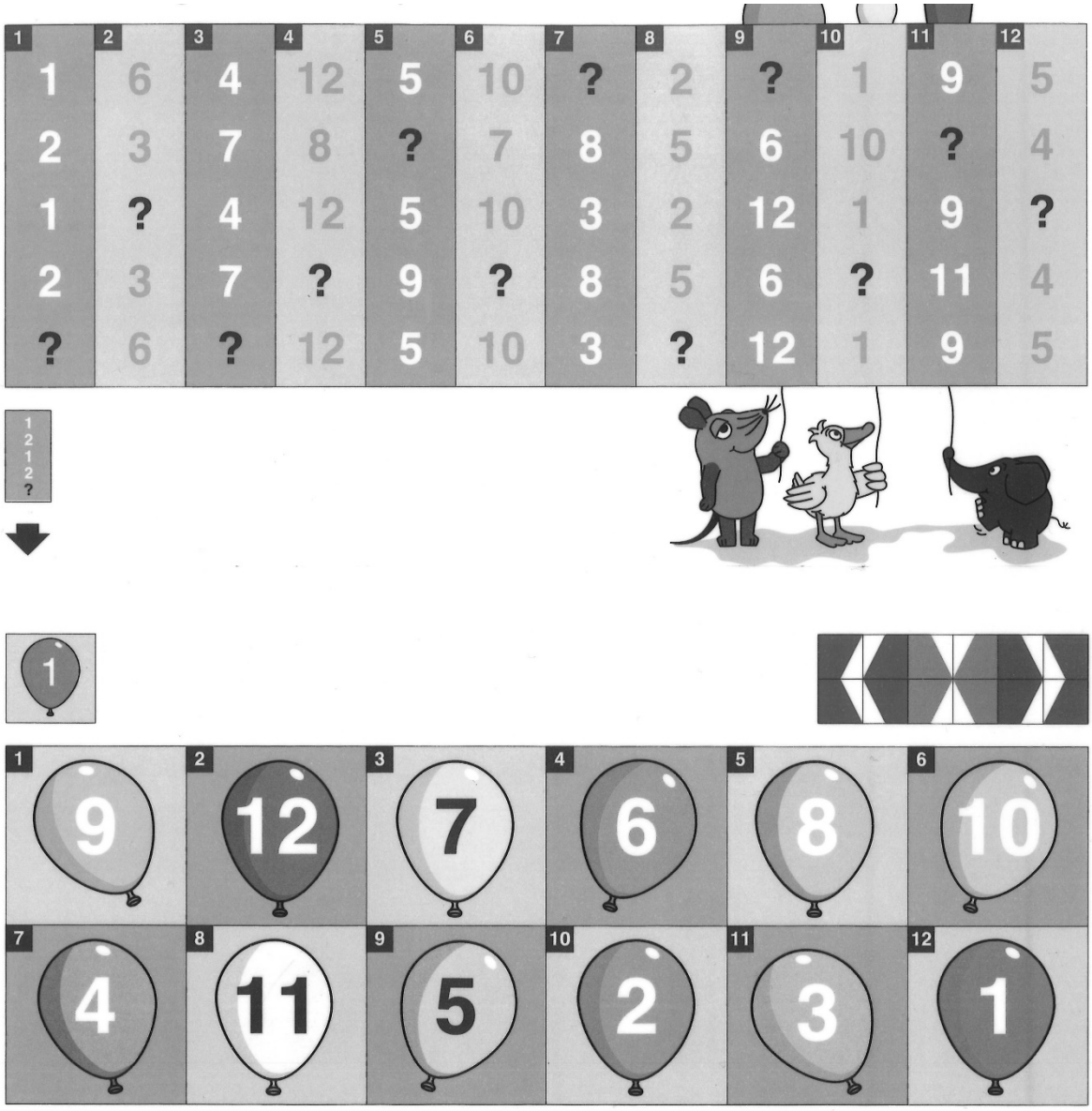
Appendix A

Task used in Experiment 1



Appendix B

Detailed Description of Procedure

The experimenter explained how the thermometer works by referring to the well-known children's game, "cold/hot hide and seek". The surer the children were, the darker the red they had to point to; the more unsure they were, the darker the blue they had to point to (as in Destan et al., 2017). The experimenter made sure the children had understood how to use the thermometer using two examples. "Are you a girl or a boy?" The child replied. "How sure are you that your answer was correct? Point to the colour." Then experimenter gave the child feedback: "Yes, you are a girl/boy. And you correctly pointed to the dark red." "Now, how sure are you that you will answer my next question correctly?" The child pointed to a colour. "How many leaves are there on that tree?" The child answered. "How sure are you that your answer was correct?" Then experimenter gave the child feedback: "You can't be very sure because we don't know the exact number of leaves. You should have shown me the dark blue." All the children understood how to use the thermometer.

The experimenter checked that the children could count to fifteen without making a mistake and then showed them an example of the numerical order task. The rule and the aim were explained to the children using an example. The children were given the items in the same order and they were shown them one-by-one.

Each child made 12 postdictive judgments and 11 predictive judgments (there was no predictive judgment after the final item). At the end, the children received a sticker for participating in the research.

Appendix C

Analytical Approach

In the results section, three main objectives were analysed. Firstly, the general effect of performance feedback on the absolute accuracy of predictive judgments was examined. To assess absolute accuracy, the average Bias Index for each participant was calculated. The Bias Index is based on the difference between judgment and performance and indicates the degree to which a child is overconfident or underconfident. The Bias Index for the predictions is based on the difference between the predictive judgment and the subsequent answer. The Bias Index ranges from -1 to 1. A value of 0 indicates accurate monitoring, while values lower than 0 indicate underconfidence and values higher than 0 indicate overconfidence. For example, when children predicted high confidence for the next item (dark red / very sure) and solved the item correctly, their Bias Index would be 0, indicating accurate monitoring. On the contrary, when they predicted high confidence but solved the item incorrectly, their Bias Index would be 1, indicating overconfidence (Schraw, 2009). To examine the effect of Performance Feedback on absolute accuracy a one-way ANOVA with Feedback Group (PF, NF) as the between-subject factor was conducted.

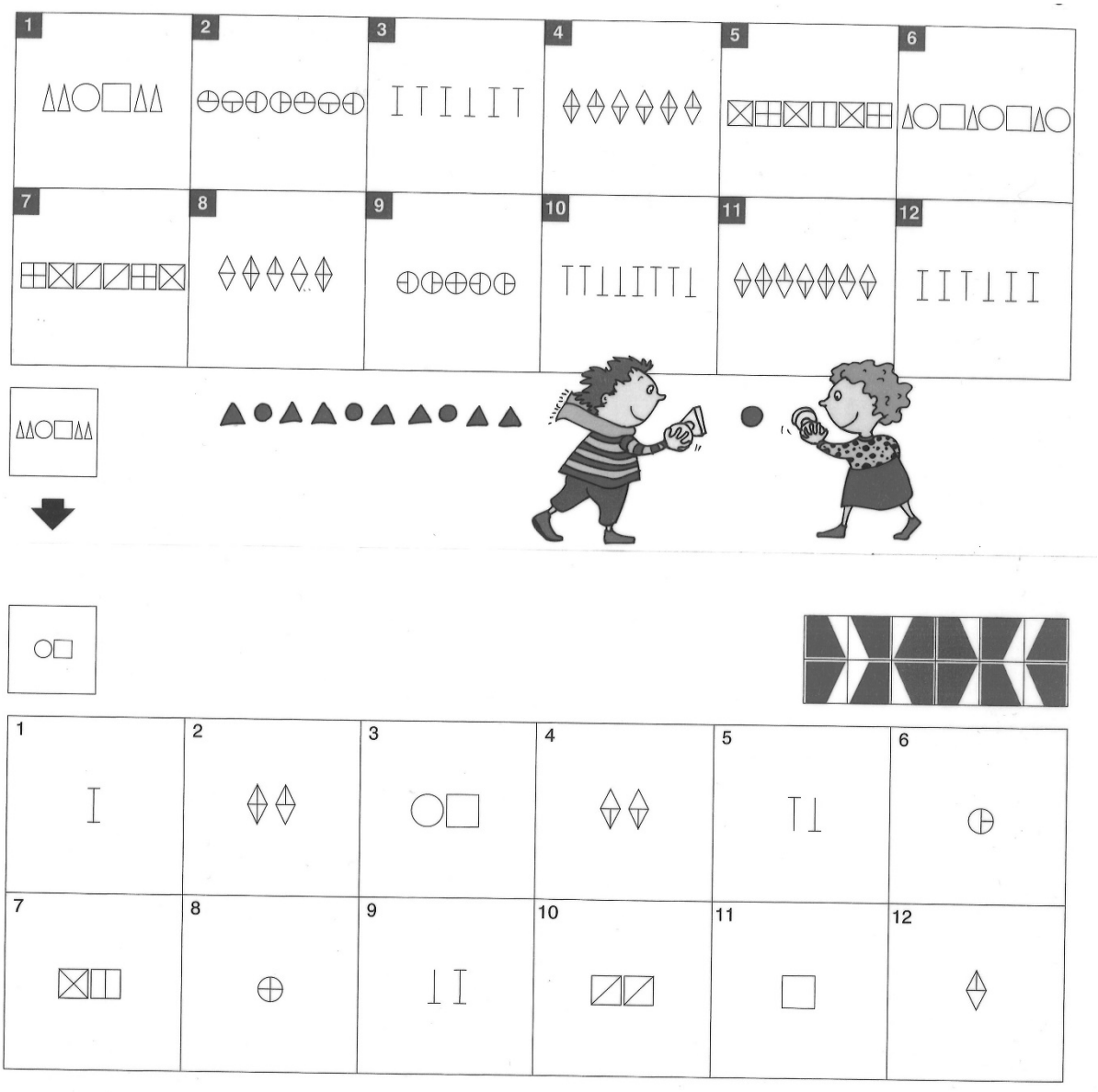
Secondly, the effect of performance feedback on the accuracy of uncertainty monitoring using predictive judgments (Model 1) was examined. Because of the nature of the data in which children could answer only correctly or incorrectly resulting in missing data where correct answers were made, a multilevel analysis (MLM) with Restricted Maximum Likelihood (REML) estimation was conducted. In order to achieve comparability between models, the data were standardized to *z*-scores. In models, Participant Numbers were utilized as the Subjects, an Item Number was used as a Repeated Condition. To build the model, the −2 restricted log likelihood estimate was used as the information criterion, and the model with the best fit was chosen as the final. All the models included random intercept and slope for the Feedback Group.

Finally, analyses to explore the anchoring effect of performance feedback were conducted. To examine the anchoring effect, the postdictive judgment given directly after the item and before the performance feedback was used. If performance feedback acts as an anchor, the effect of feedback should be transferred to the postdictive judgment that followed the performance feedback, despite the children obtaining task experience in the middle (see Figure 1 [in the main manuscript] for a description of the entire procedure, if the performance feedback acts as an anchor, the performance feedback after first task will influence the performance feedback after second task, despite the fact that children gave another predictive judgment and solved another item in the meantime). For example, when children solved the Item 1 incorrectly and obtained negative performance

feedback, performance feedback would act as an anchor when the postdictive judgment after the Item 2 would be lower, regardless of the Item 2 correct or incorrect solution. Multilevel analysis in Model 2 was used to examine the effect of the performance feedback on the postdictive judgments given immediately after an incorrect answer. Another multilevel analysis in Model 3 examined the effect of performance feedback on postdictive judgments following an incorrect answer and subsequent negative performance feedback. If the feedback effect was stronger in Model 3 than in Model 2 the anchoring effect hypothesis would be supported.

Appendix D

Task used in Experiment 2



References

Destan N., Spiess M. A., de Bruin A., van Loon M., & Roebers C. M. (2017). 6- and 8-year-olds' performance evaluations: Do they differ between self and unknown others? *Metacognition and Learning, 12*(2), 315–336. doi:10.1007/s11409-017-9170-5

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky & A. C. Graesser, (Eds.), *Handbook of metacognition in education* (pp. 415-429). New York, NY, USA: Routledge.