Supplementary Material 2

Results Study 1

Omitted items. There was a significant main effect of gender, $F_{3, 660} = 39.34$, $p < .001$, $\eta^2 = .15$, and condition, $F_{3, 660} = 239.80$, $p < .001$, $\eta^2 = .52$, as well as a significant interaction between the two factors, $F_{3, 660} = 3.72$, $p = .011$, $\eta^2 = .17$, on the number of omitted items. Between-subjects effects for each domain revealed that gender had a main effect on the number of omitted items only in the numerical reasoning task; verbal: $F_{1, 662} = 0.03$, $p = .861$, $\eta^2 < .001$; numerical: $F_{1, 662} = 111.27$, $p < .001$, $\eta^2 = .14$; figural: $F_{1, 662} = 1.37$, $p = .242$, $\eta^2 < .01$. The time condition had a main effect on the number of omitted items only in the verbal reasoning, $F_{1, 662} = 559.56$, $p < .001$, $\eta^2 = .46$, and numerical reasoning tasks, $F_{1, 662} = 28.44$, $p < .001$, $\eta^2 = .04$, but not in the figural reasoning task, $F_{1, 662} = 1.56$, $p = .21$, $\eta^2 < .01$. A significant interaction between the two factors was found for the number of omitted items only in the numerical task, $F_{1, 662} = 8.70$, $p = .003$, $\eta^2 = .01$, but not in the verbal, $F_{1, 662} = 0.11$, $p = .735$, or figural, $F_{1, 662} = 0.22$, $p = .642$, reasoning tasks.

Accuracy index. Concerning the accuracy index, there was a significant main effect of only gender, $F_{3, 660} = 3.96$, $p = .009$, $\eta^2 = .018$, but not condition, $F_{3, 660} = 2.\ 02$, $p = .110$, $\eta^2 = .009$, and there was no significant interaction between the two factors, $F_{3, 660} = 1.47$, $p = .222$, $\eta^2 = .007$. Between-subjects effects for each domain revealed that gender had a main effect on the accuracy index in all reasoning tasks; verbal: $F_{1, 662} = 8.62$, $p = .003$, $\eta^2 = .013$; numerical: $F_{1, 662} = 6.52$, $p = .011$, $\eta^2 = .010$; figural: $F_{1, 662} = 5.55$, $p = .019$, $\eta^2 = .008$. The time condition had no main effect on the accuracy index in all tasks; verbal: $F_{1, 662} = 0.05$, $p = .828$, $\eta^2 < .001$; numerical: $F_{1, 662} = 2.79$, $p = .096$, $\eta^2 = .004$; figural: $F_{1, 662} = 3.06$, $p = .081$, $\eta^2 = .005$. A significant interaction between the two factors was not found; verbal, $F_{1, 662} < 0.01$, $p = .993$, $\eta^2 < .001$; numeric: $F_{1, 662} = 1.92$, $p = .167$, $\eta^2 = .003$; figural, $F_{1, 662} = 0.91$, $p = .341$, $\eta^2 = .001$.

Criterion validity.

In order to test whether criterion validity was differed depending on time constraints, and/or sex, we first checked for mean difference in the respective predictor (results see Table 1) and the respective criterion. We performed a 3 x 2 multivariate variance analysis with grades in the three domains (German, math, and science) as the dependent variables and time constraints as the independent variable. There was a significant main effect of time constraints ($F_{3, 662} = 5.17$, $p = .002$, $\eta^2 = .023$) on the dependent variable. Subsequently performed univariate variance analysis revealed that the two groups did not differ in German ($F_{1, 664} = 2.42$, $p = .119$) and math ($F_{1, 664} = 0.70$, $p = .402$) but slightly in science grades ($F_{1, 664} = 4.76$, $p = .030$, $\eta^2 = .007$). However, as the latter effect did not even reach a small effect size, it should be ignored. We additionally performed a 3 x 2 multivariate variance analysis with grades in the three domains (Germn, math, and science) as the dependent variables and gender as the independent variable. Gender had a significant main effect on the dependent variable ($F_{3, 662} = 5.72$, $p = .001$, $\eta^2 = .027$). Subsequently performed univariate variance analysis revealed that gender did not differ in math ($F_{1, 664} = 2.67$, $p = .103$) and science ($F_{1, 664} = 2.67$, $p = .103$) but slightly in German grades ($F_{1, 664} = 8.23$, $p = .004$, $\eta^2 = .012$). Girls had better German grades than boys. Thus, gender differences in the predictor were not reflected in the criterion which has been reported before (e.g., Steinmayr & Spinath, 2009). Second, we performed simple slope analyses to test for possible interaction effects. First we considered the time limit, second gender as a moderator in both the speed and power condition, respectively. Both time limit and sex were treated as dummy variables. Before performing interaction analyses, all predictors were centred (cf. Aiken & West, 1991). In accordance with the guidelines by Aiken and West (1991), we chose one standard deviation above the mean, and one standard deviation below the mean to compute simple regression equations of grades in German, math, and science, respectively, on verbal, numeric, and figural, respectively, intelligence. The time constraint condition did not significantly moderate the association

between any of the domain-specific associations between grades and reasoning (verbal: *B* = .002, *SD Error* = .012, *t* = 0.17, *p* = .87; numeric: *B* = -.002, *SD Error* = .008, *t* = -0.34, *p* = .74; figural: *B* = -.0001, *SD Error* = .010, *t* = -0.009, *p* = .99). Gender did neither moderate the associations between grades and reasoning in the speeded (verbal: *B* = -.006, *SD Error* = .017, *t* = -0.35, *p* = .73; numeric: *B* = .013, *SD Error* = .013, *t* = 0.99, *p* = .32; figural: *B* = .015, *SD Error* = .016, *t* = 0.93, *p* = .35) nor in the power (verbal: *B* = .030, *SD Error* = .017, *t* = 1.70, *p* = .09; numeric: *B* = .002, *SD Error* = .011, *t* = 0.21, *p* = .84; figural: *B* = .014, *SD Error* = .013, *t* = 1.01, *p* = .31) condition.