Supplementary Material

Divergent thinking measures results

Fluency

The results for a univariate analysis for fluency were in line with the MANOVA reported in the main text. We found a significant main effect of Time, $F(1, 96) = 32.6$, $p < .001$, $\eta p^2 = .253$, no effect of Group, $F(2, 96) = .7$, $p = .515$, $\eta p^2 = .014$, but a significant interaction between Time and Group, $F(2, 96) = 5.4$, $p = .006$, $\eta p^2 = .101$. Follow-up analyses revealed that at T1 the difference between the groups was not significant, $F(2, 96) = .3$, $p = .753$, $\eta p^2 = .006$, but at T2 it was, $F(2, 96) = 3.1$, $p = .049$, $\eta p^2 = .061$. Post-hoc comparisons for T2 revealed that BilS children had significantly higher fluency scores than NoL2 children, $p = .018$, and marginally higher fluency scores than L2 learners, $p = .074$, while the fluency scores of L2 learners and NoL2 children did not differ, $p = .645$.

Furthermore, fluency significantly improved from T1 to T2 for BilS children by an average of 3.6 (SD 3.8), $F(1, 31) = 29.9$, $p < .001$, and for L2 learners by an average of 1.5 (SD 3.4), $F(1, 28) = 5.6$, $p = .025$. In contrast, the average fluency improvement of 0.9 (SD 3.4) for NoL2 children was not significant, $F(1, 37) = 3.0$, $p = .091$.

Flexibility

Results for a univariate analysis of creative flexibility again followed the general pattern of the MANOVA, albeit with a clearer difference between BilS and L2 learners. We found main effects of Time, $F(1, 96) = 29.4$, $p < .001$, $\eta p^2 = .234$, and of Group, $F(2, 96) = 7.0$, $p = .001$, $\eta p^2 = .128$, as well as a significant interaction between Time and Group, $F(2, 96) = 6.3$, $p = .003$, $\eta p^2 = .116$. Follow-up analyses showed again no effect of Group at T1, $F(2, 99) = .3$, $p = .721$, $\eta p^2 = .007$, but at T2, $F(2, 99) = 9.6$, $p < .001$, $\eta p^2 = .167$.

Comparisons for T2 showed that BilSs had significantly higher flexibility scores than both L2 learners, $p = .013$, and NoL2s, $p < .001$, while the scores of L2 learners and NoL2 children did not differ, $p = .112$. And similar to fluency, flexibility improved from T1 to T2 for BilS children by an average of 3.0 (SD 3.1), $F(1, 31) = 28.7$, $p < .001$, and for L2 learners by an average of 1.3 (SD 2.4), $F(1, 28) = 8.0$, $p = .008$. Again, the average improvement of .5 (SD 3.0) for NoL2 children was not signficant, $F(1, 37) = 1.1$, two-sided $p = .291$.

Originality

The results of the univariate analysis of originality scores were somewhat different from those for fluency and flexibility measures. We found no main effect of Time, $F(1, 96) = .9$, $p = .334$, $\eta p^2 = .010$, but of Group, $F(2, 96) = 3.1$, $p = .048$, $\eta p^2 = .061$, as well as a trend for an interaction between Time and Group, $F(2, 96) = 2.6$, $p = .083$, $\eta p^2 = .051$. Follow-up analyses showed no effect of Group at T1, $F(2, 98) = .8$, $p = .474$, $\eta p^2 = .015$, but at T2, $F(2, 96) = 4.0$, $p = .022$, $\eta p^2 = .077$. BilS children outperformed NoL2 children at T2, $p = .006$, while the scores of L2 learners fell in between those of the other two groups but did not significantly differ from either BilS children, $p = .284$, or NoL2s, $p = .113$. Notably, compared to fluency and flexibility scores, originality scores did not significantly change over time. Comparing values over time shows that there was no significant change for any of the groups, BilS: $F(1, 31) = 2.2$; $p = .152$; L2: $F(1, 28) = 1.4$; $p = .254$; NoL2s: $F(1, 37) = 2.6$; $p = .117$.

In sum, the groups did not differ at any measure of divergent thinking at T1. Across the two time points, only BilS children and L2 learners improved and only in terms of fluency and flexibility. This meant that the BilS children outperformed NoL2 children in terms of fluency and flexibility. L2 learners fell in between the two groups. In terms of

flexibility they were not significant from either group. In terms of fluency, they scored significantly lower than BilS children and more similarly to NoL2 children.

Executive Function measures results

Dimensional Change Card Sort task (DCCS)

Figure 2 shows the result of the DCCS. We found significant main effects of Time, $F(1, 73) = 101.5$, $p < .001$, $\eta p^2 = .582$, and Group, $F(2, 73) = 2.0$, $p = .142$, $\eta p^2 = .052$, as well as a Time x Group interaction, $F(2, 73) = 7.8$, $p < .001$, $\eta p^2 = .175$. Groups did not differ at T1, $F(2, 73) = .3$, $p = .758$, $\eta p^2 = .008$, but they did so at T2, $F(2, 73) = 9.1$, $p < .001$, $\eta p^2 = .199$. Post-hoc tests for T2 showed that BilS children had higher DCCS scores compared to L2 learners, $p = .030$, and NoL2 children, $p < .001$, while L2 learners scored marginally higher than NoL2 learners, $p = .054$. Furthermore, DCCS scores improved from T1 to T2 for all groups, BilS children: average improvement of 4.8 (SD 2.3), $t(31) = 12.0$, two-sided $p < .001$, L2 learners: average improvement of 2.9 (SD 2.9), $t(28) = 5.4$, two-sided $p < .001$, NoL2 children: average improvement of 1.7 (SD 3.0), $t(37) = 3.4$, two-sided $p = .002$.[1]

Attentional Network Task (ANT)

As indicated, the ANT provides measures for conflict, alerting and orienting. Figure 3 shows the results of the three measures. As evident in the following tests, the results for the ANT were very different from the other tests. First, children did not improve from T1 to T2. Second, BilS children scored higher on the alterting index.

Conflict

[1] Note that an analysis with all 99 participants showed the same result pattern.

For the conflict index, there was no significant effect of Time, $F(1, 73) = .003$, $p = .960$, $\eta p^2 < .001$, or Group, $F(2, 73) = 2.0$, $p = .139$, $\eta p^2 = .053$, or a Time x Group interaction, $F(2, 73) = 1.2$, $p = .317$, $\eta p^2 = .031$.

Orienting

For the orienting index, we found again no effect of Time, $F(1, 73) = .008$, $p = .930$, $\eta p^2 < .001$, or of Group, $F(2, 73) = .184$, $p = .832$, $\eta p^2 = .005$, but a significant Time x Group interaction, $F(2, 73) = 3.58$, $p = .033$, $\eta p^2 = .090$. However, groups did not differ at T1, $F(2, 75) = 2.22$, $p = .116$, $\eta p^2 = .057$, or at T2, $F(2, 73) = 1.3$, $p = .270$. Furthermore, only L2 learners showed an improvement in orienting from T1 to T2, while other groups did not show any significant change, BilS children: $t(25) < -.01$, two-sided $p = .994$, L2 learners: $t(23) = 2.1$, two-sided $p = .043$, NoL2 children: $t(25) = -1.4$, two-sided $p = .185$.

Alerting

For the alerting index, there was no significant effect of Time, $F(1, 73) = .004$, $p = .952$, $\eta p^2 < .001$, or a Time x Group interaction, $F(2, 73) = 1.8$, $p = .173$, $\eta p^2 = .047$, but a significant main effect of Group, $F(2, 73) = 4.0$, $p = .023$, $\eta p^2 = .098$. Posthoc tests (Bonferroni corrected) revealed that the alerting index was significantly higher for BilS children than for L2 learners, $p = .008$, and marginally higher than for NoL2 children, $p = 0.055$. L2 learners and NoL2 children did not differ from each other, $p = .416$.

Supplementary Table 1. Categories for flexibility scoring

Categories identified at T1	Categories identified at T2
Buildings (e.g., schools, churches, castles)	Buildings
Boxes and containers	Boxes and containers
Nests/homes/shelters for animals/people	Lines used as roads/paths/bases (with someone walking/standing on them or a vehicle driving on it)
Squares used as body of animals/people	Lines used as the body for animals/people/monsters
Squares used as fences/compounds/paddocks to deliminate an area (e.g. gardens, parks)	Electronic devices (e.g., Phones/tablets/laptops)
Stationary (e.g., cards, photos, papers, pictures)	Stationary
Furniture (e.g., chairs, tables, cupboards)	Furniture
Playground equipment (e.g., slides, swings)	
Vehicles (e.g., cars, lorries, vans)	Vehicles
Squares used as faces of animals/people/monsters	