935), and goal maintenance, which amounts to prospective keeping in mind personal intentions or external instructions (e.g., Au et al., 2014; Melby-Lervåg & Hulme, 2013). Since these functions have been less frequently used in training studies, they are worth considering. Additionally, we chose the “standard executive functioning” – prepotent response inhibition and task switching. We intended to explore which functions would be easier to train, as well as which of them, if any, would produce far-transfer effects on Gf.

### Method

#### Participants

One hundred and twenty individuals were randomly assigned to training and control groups. Due to dropout, the final participant group included 85 individuals (training group: \(N = 46\), 25 women; control group: \(N = 39\), 22 women). The mean age of the participants was 17.32 \((SD = 0.47,\ range\ 17–18)\). The members of both groups received a 100 PLN (20–25 €) reward. Additionally, the highest-scoring participant got a personal computer (PC) and one randomly chosen person received a digital camera. The former reward was introduced in order to encourage all participants to exert effort in the training, and the latter to motivate those who felt they had not done well enough to win the main prize.

#### Tests and materials

The Gf level was assessed using Raven’s Standard Progressive Matrices Plus (Raven, Raven, & Court, 2003), adapted into Polish by Jaworowska and Szustrowa (2000). The test does not have any alternate versions, and therefore it was used twice. Standard version, which has its parallel counterpart, was rejected due to difficulty level considered to be too high for the participants’ population (Jaworowska & Szustrowa, 2000).

Four computerized tasks were employed as the training tools of executive functions. The non-verbal version of Eriksen flanker task (Eriksen & Eriksen, 1974) was used to train interference resolution. The participants were categorizing figures as quadrangles or triangles by clicking right or left mouse button, respectively. In the quadrangle condition the triangles appeared as the interfering stimuli, and the other way round. Each trial included six interfering stimuli presented symmetrically on both sides of the target stimulus. The distance between the target stimulus and the interfering stimuli was 1 cm. The exposure time of each stimulus was 2000 ms and the interval between them was 500 ms. Each session consisted of one series of 120 trials, including 10 trials in which the interfering stimuli were present.

The response inhibition training task (De Jong, Coles, & Logan, 1995) involved both inhibition and execution of alternative responses. The participants were categorizing numbers as even or odd using the right or left mouse button, respectively. However, when a given number was presented within a square, the button-parity match was reversed. Each stimulus was exposed for 2000 ms and the interval between the stimuli was 500 ms. Each session consisted of one series of 135 trials out of which 32 required the execution of an alternative response.

The task switching training involved letter categorization. Each character could be categorized either as a vowel/consonant, or as having/not having angles (e.g., K, H/S/O). The participants were categorizing the stimuli by clicking the right mouse button when a letter had no angles or when it was a consonant, and by clicking the left mouse button when a letter had angles or was a vowel. Hints concerning the valid classification rule...
were displayed between the letter sequences. The rule could change after one, two, three, four, or five trials, but the duration of each sequence was unpredictable. The exposure time of each stimulus was 2000 ms, the exposure time of a hint was 1000 ms, and the interval between the stimuli was 500 ms. Each session consisted of one series of 144 trials, with 48 trials occurring after a switch.

In the goal monitoring training task two types of stimuli (quadrangles and triangles) were simultaneously presented on a screen. The participants were categorizing a stimulus that appeared on a given side of a screen as a quadrangle or as a triangle on the basis of one of the four hints describing the current target shape. The hints were appearing between the sequences of stimuli, varying in number from 1 to 5. The hint could direct the participant to attend to the left/right side of a screen, or to the same/opposite side of a screen. Hints indicating left or right side were called “explicit”, whereas those indicating the same or the opposite side of a screen were called “implicit”. Implicit cueing trials were supposed to be more demanding because of increased cognitive load involved in the processes of categorization. They also required the constant updating of the location to which one’s attention should be directed. The hint was visible for 1000 ms, the exposure time of each stimulus was 2000 ms, and the interval between the stimuli was 500 ms. Each session consisted of one series of 326 trials.

Procedure
The Raven’s Progressive Matrices Plus test was administered as pretest in both groups in the first session. The cognitive training took place for 14 consecutive working days. Each session lasted for about 60 minutes and tasks took about 12 minutes each. The day after the training sessions ended, Raven’s test was administered in both groups again. The time for pretest and posttest completion of cognitive training tasks was limited to 60 minutes.

Results
First, the average progress in training tasks was examined via trend analysis of average results obtained in subsequent training sessions. ANOVA with linear contrasts (contrast coefficients: -7, -6, -5, -4, -3, -2, -1, 1, 2, 3, 4, 5, 6, 7) was conducted for the indicators of error rate and response time in all training tasks. The indices of training effectiveness were created for each task separately.

A gradual linear decrease in the proportion of errors \( F(1, 19) = 6.24; p = .022; \eta^2 = 0.257 \) and in the average response time \( F(1, 19) = 90.999; p < .001; \eta^2 = 0.835 \) was observed for task requiring interference resolution. Similarly, a gradual linear decrease in the number of errors \( F(1, 19) = 79.761; p < .001; \eta^2 = 0.808 \) and in the average response time \( F(1, 19) = 51.427; p < .001; \eta^2 = 0.730 \) was observed in task switching. Also, a gradual linear decrease in the number of errors \( F(1, 19) = 26.548; p < .001; \eta^2 = 0.610 \) and in the average response time \( F(1, 19) = 23.12; p < .001; \eta^2 = 0.576 \) was observed for response inhibition task. These findings support the first hypothesis. However, no significant changes in the level of performance were observed for the goal-monitoring task. These results are presented in Figures 1 and 2.

Two-way mixed design analysis of variance was run to verify the Hypothesis 2. The between-subjects factor was the group allocation, whereas the within-subjects factor referred to the pretest and posttest measures. The results showed a significant interaction effect, \( F(1, 83) = 13.843; p < .001; \eta^2 = 0.143 \) (Figure 3). The analysis of simple main effects showed that the cognitive training group scored higher on the posttest \( (M = 46.23; SE = 0.79) \) than on the pretest \( (M = 43.67; SE = 0.82) \), \( F(1, 83) = 79.397; p < .001; \eta^2 = 0.489 \). Importantly, the control group also scored higher on the posttest \( (M = 44.46; SE = 0.72) \) than on the pretest \( (M = 43.35; SE = 0.75) \), \( F(1, 83) = 17.509; p < .001; \eta^2 = 0.174 \), but the effect size in the training group was significantly larger. The experimental group improved their Raven’s scores for 54% of their initial score’s standard deviation and the control group showed only a 21% increase. No significant differences between groups were observed in the pretest \( (F(1, 83) = 0.082; p = .775; \eta^2 = 0.001) \) and in the posttest \( (F(1, 83) = 2.754; p = .101; \eta^2 = 0.032) \).

The analysis showed a very strong correlation between two fluid intelligence measurements \( (r = 0.927; p < .001) \). We also observed a negative correlation \( (r = -0.255, p < 0.018, \text{two-tailed}) \) between intelligence level in the first measurement and the amount of increase in intelligence level, estimated as the posttest minus pretest difference. This finding suggests that the training procedures may have been more beneficial to participants who obtained relatively lower scores in the initial intelligence assessment.

In order to test whether the increase in intelligence level depended on the results of cognitive training tasks, we created the following indicators:

(1) **Faultiness**, by calculating the average number of errors committed in each session and the average correct response time in each session. The lower these indices, the better the results in the tasks.

(2) **Progress**, by subtracting the number of errors in the last session from the number of erroneous responses in the first session, and subtracting the average response time in the last session from the average response time in the first session. Positive values indicated that a participant benefitted from the training on the start-to-finish basis.

(3) **Dynamics**, by calculating the average change in the number of errors, or response time, between subsequent sessions. Positive values indicated that a participant benefitted from the training on the session-to-session basis.

As mentioned above, we found that the training effects were more salient in participants scoring low in the pretest. Therefore, we expected positive correlations between IQ gains and the indices of faultiness, progress, and dynamics. We supposed that participants with large IQ gains would be characterized by poor performance throughout the training, due to lower initial IQ level. However, we expected that they would obtain higher indices of progress and dynamics, to compensate the relatively poor performance in the
training tasks, and to benefit from training in order to increase the general intelligence level.

The results (Table 1 and Table 2) confirmed the predictions concerning the average number of errors in the interference resolution ($r = .417; p = 0.014$) and switching tasks ($r = .522, p = .002$). No significant relationships between the increase in intelligence level and the remaining indicators of performance in the training tasks have been observed. Surprisingly, we did not find systematic relationships between faultiness and Raven’s score in the pretest, although negative correlations should be expected if intelligence is rooted, to some extent, in the effectiveness

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Raven Pretest</th>
<th>Raven Posttest</th>
<th>Raven progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interference (ER)</td>
<td>.253#</td>
<td>.422*</td>
<td>.417*</td>
</tr>
<tr>
<td>$p$</td>
<td>.097</td>
<td>.012</td>
<td>.014</td>
</tr>
<tr>
<td>Interference (RT)</td>
<td>.077</td>
<td>.130</td>
<td>.129</td>
</tr>
<tr>
<td>$p$</td>
<td>.197</td>
<td>.256</td>
<td>.258</td>
</tr>
<tr>
<td>Inhibition (ER)</td>
<td>.362*</td>
<td>.341*</td>
<td>-.080</td>
</tr>
<tr>
<td>$p$</td>
<td>.029</td>
<td>.038</td>
<td>.346</td>
</tr>
<tr>
<td>Inhibition (RT)</td>
<td>-.051</td>
<td>.015</td>
<td>.174</td>
</tr>
<tr>
<td>$p$</td>
<td>.398</td>
<td>.470</td>
<td>.188</td>
</tr>
<tr>
<td>Switching (ER)</td>
<td>.125</td>
<td>.328*</td>
<td>.522**</td>
</tr>
<tr>
<td>$p$</td>
<td>.270</td>
<td>.044</td>
<td>.002</td>
</tr>
<tr>
<td>Switching (RT)</td>
<td>.111</td>
<td>.203</td>
<td>.233</td>
</tr>
<tr>
<td>$p$</td>
<td>.283</td>
<td>.145</td>
<td>.122</td>
</tr>
<tr>
<td>Monitoring (ER)</td>
<td>.165</td>
<td>.128</td>
<td>-.109</td>
</tr>
<tr>
<td>$p$</td>
<td>.200</td>
<td>.258</td>
<td>.291</td>
</tr>
<tr>
<td>Monitoring (RT)</td>
<td>-.0109</td>
<td>-.047</td>
<td>.167</td>
</tr>
<tr>
<td>$p$</td>
<td>.290</td>
<td>.405</td>
<td>.198</td>
</tr>
</tbody>
</table>

Note: ER – error rate averaged across 14 sessions; RT – reaction time averaged across 14 sessions; Raven progress – difference between posttest and pretest in Raven’s test; * $p < 0.05$; ** $p < 0.01$.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Raven Pretest</th>
<th>Raven Posttest</th>
<th>Raven progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interference (ER)</td>
<td>.431*</td>
<td>.328*</td>
<td>.298#</td>
</tr>
<tr>
<td>$p$</td>
<td>.011</td>
<td>.045</td>
<td>.062</td>
</tr>
<tr>
<td>Interference (RT)</td>
<td>.110</td>
<td>.080</td>
<td>-.085</td>
</tr>
<tr>
<td>$p$</td>
<td>.290</td>
<td>.344</td>
<td>.334</td>
</tr>
<tr>
<td>Inhibition (ER)</td>
<td>.233</td>
<td>.278#</td>
<td>.099</td>
</tr>
<tr>
<td>$p$</td>
<td>.117</td>
<td>.076</td>
<td>.309</td>
</tr>
<tr>
<td>Inhibition (RT)</td>
<td>.033</td>
<td>.120</td>
<td>.221</td>
</tr>
<tr>
<td>$p$</td>
<td>.435</td>
<td>.272</td>
<td>.129</td>
</tr>
<tr>
<td>Switching (ER)</td>
<td>.333*</td>
<td>.518**</td>
<td>.446***</td>
</tr>
<tr>
<td>$p$</td>
<td>.045</td>
<td>.003</td>
<td>.010</td>
</tr>
<tr>
<td>Switching (RT)</td>
<td>.211</td>
<td>.302#</td>
<td>.216</td>
</tr>
<tr>
<td>$p$</td>
<td>.145</td>
<td>.063</td>
<td>.140</td>
</tr>
<tr>
<td>Monitoring (ER)</td>
<td>-.303#</td>
<td>-.420*</td>
<td>-.279#</td>
</tr>
<tr>
<td>$p$</td>
<td>.058</td>
<td>.013</td>
<td>.076</td>
</tr>
<tr>
<td>Monitoring (RT)</td>
<td>.152</td>
<td>.129</td>
<td>-.069</td>
</tr>
<tr>
<td>$p$</td>
<td>.221</td>
<td>.257</td>
<td>.363</td>
</tr>
</tbody>
</table>

Note: ER – error rate averaged across 14 sessions; RT – reaction time averaged across 14 sessions; Raven progress – difference between posttest and pretest (the second minus the first assessment); GAIN – difference between the first and the last day of training (1st session minus 14th session); * $p < 0.05$; ** $p < 0.01$; *** $p <0.001$. 
of executive control. Instead, we found positive correlations of pretest and posttest Raven’s scores with the average number of errors in the inhibition task ($r = .362, p = .029$, and $r = .341, p = .038$, respectively). Table 1 also shows an unexpected positive correlation ($r = .328, p = .044$) of the average number of errors in the switching task and posttest Raven’s scores.

The significant relationship between the increase in intelligence level and the indicators of progress (Table 2), calculated by error subtraction, was observed for the switching task ($r = .446; p = .01$). Similar relationships regarding interference resolution ($r = .298, p = .062$) and goal monitoring ($r = -.279, p = .076$) were only marginally significant, the latter being unexpectedly negative. The positive correlations indicate that people who showed progress during the training in the switching task tended to gain more in the Raven’s posttest measurement in comparison to the pretest measurement, whereas the marginally significant negative correlation supports the opposite conclusion. Finally, we did not observe any significant relationships between the increase in intelligence level and the dynamics of change. No significant relationships between the first intelligence measurement and the dynamics indicators could be observed, either.

**Discussion**

We trained 39 participants with four cognitive control tasks: interference resolution, response inhibition, task switching, and goal monitoring. Before and after the training, participants took part in intelligence assessment sessions. Participants in the control group did not train anything but underwent both intelligence assessment sessions. We found that performance in three tasks showed significant training-related improvement. We also found that both groups improved their intelligence test scores but this effect was stronger for the experimental group. The IQ rises were negatively correlated with the initial level of intelligence, suggesting that low-scoring participants gained more than high-scoring ones. Additionally, we found that the IQ gains were positively correlated with the overall number of errors in two tasks: interference resolution and task switching. They were also correlated with progress made throughout training of the task switching function. Finally, we were unable to find any relationships between IQ gains and the parameters of skill acquisition dynamics.

These findings bring some support to both hypotheses of our research. The first hypothesis presumed that systematic training would improve executive control and, indeed, we can see systematic improvement of performance in three out of four training tasks (Fig. 1 and Fig. 2). The second hypothesis claimed that the training effects would generalize to Gf level. The training group improved their IQ scores to a greater extent than did the control group, therefore we assume that this effect could be at least partly attributable to the improvement in Gf level (Fig. 3). Thus, we can conclude that our study contributes to the debate on malleability of Gf through short-term interventions.

However, several caveats seem necessary. Firstly, we did not train our control group. Should the control group be engaged in some training, our findings might be more persuasive. Substantial dropout rate suggests that the level of motivation was a factor that might have contributed to the results. Therefore, our findings are prone to criticism that the observed outcomes resulted from increased motivation of the active group rather than the training itself. However, a recent meta-analysis of studies aimed at improving Gf through WM training (Au et al., 2014) found that the type of control group (passive or active) is not a moderator of the training effects. Additionally, the passive control group in our study was assigned at random. Secondly, relationships between IQ gains and the indices
of training effectiveness were weak and inconsistent. The experimental group improved their IQ scores to a greater extent than the control group, but these effects are rather difficult to bind with the effectiveness of training. Rather, it seems that the sheer participation in the procedures helped the experimental group to improve their intelligence test scores. As to the control group, the observed increase can be easily explained: they were tested twice with the same assessment tool, and the effect of retest usually brings about some rise of raw scores. In the case of experimental group, the effect of posttest probably added to the effect of training, thus producing the hypothesized interaction shown in Fig. 3.

The lack of relationships between the range of IQ increase and the indices of training effectiveness (except for task switching) is an issue pertaining to the theoretical aspects of our study. Executive control is believed to underlie general mental ability. If so, the better the indices of executive control, the higher the intelligence test scores. Moreover, the better the effects of training of executive control, the more salient the IQ score gains. The above-outlined reasoning motivated our study from the very beginning. Since we did not find strong arguments confirming that IQ gains would be related to effectiveness of training, we have to conclude that executive control probably does not underlie general mental ability. This conclusion is supported by the lack of systematic relationships between the results of the first intelligence assessment and executive control indices. We found just one weak correlation between intelligence test scores and the inhibition task, but this correlation is positive, meaning that intelligent people committed more errors than less intelligent ones. Altogether, our findings did not support the hypothesis that Gf is subserved by effectiveness of executive control. However, this could also be due to the fact that only one measure of Gf was used in the pretest and posttest – employment of a wider range of tools needs to be taken into consideration in future studies.

As to the practical aspects of our study, we can conclude that systematic training of cognitive functions may serve as an effective way to improve one’s intellectual capacity. The effect size was not strikingly large, probably because of the fact that the training was relatively short-term. Cognitive training organized for months or years could be much more effective and long-lasting, if only high level of participants’ motivation is ensured. It is worth mentioning that the effects of training were stronger in the case of people whose intelligence test scores were relatively lower at the beginning of this intervention. Due to the fact that we assessed IQ levels at two time points, it cannot be excluded that the observed effects are the result of regression to the mean. However, it is possible that cognitive trainings may be beneficial first of all to those whose cognitive abilities are below the average. This line of reasoning also suggests that for bright people, trainings may be not that beneficial.

References


