Supplementary Material

To the article:

Poloczek, S., Hammerstein, S. & Büttner, G. (in press). Children's mixed-rounding strategy use in computational estimation. *Journal of Numerical Cognition.*

[29.07.2021]

Content

If you are interested in any further information, please contact the first author (poloczek@paed.psych.uni-frankfurt.de).

The dataset and R-scripts to re-run the analyses reported in the article or to set up own further analyses can be found at www.psycharchives.org and are linked to this supplementary material file at Psycharchives.

1 Computational Estimation Experiment

1.1 Further Constraints in Item Construction

As in previous research on computational estimation (e.g. Lemaire & Brun, 2016; Lemaire & Lecacheur, 2011), estimation problems were created with the following constraints: (1) operands never included 0 or 5 as unit digit, (2) no pair of operands included equal unit digits (e.g., never 32 + 62), (3) no pair of operands would result in equal decade digits when using the rounding strategy indicated by the main category (e.g., never 39 + 42, because it would lead to 40 + 40), (4) no trials with reverse order of the operands were presented (e.g., not 68 + 24 in one trial and 24 + 68 in another) and (5) no rounded operand was equal to 0, 10 or 100 if rounded with the best strategy (i.e., range of operands from 16 to 94). Therefore, problem sizes of the estimation items ranged from estimates of 50 to 170.

Table S1. Exact Pseudorandom Sequences of Computational Estimation Problems

Testblock 1a		**Testblock 1b**		**Testblock 2a**		**Testblock 2b**	
Problem	**Category**	**Problem**	**Category**	**Problem**	**Category**	**Problem**	**Category**
48 + 72	mixed2	42 + 18	mixed2	43 + 61	small1	29 + 42	mixed3
22 + 57	mixed1	43 + 48	mixed3	68 + 16	large1	59 + 87	large3
62 + 54	small3	18 + 29	large3	39 + 48	large3	79 + 16	large2
84 + 71	small2	78 + 37	large2	94 + 32	small3	53 + 94	small3
56 + 28	large1	54 + 93	small3	47 + 19	large3	22 + 74	small3
77 + 29	large3	81 + 24	small2	68 + 59	large3	43 + 22	small2
47 + 28	large2	27 + 73	mixed2	72 + 93	small2	17 + 58	large2
93 + 77	mixed2	41 + 76	mixed1	54 + 33	small3	87 + 76	large1
32 + 83	small2	93 + 31	small1	62 + 81	small1	46 + 84	mixed2
19 + 76	large2	41 + 63	small1	64 + 28	mixed3	16 + 93	mixed1
38 + 86	large1	63 + 24	small3	91 + 47	mixed1	92 + 59	mixed3
88 + 43	mixed3	27 + 36	large1	19 + 43	mixed3	41 + 32	small1
23 + 52	small2	69 + 88	large3	78 + 47	large2	16 + 37	large1
34 + 62	small3	76 + 57	large1	46 + 58	large1	36 + 69	large2
81 + 92	small1	87 + 52	mixed1	27 + 51	mixed1	73 + 26	mixed1
66 + 21	mixed1	39 + 34	mixed3	54 + 36	mixed2	28 + 94	mixed3
32 + 21	small1	59 + 37	large3	71 + 43	small1	31 + 74	small2
74 + 79	mixed3	56 + 69	large2	84 + 41	small2	93 + 69	mixed3

Note. Items were presented to all participants in the same pseudorandom, but carefully balanced order (in each testblock from top to bottom).

1.2 Sequence of Trials

All children received the same 72 items in the same pseudorandom order. As we were interested in comparing strategy selection and estimation latencies between categories of problems, the pool of estimation problems as well as the pseudorandom test order were carefully constructed to reduce the impact of potential confounding variables. The 72 items were distributed onto the four test blocks so that test blocks contained six problems of each main category and that the blocks were well matched on the other dimensions listed above. Within each test block, problems were put into random order. This random sequence was adjusted so that there were about 50% of trials in which the same main task category (small, mixed, large) was repeated. If children always or mostly used the rounding strategy suggested by the main category, the number of strategy repetitions and switches was also balanced or nearly balanced. Additionally, the split between suggested strategy repetition and suggested strategy switches was similar for all nine subcategories. Finally, after rounding no two consecutive trials could be the same (e.g., 61 + 43 could not be followed by 57 + 41 because both would result in 60 + 40; for the exact sequence of items see Table S1).

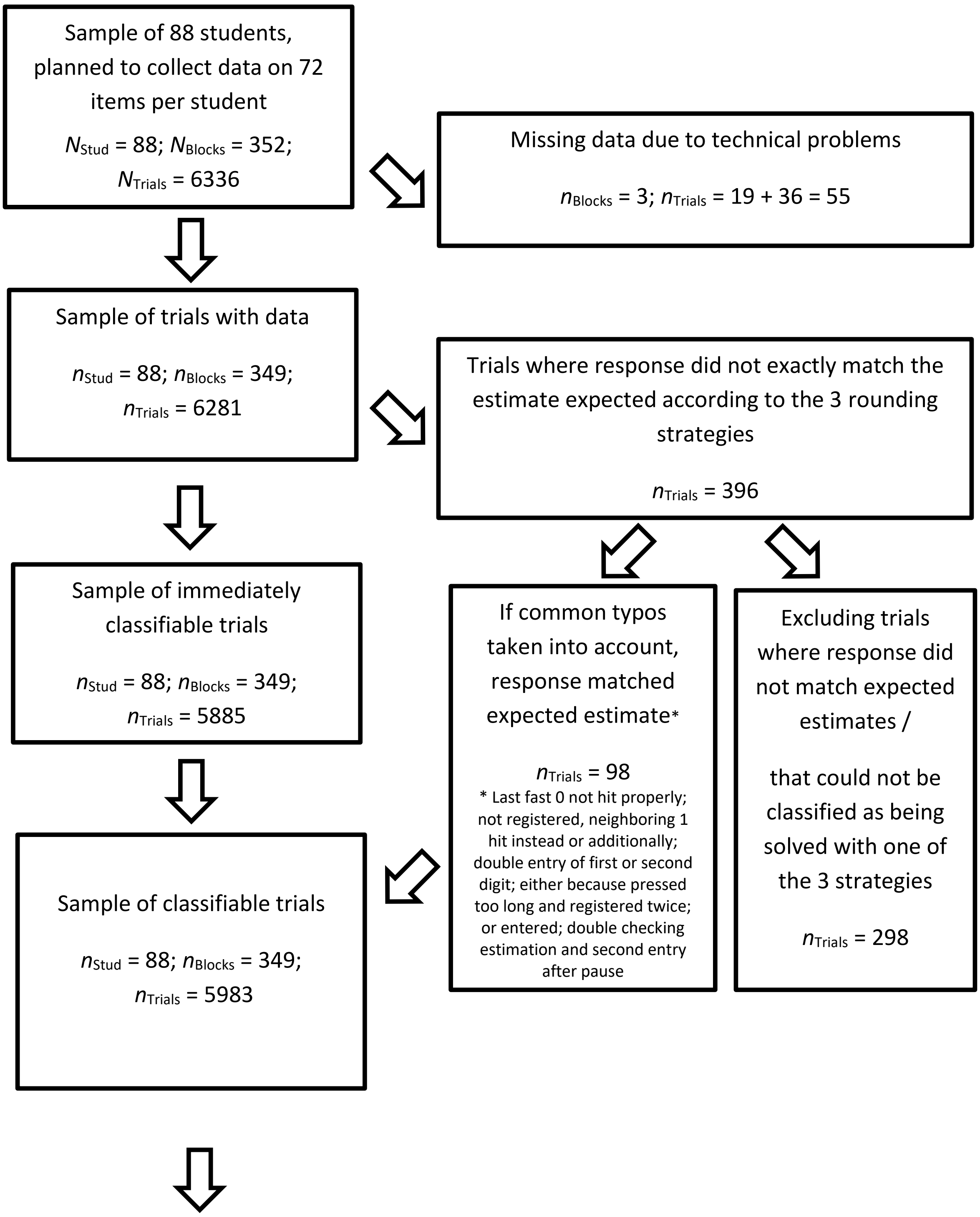
1.3 Inferring Strategy Use

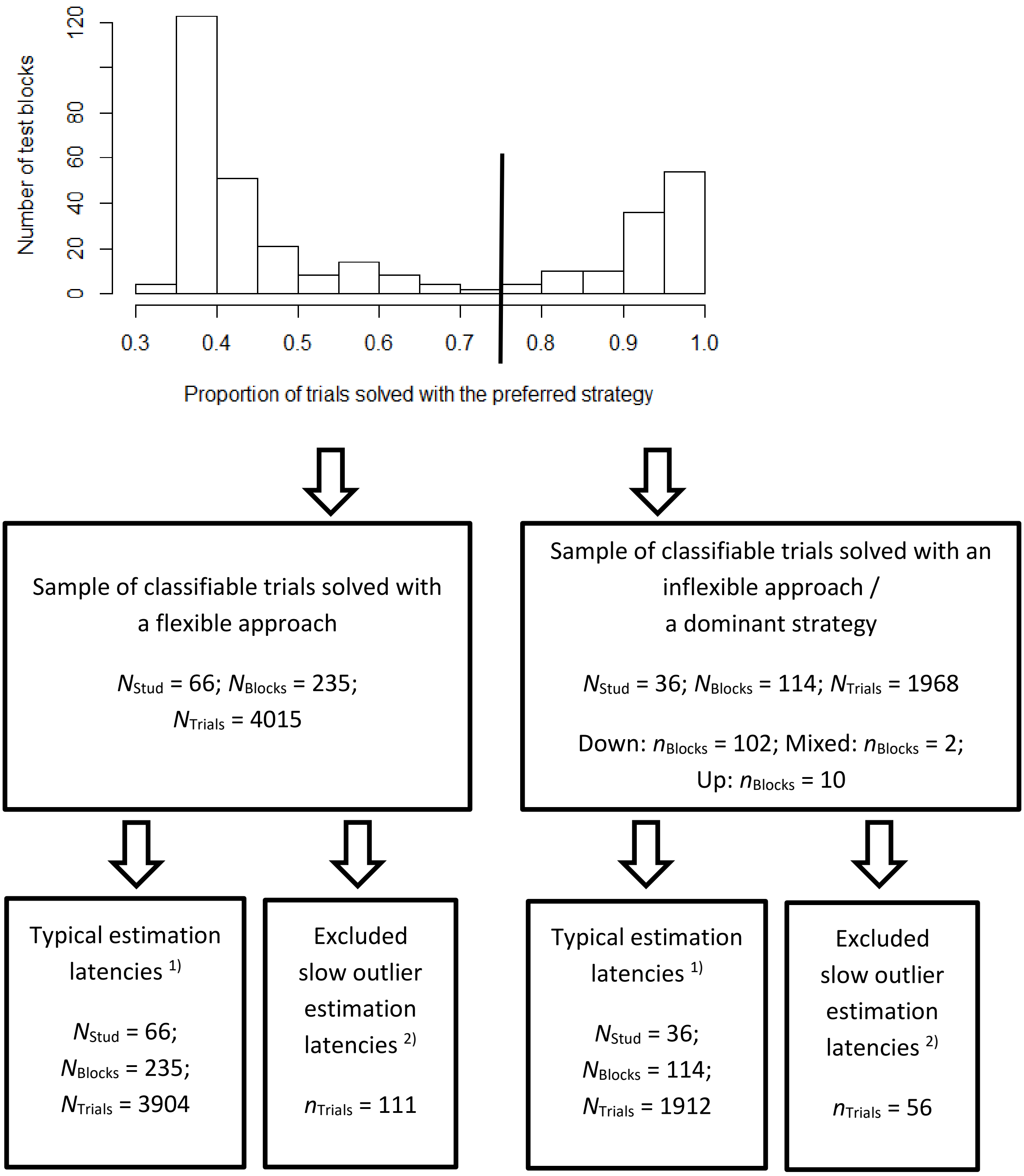
The rounding strategy selected by a child for a given trial was inferred from the estimate for that trial. If for example a child typed in 70 as response to 22 + 57, the strategy was classified as rounding-down; if the response was 80 as mixed-rounding; and if the response was 90 as rounding-up. In 93.7% of all 6281 trials the response directly matched one of expected estimates. Of the remaining 396 responses 98 very likely were due to common typos like not properly hitting the last 0 digit (entries like 71 or 710 or 7 for the example above) or having either the first digit or the second digit logged twice (entries like 770 or 177 for problems in which 70 or 170 were plausible estimates). The syntax to classify responses was extended so that trials with common typos also were classified, resulting in 95.2% of trials having a strategy classification.

To assess whether determining strategy choice based on the estimates would likely lead to misclassifications, we examined children's calculation errors during a pure calculation task of two-digit additions (e.g., 40 + 70; data gathered to estimate children's addition speed, as this task was not included in the analyses for the paper, it has not been described further). Within a total of 1056 trials across 88 participants, 93.9% of entered responses were correct, a further 2.7% contained an error that would not lead to a strategy-misclassification (e.g., first digit not registered, typo or calculation error not close to correct result). The percentage of +10 errors was 1.6% (i.e., answers with 10 units above the correct result, e.g., 120 instead of 110) and 0.9% for the -10-errors (i.e., answers with 10 units below the correct result). These errors would result in classification errors, if they were paired with a mixed-unit problem and would lead to classification errors with a 50% chance when paired with small-unit or large-unit problems. +20-errors and -20-errors were very rare (0.3% and 0.5%, respectively) and just one in three of those errors would lead to a misclassification. Therefore, inferring strategy use from the given responses in the computational estimation task should only be associated with a small error.

1.4 Flowchart Inclusion/Exclusion of Trials

Figure S1. Flowchart illustrating decisions to exclude and include trials.





Note. To reduce the impact of outliers on the analyses of estimation latencies, only latencies within the boundaries of +/-2.5 standard deviations (computed for each individual) around the individuals mean latency were included (considered typical estimation latencies). No fast outliers were present. Slow outliers were present on 2.7% of trials and were excluded.

1) RTs or estimation latencies that were within +/- 2.5 (individual) SDs from the individual's mean
2) RTs or estimation latencies that were slower than + 2.5 (individual) SDs than the individual's mean

2 Further Details on the Results-Section

2.1 Dummy Coding to Compare Main Task Categories and Subcategories

The set of eight predictors for the comparison between nine types of task was constructed in a way to be able to compare strategy use at the same time

- between main task categories
  (small vs. mixed vs. large unit problems)
- and within main task categories
  but between subcategories
  (small1 vs. small2 vs. small3 |
  mixed1 vs. mixed2 vs. mixed3 |
  large1 vs. large2 vs. large3).

To achieve this the predictors were set-up as demonstrated in the Table S2.

The first two predictors ensure, that all mixed-unit problems are contrasted with all small-unit problems as main reference category, and likewise large-unit problems are contrasted with small-unit problems.

In order to compare the three subcategories within the main categories, for each main category two subcategory predictors were created, e.g. both small2 and small3 problems (both can be either solved by rounding down or mixed rounding) are contrasted with the reference category small1 (twice 0 as dummy predictor; reference category as rounding down unambiguously best strategy).

For each of the nine subcategories a unique pattern of 8 predictor values exists. All predictors are independent; none can be expressed as a function of the other predictors.

Table S2. Set of Predictors to Compare Strategy Use Between Main Task Categories and Within Task-Categories to Examine the Potential Effect of Unit Sums

Estimation problems		Main category predictors		Subcategory predictors					
Main categories	**Sub-categories**	**Mixed** (vs. small)	**Large** (vs. small)	**Small2** (vs. small1)	**Small3** (vs. small1)	**Mixed1** (vs. mixed2)	**Mixed3** (vs. mixed2)	**Large1** (vs. large3)	**Large2** (vs. large3)
small	**small1**	0	0	0	0	0	0	0	0
	small2	0	0	1	0	0	0	0	0
	small3	0	0	0	1	0	0	0	0
mixed	**mixed1**	1	0	0	0	1	0	0	0
	mixed2	1	0	0	0	0	0	0	0
	mixed3	1	0	0	0	0	1	0	0
large	**large1**	0	1	0	0	0	0	1	0
	large2	0	1	0	0	0	0	0	1
	large3	0	1	0	0	0	0	0	0

2.2 Calculating Predicted Probabilities with 95% Credible Intervals

The predicted probability of choosing a certain strategy for item *i* by participant *j* is

$$\pi_{ij} = \frac{\exp\left(\beta_0 + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + u_i + u_j\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + u_i + u_j\right)}$$

where, β_0, β_1 up to β_k for the k^{th} predictor are replaced by the estimates from the fitted GLMM, u_i by the random item intercept residuals and u_j by the random participant intercept residuals.

The predicted probability, that an average person (u_j = 0) will choose rounding-down when giving an estimate for an average item (u_i = 0) that belongs to the subcategory *small1* (all predictors x_1 to x_k = 0) would only include the intercept (see Table 2 in the paper); so, it would be $\exp(3.20)/(1+\exp(3.20)) = 0.961$.

The formula for the predicted probability for an average person and an average item belonging to the subcategory *mixed1* would additionally include the estimates for main category *mixed*-unit and the subcategory *mixed1*: $\exp(3.20 - 5.70 + 0.19)/(1 + \exp(3.20 - 5.70 + 0.19)) = 0.090$.

To estimate the mean predicted probability for our population of participants and items it is not sufficient to just take the values estimated for an average person and item. Due to the non-linear inverse-logit transformation the mean predicted probability is not equal to the predicted probability of an average participant, the median predicted probability (see Steele, 2009). Mean and median predicted probabilities can be close if probabilities are in the range of 0.20 to 0.80 and if random intercept variance is low. But given the predicted probabilities were more extreme in this study and given that there was considerable random participant variance, mean predicted probabilities with a population-averaged interpretation were calculated. To do so, for each of the 9 task-subcategories 1000 predicted probabilities were calculated, each with a different value for u_i and u_j. These values were drawn from a random distribution with *M* = 0 and variances corresponding to the random item and the random participant variance of the model (compare Steele, 2009). Taking the mean of the 1000 predicted probabilities yields the probability with the population-averaged interpretation.

To provide 95% credible intervals around the mean, this process of simulating 1000 predictions and computing the mean was not performed once with the values provided in Table 2, but 5000 times with the value combinations of the 5000 stored MCMC iterations. Computing the mean of the 5000 mean predicted probabilities per subcategory and extracting the 2.5th and the 97.5th percentile of these 5000 predicted probabilities resulted in the reported mean predicted probabilities with 95% credible intervals.

We are grateful to William Browne, Centre for Multilevel Modelling at the University of Bristol, for his valuable comments on our thoughts on computing credible intervals for predicted probabilities.

2.3 Complete Result Tables Best Strategy Selection

Table S3a. Logistic GLMM Examining Predictors of Best Strategy Use (Small-Unit as Reference)

Fixed Part	**β**	**95% CI**	***p***
Intercept	3.29	[2.74, 3.87]	< .001
Mixed-unit problem (vs. small-unit)	-0.88	[-1.23, -0.53]	< .001
Large-unit problem (vs. small-unit)	-1.11	[-1.46, -0.76]	< .001
Problem size (per 10 unit increase)	-0.09	[-0.13, -0.04]	< .001
Random Part	***u***	**95% CI**	**Δ *DIC***
Participant intercept variance	3.76	[2.36, 5.76]	863.3
Item intercept variance	0.12	[0.02, 0.27]	14.5

Note. 95% CI = 95% credible interval; Δ DIC = change in Deviance Information Criterion if random intercept dropped from model. Unless Δ DIC is small (below 5) including the random effect variance improves model fit (see Zhang et al., 2016).

Table S3b. Re-Parameterized Logistic GLMM Examining Predictors of Best Strategy Use (Large-Unit Problems as Reference Category)

Fixed Part	**β**	**95% CI**	***p***
Intercept	*2.19*	*[1.66, 2.74]*	*< .001*
Small-unit problem (vs. large-unit)	*1.10*	*[0.75, 1.45]*	*< .001*
Mixed-unit problem (vs. large-unit)	*0.23*	*[-0.10, 0.54]*	*.15*
Problem size (per 10 unit increase)	-0.09	[-0.13, -0.04]	< .001
Random Part	***u***	**95% CI**	**Δ *DIC***
Participant intercept variance	3.75	[2.41, 5.75]	*
Item intercept variance	0.12	[0.01, 0.28]	*

Note. 95% CI = 95% credible interval; * Δ DIC = change in Deviance Information Criterion not calculated, similar to Table S3a, as only re-parameterized model; *different parameter-estimates due to re-parameterization*

2.4 Complete Result Tables Estimation Latencies

Table S4a. LMM Examining Predictors of Estimation Latency (Rounding-Down as Reference)

Fixed Part	**β**	**95% CI**	***p***
Intercept (reference strategy rounding-down)	5.20	[4.84, 5.59]	< .001
Mixed-rounding (vs. rounding-down)	0.43	[0.23, 0.62]	< .001
Rounding-up (vs. rounding-down)	0.88	[0.67, 1.09]	< .001
Problem size (per 10 unit increase)	0.16	[0.12, 0.20]	< .001
Random Part	***u***	**95% CI**	**Δ *DIC***
Participant intercept variance	1.77	[1.23, 2.50]	1659.7
Item intercept variance	0.25	[0.16, 0.37]	251.9
Residual variance	2.77	[2.64, 2.90]	

Note. 95% CI = 95% credible interval; Δ DIC = change in Deviance Information Criterion if random intercept dropped from model. Unless Δ DIC is small (below 5) including the random effect variance improves model fit (see Zhang et al., 2016).

Table S4b. Re-Parameterized LMM Examining Predictors of Estimation Latency (Rounding-Up Strategy as Reference Category)

Fixed Part	**β**	**95% CI**	***p***
Intercept (reference strategy rounding-down)	*6.08*	*[5.71, 6.46]*	*< .001*
Rounding-down (vs. rounding-up)	*-0.88*	*[-1.09, -0.66]*	*< .001*
Mixed-rounding (vs. rounding-up)	*-0.45*	*[-0.64, -0.26]*	*< .001*
Problem size (per 10 unit increase)	0.16	[0.12, 0.20]	< .001
Random Part	***u***	**95% CI**	
Participant intercept variance	1.77	[1.23, 2.50]	*
Item intercept variance	0.25	[0.16, 0.37]	*
Residual variance	2.77	[2.65, 2.90]	

Note. 95% CI = 95% credible interval; * Δ DIC = change in Deviance Information Criterion not calculated, similar to Table S4a, as only re-parameterized model; *different parameter-estimates due to re-parameterization*

3 Guide to further files and variables

The ***data*** to reproduce the analyses or to run further analyses are available at psycharchives.org:

Poloczek-et-al-2021_JNC_Data_CompEst.csv

Poloczek-et-al-2021_JNC_Data-Codebook_CompEst.csv

Variable and value labels are also given in a more reader friendly format in Table S5a and Table S5b.

R-code to reproduce the analyses is also available at psycharchive.org.

Poloczek-et-al-2021_JNC_Analyses_StrategicFlexibility.R
to reproduce the distribution of proportion of preferred strategy use and the analyses on strategic flexibility.

Poloczek-et-al-2021_JNC_Analyses_GLMM-SelectMix-Table2.R
to reproduce the GLMMs on which strategies were selected for which problem categories (Table 2) and
Poloczek-et-al-2021_JNC_Analyses_GLMM-SelectMix-PP-Figure1.R
to reproduce the predicted probabilities which are based on the analyses presented in Table 2 and which are presented in Figure 1.

Poloczek-et-al-2021_JNC_Analyses_GLMM-BestStrat-TableS3.R
to reproduce the GLMMs on best strategy use which are presented in the result section
(and in Tables S3a & S3b) and
Poloczek-et-al-2021_JNC_Analyses_GLMM-BestStrat-PP-Text.R
to reproduce the predicted probabilities which are based on the analyses in Table S3.

Poloczek-et-al-2021_JNC_Analyses_LMM-EstLatencies-TableS4.R
to preproduce the LMMs on estimation latencies which are presented in the result section
(and in Tables S4a & S4b)

Table S5a. Guide to Variables in Dataset (Variable Labels and Variable type)

Variable name	Variable label	Variable type
scode	ID for each student	numeric
scodeXtrial	ID for each trial taken by a student	numeric
sex_ch	Sex of the student	string
age_month	Age of the student (in month)	numeric
cest_tb	Test block in Computational Estimation experiment	numeric
cest_trial	ID for each presented estimation problem	numeric
cest_stim	Stimulus - presented estimation problem	string
cest_sumex	Exact sum of addition task	numeric
cest_tcat3	Main task category of estimation problem	string
cest_tcat9	Subcategory of estimation problem	numeric
cest_tcat9s	Subcategory of estimation problem (string)	string
cest_tswi	Task category switch (different main category to preceding problem)	string
cest_estop	Optimal problem-based estimate (according to main category)	numeric
cest_prsi_best_c	Problem size of estimate (based on best strategy - centered - divided by 10)	numeric
cest_tcat_sm	Small-unit problem	numeric
cest_tcat_mi	Mixed-unit problem	numeric
cest_tcat_la	Large-unit problem	numeric
cest_tcat_sm2	Small-unit problem subcategory 2 (unit sum 5)	numeric
cest_tcat_sm2	Small-unit problem subcategory 2 (unit sum 5)	numeric
cest_tcat_sm3	Small-unit problem subcategory 3 (unit sum 6 or 7)	numeric
cest_tcat_mi1	Mixed-unit problem subcategory 1 (unit sum 7 to 9)	numeric
cest_tcat_mi3	Mixed-unit problem subcategory 3 (unit sum 11 to 13)	numeric
cest_tcat_la1	Large-unit problem subcategory 1 (unit sum 13 or 14)	numeric
cest_tcat_la2	Large-unit problem subcategory 2 (unit sum 15)	numeric

Variable name	**Variable label**	**Variable type**
cest_resp	Response to estimation problem	numeric
cest_rcat	Response category - rounding strategy	string
cest_accl	Accurate response (lenient scoring - all rounding results accurate - not only best)	numeric
cest_RTms	Response latency (in ms)	numeric
cest_rcat_do	Trial solved with rounding-down	numeric
cest_rcat_mi	Trial solved with mixed-rounding	numeric
cest_rcat_up	Trial solved with rounding-up	numeric
cest_rcat_best	Trial solved with best strategy (optimal problem-based strategy)	numeric
cest_sswi	Strategy switch (different strategy to preceding trial used)	string
cest_prefStr_pr	Proportion of preferred strategy within test block	numeric
cest_prefStr_4c	Preferred strategy within test block	string
cest_domStr	Dominant strategy within test block (more than 75% of trials with preferred strategy)	numeric
cest_sRTacc	Estimation latency (in s) (only trials with classifiable strategy - outliers removed)	numeric

Table S5b. Guide to Variables in Dataset (Value Labels for String Variables)

Variable name	Value labels
sex_ch	"f" = "female", "m" = "male"
cest_tcat3	"la" = "large-unit problem", "mi" = "mixed-unit problem", "sm" = "small-unit problem"
cest_tcat9s	"sm1" = "small-unit subcat. 1", "sm2" = "small-unit subcat. 2", "sm3" = "small-unit subcat. 3", "mi1" = "mixed-unit subcat. 1", "mi2" = "mixed-unit subcat. 2", "mi3" = "mixed-unit subcat. 3", "la1" = "large-unit subcat. 1", "la2" = "large-unit subcat. 2", "la3" = "large-unit subcat. 3" *
cest_tswi	"rep" = "Repetition of main task category", "swi" = "Switch of main task category", "NA" = "not applicable (no preceding trial)"
cest_rcat	"do" = "rounding-down", "mi" = "mixed-rounding", "up" = "rounding-up", "xx" = "not classifiable"
cest_sswi	"rep" = "Repetition of strategy", "swi" = "Switch of strategy", "NA" = "not applicable (no preceding trial or strategy not classifiable)"
cest_prefStr_4c	"do" = "rounding-down", "mi" = "mixed-rounding", "up" = "rounding-up", "shared pref" = "shared preference"

*To interpret the subcategories, see Table 1 in the method section of the paper

4 References

Lemaire, P., & Brun, F. (2016). Age-related differences in children's strategy repetition: A study in arithmetic. *Journal of Experimental Child Psychology, 150,* 227–240. doi:10.1016/j.jecp.2016.05.014

Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: A study in computational estimation. *Cognitive Development, 26,* 282–294. doi:10.1016/j.cogdev.2011.01.002

Steele, F. (2009). Module 9: Single-level and multilevel models for ordinal responses concepts. In Centre for Multilevel Modelling (Eds.). *LEMMA (Learning environment for multilevel methods and applications.* University of Bristol. Available at: https://www.cmm.bris.ac.uk/lemma/