Supplemental Material for Wiedermann, Artner, and von Eye, Heteroscedasticity as a Basis of Direction Dependence in Reversible Linear Regression Models, Multivariate Behavioral Research

Supplementary Subgroup Analyses

To be placed on MBR website.

In this supplement, we present results of a subgroup analysis on the direction of effect of Analog Magnitude Code (AMC) and Auditory Verbal Code (AVC) ability scores obtained from 341 second to fourth grade elementary school children. Subgroup analyses were performed to address a potential ceiling effect in AMC scores which may affect results of Direction Dependence Analysis (DDA). Descriptive analyses based on the entire sample showed that 33 out of 341 children (i.e., 9.7%) obtained a maximum AMC score. Although this is below the cut-off of 15 – 20% commonly used to determine the presence of a ceiling effect, we re-ran linear regression models and DDA for a subgroup of the sample. Because AMC ability scores systematically increased with school grade, we only included second- and third-grade children (n = 221). Figure S1 shows the distributions and scatterplots of AMC and AVC scores for the sub-sample. Both, AMC and AVC scores, deviate from normality according to the Shapiro-Wilk test (both p's < .001). Thus, distributional requirements for DDA are fulfilled for the subgroup. For AVC, again no subject obtained the maximum score. For AMC, the portion of subjects reaching the maximum scores decreased to 6.3%.

Again, for the target model, AVC was regressed on AMC while adjusting for age (in years), the amount of time to complete the test (in minutes), and preexisting difficulty with numbers (0 = no, 1 = yes). Regression diagnostics for the target model suggested normality of the error term (Shaprio-Wilk's W = 0.99, p = .904) and empirically confirmed the absence of outliers (largest studentized deleted residual = –3.160, Bonferroni adjusted p = .391). However, five observations again showed hat-values larger than three times the average hat-value (one of the five observations also showed the largest Cook's distance of 0.161). To lower the impact of highly influential scores, these five observations were excluded from further analyses. Results of the target model (Model I: AMC → AVC) and the alternative model (Model II: AVC → AMC)

for the remaining 216 subjects are shown in Table S1. Scatterplots together with the corresponding simple and multiple linear regression lines are given in Figures S1c and S1d.

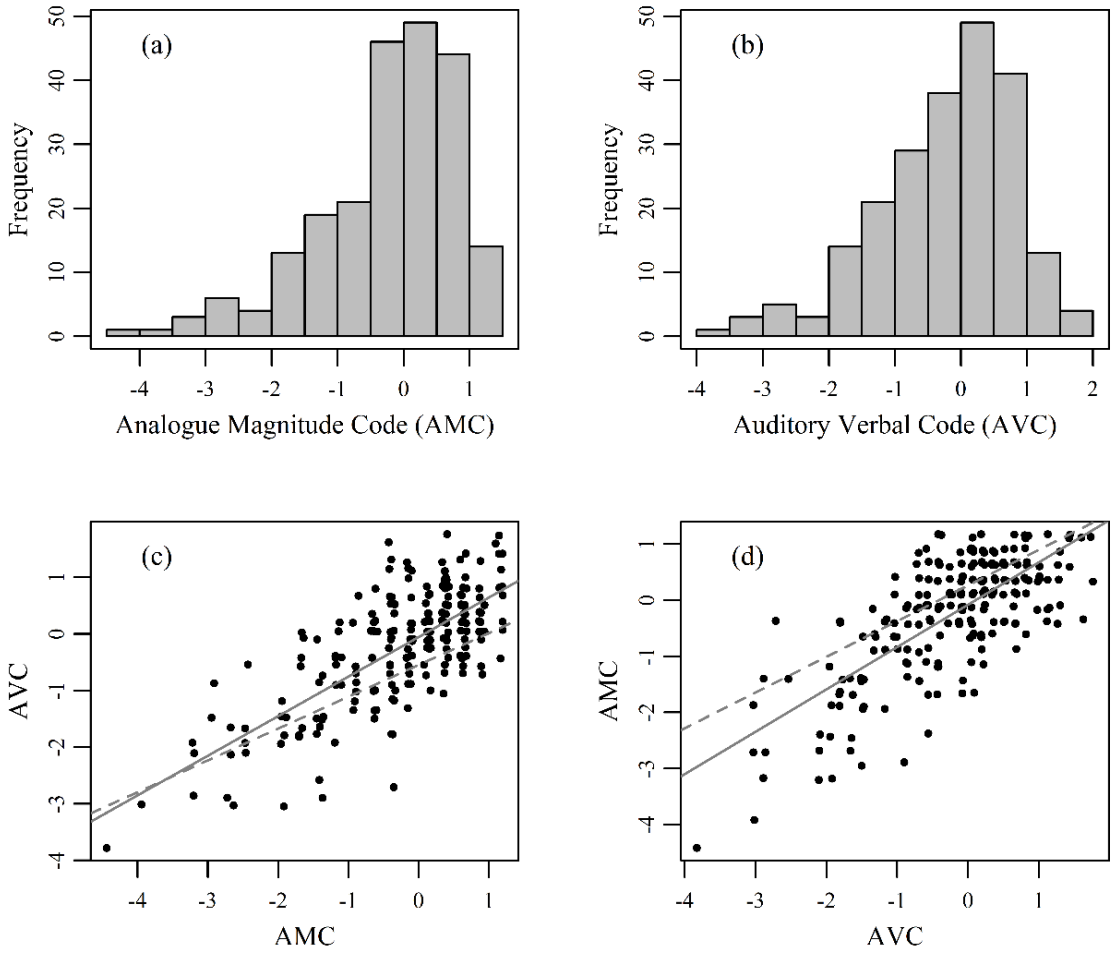

Figure S1: Distributions and scatterplots of AMC and AVC (solid and dashed lines refer to simple and multiple linear regression lines).

Next, we ask questions concerning the direction of effect. Visual diagnostics and homoscedasticity tests were used to evaluate the assumption of constant error variance for both competing models. Figure S2 shows the estimated regression residuals as a function of the predicted values for the two models. In general, residuals obtained from the target model do not show a clear trend over the range of predicted values. In contrast, a triangle-like pattern is again

observed for the alternative model. Thus, violations of the homoscedasticity assumption are more likely to occur in the alternative model which uses AVC as the outcome variable.

Table S1: Subgroup Results of competing multiple linear regression models

Source	β	S.E.	*t*-value	*p*-value
Model I: AMC → AVC (multiple R^2 = 0.539)				
Analog Magnitude Code (AMC)	0.51	0.06	9.18	<.001
Age in years	0.22	0.07	3.31	0.001
Amount of time to complete the test	–0.02	0.01	–2.18	0.031
Preexisting difficulties with numbers	–0.33	0.11	–3.07	0.002
Model II: AVC → AMC (multiple R^2 = 0.517)				
Auditory Verbal Code (AVC)	0.57	0.06	9.18	<.001
Age in years	0.06	0.07	0.90	0.370
Amount of time to complete the test	–0.03	0.01	–2.90	0.004
Preexisting difficulties with numbers	–0.27	0.12	–2.34	0.020

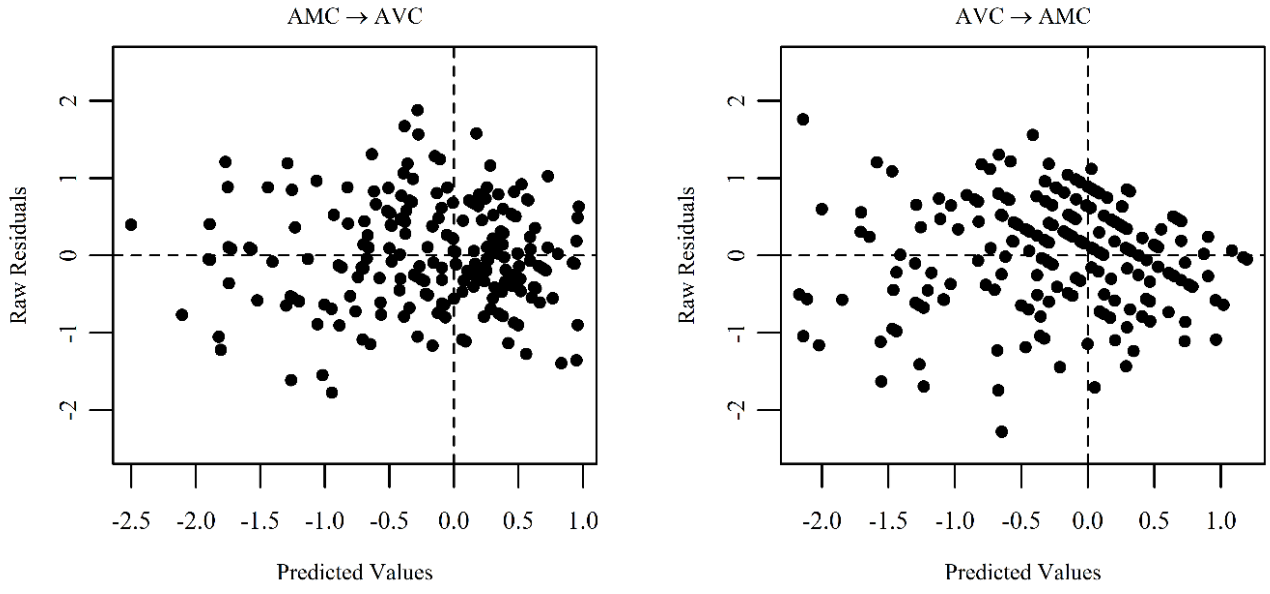

Figure S2: Estimated residuals and predicted values of both competing models.

To complete the analysis, we applied the nine homoscedasticity tests and used the proposed decision rules for model selection (see Table S2). Overall, we arrive at the same conclusions as reported for the entire sample: Six out of nine tests suggest that a model of the form AMC →AVC is more likely to approximate the data generating process. Note that for the other three procedures, no distinct decisions can be made. Specifically, Goldfeld-Quandt's and Harrison-McCabe's procedures reject the null hypothesis for both models and Park's test retains the null hypotheses for both models. Taken together, the majority of test results are in accordance with the hierarchical development of the triple code model which suggests that the AMC reflects a core system which is necessary to develop the AVC (i.e., AMC →AVC).

Table S2: Result of homoscedasticity tests for the two competing multiple linear regression models based on subgroup analyses.

	Model I	Model II	
Significance test	Response: AVC Predictors: AMC Age Time Difficulty	Response: AMC Predictors: AVC Age Time Difficulty	Decision based on 5% nominal significance level
Breusch-Pagan	$\chi^2(4) = 3.83$, $p = .430$	$\chi^2(4) = 13.28$, $p = .010$	AMC → AVC
robust Breusch-Pagan	$\chi^2(4) = 4.04$, $p = .400$	$\chi^2(4) = 13.93$, $p = .008$	AMC → AVC
Goldfeld-Quandt	$F(101, 100) = 1.65$, $p = .013$	$F(101, 100) = 4.58$, $p < .001$	undecided
Harrison-McCabe	$HMC = 0.38$, simulated $p = .006$	$HMC = 0.36$, simulated $p < .001$	undecided
White	$\chi^2(12) = 17.83$, $p = .121$	$\chi^2(12) = 34.83$, $p < .001$	AMC → AVC
Glejser	$\beta = –0.039$, $t = –1.41$, $p = .161$	$\beta = –0.115$, $t = –4.14$, $p < .001$	AMC → AVC
Park	$\beta = 0.066$, $t = 0.13$, $p = 0.898$	$\beta = –1.023$, $t = –1.80$, $p = 0.075$	undecided
Szroeter	$z = 1.70$, $p = 0.090$	$z = 3.78$, $p < .001$	AMC → AVC
Horn	$d = –1.68$, $p = 0.093$	$d = –3.85$, $p < .001$	AMC → AVC