Supplementary A

We asked all participants to estimate their performance on the spatial (Fig. S1) and timing (Fig. S2) components of the motor task after each trial in retention and transfer, similar to past research (e.g., Barros et al., 2019; Carter et al., 2014; Carter & Patterson, 2012). Additionally, half of the participants in each group were randomly selected to also estimate their performance after each trial in the pre-test (not shown). To assess error estimation for the spatial and timing components, we first calculated the difference between a participant's estimated performance and their actual performance on each trial. Next, we computed total estimation error (EE) using the equation:

$$EE = \sqrt{CE^2 + VE^2} \tag{1}$$

where CE was the average estimation bias and VE was the standard deviation of these errors. This approach is consistent with that used by Bruechert et al. (2003).

Error estimation in retention and transfer

Total estimation error was generally lower in retention compared to transfer for both the spatial (Fig. S1) and timing (Fig. S2) domains in Experiments 1 and 2. For the spatial estimation error, we found a significant main effect of Test in Experiment 1, $F(1, 148) = 32.40$, $p < .001$, $\eta_G^2 = .060$, and in Experiment 2, $F(1, 74) = 18.79$, $p < .001$, $\eta_G^2 = .026$, with more accurate estimations in retention than transfer. Although there was a significant Choice x Feedback x Test interaction in Experiment 1 for timing estimation error, $F(1, 148) = 7.12$, $p = .008$, $\eta_G^2 < .001$, we did not decompose the interaction as the effect size estimate was less than .001. There were no significant main effects or interactions for timing estimation error in Experiment 2. A potential explanation for the difference in estimation accuracy between retention and transfer for the spatial component and not the the timing component is that only the spatial goal changed. Specifically, it was 40 deg in retention and 60 deg in transfer whereas the timing goal remained 225 ms in both tests.

Development of error estimation skills

In Experiment 1, the subset of participants in each group who estimated their error after each pre-test trial improved the accuracy of their estimation skills throughout the experiment. We found significant main effects of Test for both the spatial, $F(1.45, 104.67) = 20.72$, $p < .001$, $\eta_G^2 = .144$, and timing, $F(1.10, 79.40) = 46.88$, $p < .001$, $\eta_G^2 = .060$, error estimations. For the spatial error estimations, Holm-Bonferonni post-hoc tests revealed that pre-test had higher error than retention and transfer (p's $< .001$), and transfer had higher error than retention ($p < .001$). Similarly, timing error estimations were more accurate in retention and transfer compared to the pre-test (p's $< .001$). In Experiment 2, none of the main effects or interactions were significant.
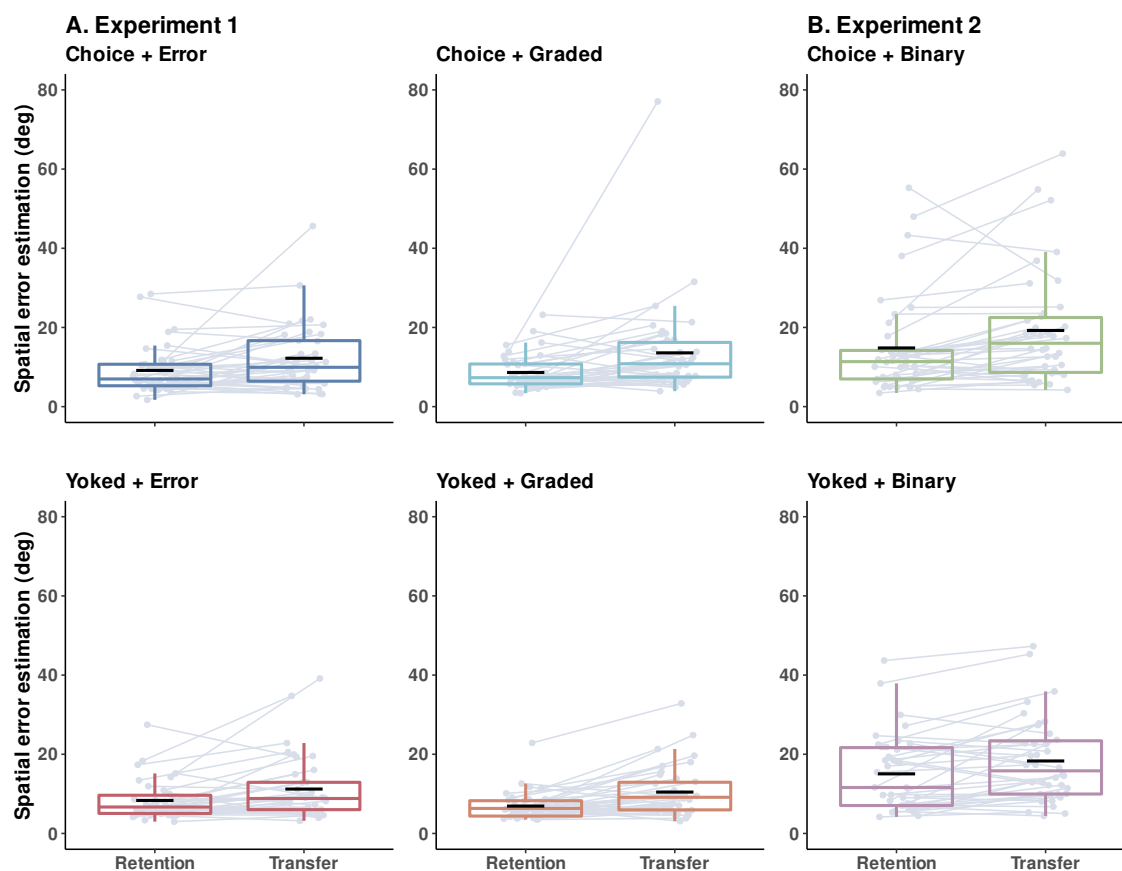
Figure S1

Total spatial error estimation data. *Boxplots of retention and transfer data from all participants for Experiment 1 (**A**) and Experiment 2 (**B**). The Choice with error feedback (Choice+Error) group is shown in dark blue, the Choice with graded feedback (Choice+Graded) group is shown in light blue, the Choice with binary feedback (Choice+Binary) group is shown in green, the Yoked with error feedback (Yoked+Error) group is shown in red, the Yoked with graded feedback (Yoked+Graded) group is shown in yellow, and the Yoked with binary feedback (Yoked+Binary) group is shown in purple. Boxplots represent 25th, 50th, and 75th percentiles, and the solid black line denotes the group mean. Grey connected dots represent individual data for participants in each group.*
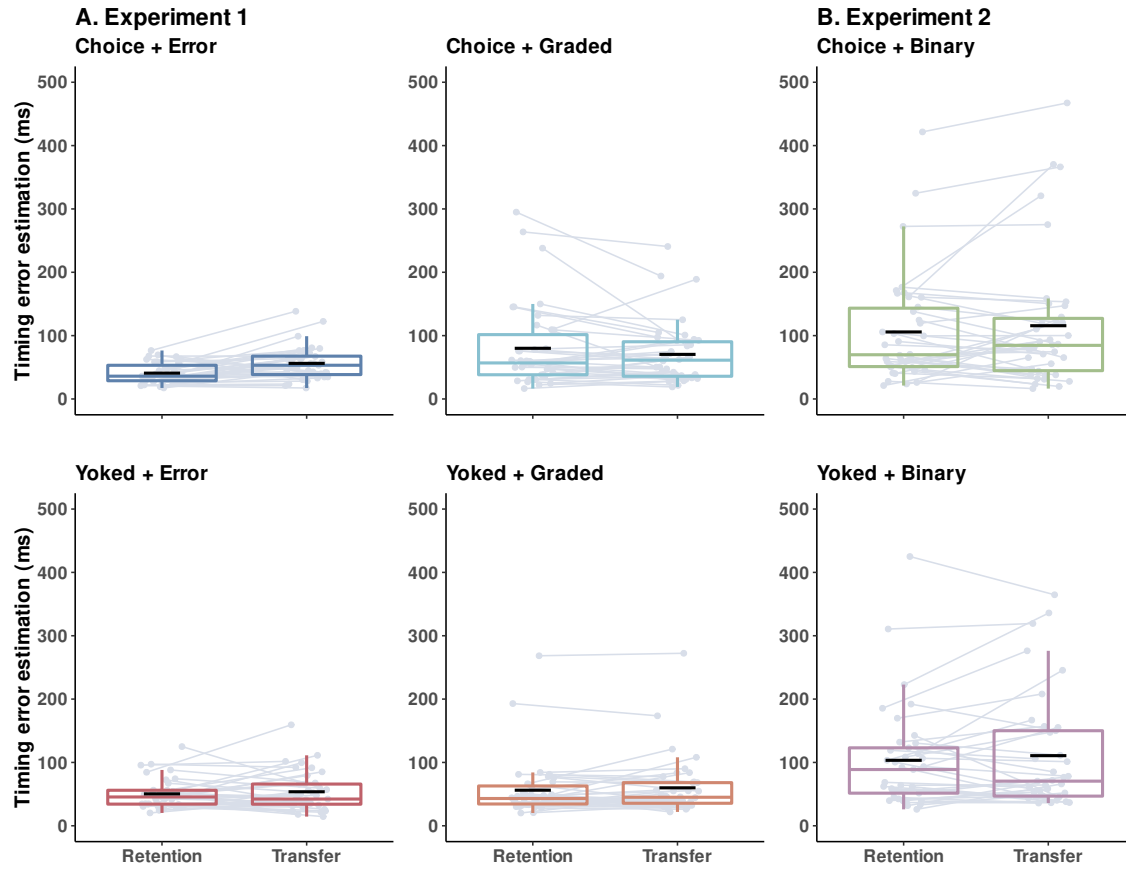
Figure S2

Total timing error estimation data. *Boxplots of retention and transfer data from all participants for Experiment 1 **(A)** and Experiment 2 **(B)**. The Choice with error feedback (Choice+Error) group is shown in dark blue, the Choice with graded feedback (Choice+Graded) group is shown in light blue, the Choice with binary feedback (Choice+Binary) group is shown in green, the Yoked with error feedback (Yoked+Error) group is shown in red, the Yoked with graded feedback (Yoked+Graded) group is shown in yellow, and the Yoked with binary feedback (Yoked+Binary) group is shown in purple. Boxplots represent 25th, 50th, and 75th percentiles, and the solid black line denotes the group mean. Grey connected dots represent individual data for participants in each group. Individual data of 1 participant in each of the Choice+Error, Yoked+Error, and Yoked+Binary groups is not shown as their error exceeded 500 ms.*

Supplementary B

Model diagnostics of total error (E) for the spatial and timing goals revealed skewed distributions. As a result, we carried out sensitivity analyses using shift functions, which is a robust technique that is well-suited for skewed distributions (Rousselet & Wilcox, 2020; Wilcox, 2021). For all shift functions, we collapsed across retention and transfer data to compute a single post-test score for both the spatial and timing goals. Given our primary interest was related to the role of choice during practice for motor learning and shift functions only compare two groups, we only ran these analyses on the choice factor for Experiment 1 (collapsed across feedback) and Experiment 2. Conducting a shift function analysis is a multi-step process that first involved calculating the 20% trimmed means for each participant and time point. Next, deciles for the factor of choice (choice versus yoked) in Experiments 1 and 2 were calculated using the Harrell-Davis estimator (Harrell & Davis, 1982; Rousselet et al., 2017). Lastly, differences at each decile were evaluated based on the 95% confidence intervals corrected for multiple comparisons using Hochberg's method (Hochberg, 1988). Spatial and timing measures were analyzed separately for Experiments 1 and 2.

Choice versus yoked

For Experiment 1, the 95% confidence intervals overlapped with zero at each decile when comparing the choice and yoked groups for both the spatial (Fig. S3A) and timing (Fig. S3B) goals. Similarly, in Experiment 2 there were no significant differences between the choice and yoked groups in any decile for the spatial (Fig. S3C) and timing (Fig. S3D) goals. The results from these shift function analyses are consistent with the analyses reported in the main manuscript and a recent meta-analysis (McKay et al., in-press), suggesting that giving learners choice during practice is not advantageous for motor learning.
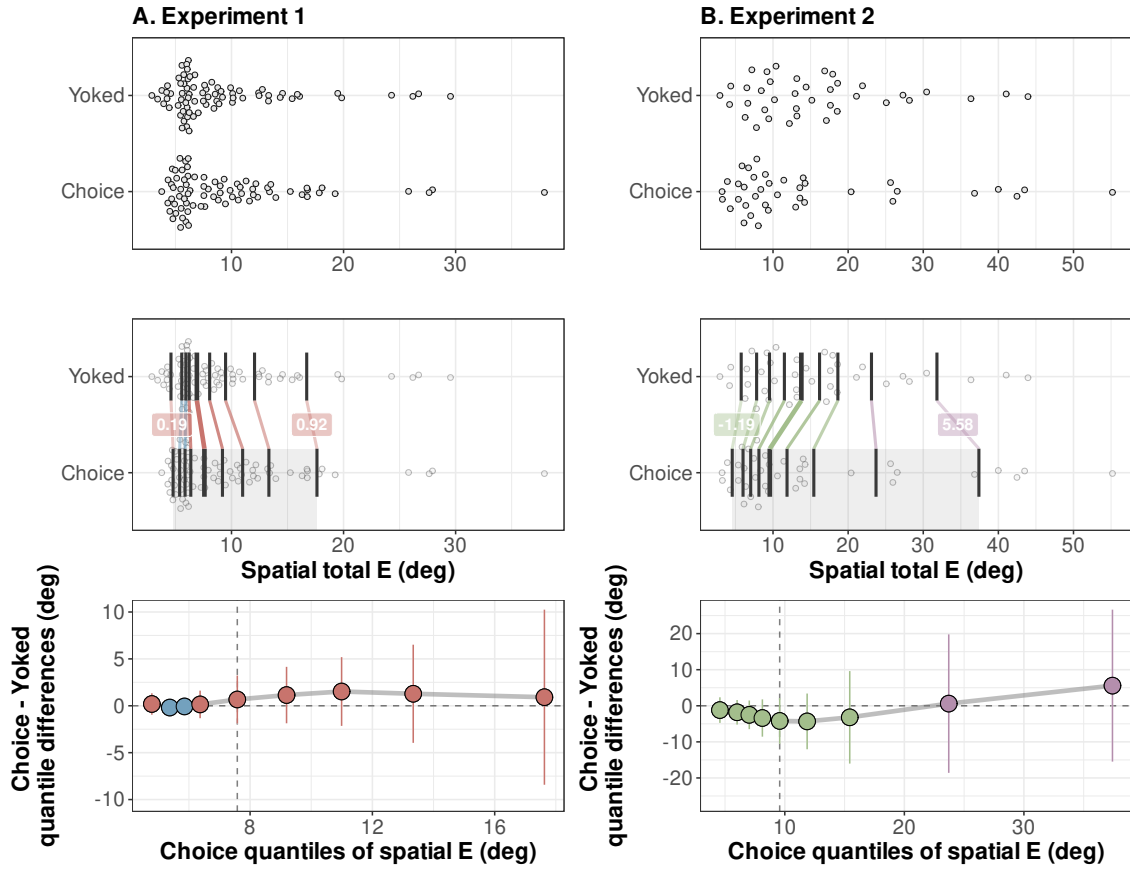
Figure S3

Post-test (collapsed across retention and transfer) spatial E (deg) shift function. *The top row illustrates scatter plots of individual mean spatial E for each group in Experiment 1 **(A)** and Experiment 2 **(B)**. The middle row illustrates the same scatter plots as the top row, but with the deciles of each distribution represented by the black lines. The thick black line represents the the median of each distribution. The deciles from each group are joined by colored lines, with blue (Experiment 1) and green (Experiment 2) indicating lower error for the choice group deciles, and red (Experiment 1) and green (Experiment 2) indicating lower error for the yoked group deciles. The bottom row illustrates the shift function, which focuses on the grey shaded region of the x-axis in the middle row. The deciles for the choice group are plotted on the x-axis and the difference in deciles between the choice and yoked group are plotted on the y-axis. The vertical dash line represents the median of the choice distribution. The same color coding for differences in deciles from the middle row is used in the bottom row. Error bars represent 95% confidence intervals, corrected for multiple comparisons. The horizontal dashed line represents zero differences between the deciles of the two group. If a 95% confidence interval overlaps with zero, there is no significant difference between the two groups on that decile of the distribution. There were no significant differences for any decile in either experiment.*
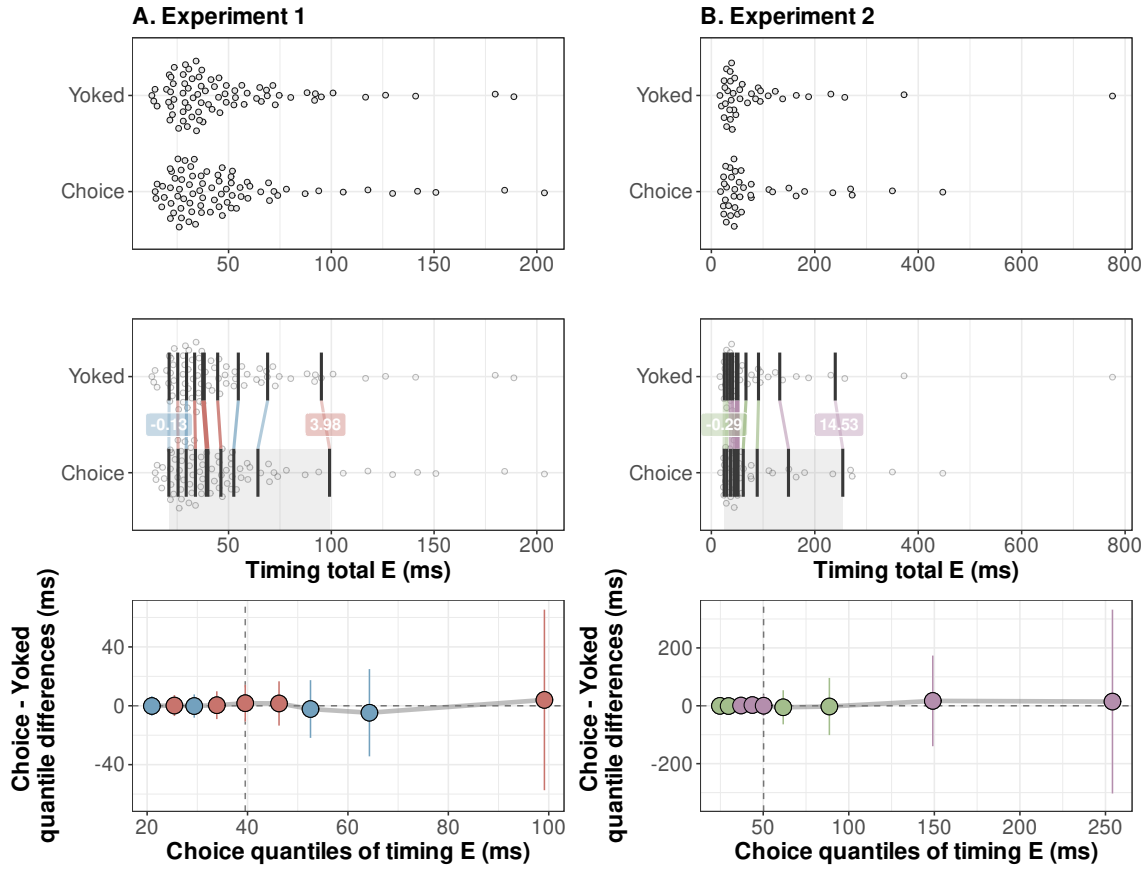
Figure S4

Post-test (collapsed across retention and transfer) timing E (ms) shift function. *The top row illustrates scatter plots of individual mean timing E for each group in Experiment 1 (A) and Experiment 2 (B). The middle row illustrates the same scatter plots as the top row, but with the deciles of each distribution represented by the black lines. The thick black line represents the the median of each distribution. The deciles from each group are joined by colored lines, with blue (Experiment 1) and green (Experiment 2) indicating lower error for the choice group deciles, and red (Experiment 1) and green (Experiment 2) indicating lower error for the yoked group deciles. The bottom row illustrates the shift function, which focuses on the grey shaded region of the x-axis in the middle row. The deciles for the choice group are plotted on the x-axis and the difference in deciles between the choice and yoked group are plotted on the y-axis. The vertical dash line represents the median of the choice distribution. The same color coding for differences in deciles from the middle row is used in the bottom row. Error bars represent 95% confidence intervals, corrected for multiple comparisons. The horizontal dashed line represents zero differences between the deciles of the two group. If a 95% confidence interval overlaps with zero, there is no significant difference between the two groups on that decile of the distribution. There were no significant differences for any decile in either experiment.*

References

Barros, J. A. C., Yantha, Z. D., Carter, M. J., Hussien, J., & Ste-Marie, D. M. (2019). Examining the impact of error estimation on the effects of self-controlled feedback. *Human Movement Science*, *63*, 182–198. https://doi.org/10.1016/j.humov.2018.12.002

Bruechert, L., Lai, Q., & Shea, C. H. (2003). Reduced knowledge of results frequency enhances error detection. *Research Quarterly for Exercise and Sport*, *74*(4), 467–472. https://doi.org/10.1080/02701367.2003.10609116

Carter, M. J., Carlsen, A. N., & Ste-Marie, D. M. (2014). Self-controlled feedback is effective if it is based on the learner's performance: A replication and extension of Chiviacowsky and Wulf (2005). *Frontiers in Psychology*, *5*, 1–10. https://doi.org/10.3389/fpsyg.2014.01325

Carter, M. J., & Patterson, J. T. (2012). Self-controlled knowledge of results: Age-related differences in motor learning, strategies, and error detection. *Human Movement Science*, *31*(6), 1459–1472. https://doi.org/10.1016/j.humov.2012.07.008

Harrell, F. E., & Davis, C. (1982). A new distribution-free quantile estimator. *Biometrika*, *69*(3), 635–640. https://doi.org/10.1093/biomet/69.3.635

Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–802. https://doi.org/10.1093/biomet/75.4.800

McKay, B., Yantha, Z. D., Hussien, J., Carter, M. J., & Ste-Marie, D. M. (in-press). Meta-analytic findings in the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value. *Meta-Psychology*. In-press. https://doi.org/10.31234/osf.io/8d3nb

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, *46*(2), 1738–1748.

Rousselet, G. A., & Wilcox, R. R. (2020). Reaction times and other skewed distributions: Problems with the mean and the median. *Meta-Psychology*, *4*, 1–39.

Wilcox, R. R. (2021). *Introduction to robust estimation and hypothesis testing* (5th ed.). Academic press.