Autonomy-supportive instructional language does not enhance skill acquisition compared to controlling instructional language

Laura St. Germain[1], Brad McKay[1], Lidia Barbera[2], Chitrini Tandon[2], Jeswende Seedu[3], Chantal Carrillo[4], Denver M.Y. Brown[5], and Michael J. Carter[1]

[1]Department of Kinesiology, McMaster University

[2]School of Interdisciplinary Science, McMaster University

[3]Faculty of Health Sciences, McMaster University

[4]Department of Psychology, Neuroscience, and Behaviour, McMaster University

[5]Department of Psychology, The University of Texas at San Antonio

Author Note

Laura St. Germain https://orcid.org/0000-0002-5513-4183

Brad McKay https://orcid.org/0000-0002-7408-2323

Devnver M.Y. Brown https://orcid.org/0000-0003-4078-8253

Michael J. Carter https://orcid.org/0000-0002-0675-4271

Correspondence concerning this article should be addressed to Michael J. Carter, Department of Kinesiology, McMaster University. 1280 Main St. W., Hamilton, ON, Canada. L8S 4K1. Email: cartem11@mcmaster.ca; motorlab@mcmaster.ca

Abstract

Instructional language is one of three techniques in OPTIMAL theory that can be manipulated to foster an autonomy-supportive practice environment to enhance motor performance and learning. While autonomy-supportive language has been shown to be beneficial in educational psychology, coaching, and health settings, the wording of task instructions has received minimal attention in the motor learning literature to date. We investigated the influence of two instructional language styles on skill acquisition in a preregistered experiment. Participants ($N = 156$) learned a speed cup stacking task and received instructions throughout practice that used either autonomy-supportive or controlling language. Although the autonomy-supportive instructions resulted in higher perceptions of autonomy, there were no group differences for motor performance in acquisition or retention. Perceptions of competence and intrinsic motivation did not differ between groups at any time point. These data are difficult to reconcile with key predictions in OPTIMAL theory regarding a direct and causal influence of motivational factors on performance and learning. However, our equivalence test suggests these effects on skill acquisition may be smaller than what we were powered to detect. These findings are consistent with a growing body of evidence highlighting the need for much larger N experiments in motor learning research.

Keywords: Motor learning, Retention, OPTIMAL theory, Preregistered

Autonomy support refers to a teaching style or approach that fosters self-determination and intrinsic motivation in learners by providing them with choices, respect, and opportunities to make decisions. In Self-Determination Theory (Deci & Ryan, 2012; Ryan & Deci, 2020), autonomy is broadly defined as the sense of ownership and initiative over one's behaviors. Within the Basic Psychological Needs Theory (Ryan & Deci, 2000, 2017) of Self-Determination Theory, humans have inherent psychological needs for autonomy, competence, and relatedness. When these needs are satisfied, individuals experience positive outcomes such as enhanced performance and increased intrinsic motivation; accompanied by a greater sense of interest, enjoyment, and inherent satisfaction (Ryan & Deci, 2000). Autonomy support has been shown to be efficacious in a variety of contexts, including educational psychology (see Reeve, 2009; Ryan & Deci, 2020 for reviews; see Su & Reeve, 2011 for a meta-analysis), coaching (see Mossman et al., 2022 for a meta-analysis), and health (see Okada, 2021 for a meta-analysis). During the last decade, motor learning scientists have become increasingly interested in the use of autonomy-supportive practice conditions for skill acquisition (see Sanli et al., 2013; Ste-Marie et al., 2020; Wulf & Lewthwaite, 2016 for reviews).

Within the OPTIMAL theory of motor learning, Wulf and Lewthwaite (2016) proposed that autonomy-supportive practice conditions benefit motor performance and learning through enhanced expectancies, efficient goal-action coupling, and dopamine availability for memory consolidation and neural pathway development (p. 1404). The main autonomy-supportive manipulation used in the motor learning literature to date has been providing learners with opportunities for choice either before or during practice. In OPTIMAL theory, Wulf and Lewthwaite (2016) highlighted two ways to support autonomy through choice: *control over practice conditions* (i.e., task-relevant choices) and *incidental choices* (i.e., task-irrelevant choices). The dominant view over the years has been that both choice manipulations are effective for skill acquisition (e.g., Carter et al., 2014; Carter & Ste-Marie, 2017; Chiviacowsky & Wulf, 2002, 2005; Lewthwaite et al., 2015; Wulf et al., 2014; Wulf et al., 2018); commonly referred to as the self-controlled learning advantage. Recently, however, this so-called self-controlled learning advantage has failed to be replicated in several large N—and often pre-registered—experiments (Bacelar et al., 2022; Leiker et al., 2019; McKay & Ste-Marie, 2020, 2022; St. Germain et al., 2023; St. Germain et al., 2022; Yantha et al., 2022). A recent meta-analysis found that estimates of the self-controlled learning effect could range from $g = -0.11$ to 0.26 after correcting for publication bias (McKay et al., 2022). McKay, Bacelar, et al. (2023) re-analyzed this meta-analysis using a robust Bayesian approach (Bartoš et al., 2023; Maier, 2023) and found the overall model ensemble estimated the effect as $d = .034$ (95% credible interval [.0, .248]). Taken together, these studies suggest that the true effects of these choice manipulations for motor learning are uncertain, small, and potentially null. Wulf and Lewthwaite (2016) also highlighted instructional language as a third practice variable that can be manipulated to enhance learning through autonomy-support. Yet, the wording of such task instructions has received minimal attention in the motor learning literature to date.

The language used in task instructions exists on a continuum ranging from highly controlling to autonomy-supportive (Reeve, 2009). Factors that contribute to autonomy-supportive instructions are prioritization of the learner's perspective and goals, openness to learner initiative, and support for learner self-direction (Reeve, 2009; Reeve & Tseng, 2011). Reeve and Tseng (2011) provided participants with either controlling, autonomy-supportive, or neutral instructions about how to solve near-unsolvable puzzles. Despite performance on the puzzles being the same between groups (i.e., the puzzles were not solved), the controlling language group had the lowest perceptions of autonomy while the autonomy-supportive group had the highest perceptions of competence. Thus, autonomy-supportive instructions can exert affective benefits even in the absence of performance gains. Hooyman et al. (2014) extended this work to the motor learning literature by providing participants with either autonomy-supportive, controlling, or neutral instructions about how to perform a modified cricket bowl to a target. Compared to the controlling language group, the autonomy-supportive group performed with less error in practice and in a delayed retention test, and also had higher ratings for perceived choice, self-efficacy, and positive affect at the end of practice.

Although the results of Hooyman et al. (2014) suggested a motor performance and learning benefit of autonomy-supportive language, there are some methodological limitations that warrant consideration. First, the autonomy-supportive instructions were confounded with an analogy; previously shown to also facilitate motor performance and learning (e.g., Liao & Masters, 2001; see Masters et al., 2020 for a review). As the analogy was not part of the controlling or neutral language instructions, it is impossible to disentangle whether the benefits in the autonomy-supportive group resulted from the instructional language, the analogy, or some combination of the two. Second, the authors' measure of perceived choice to capture autonomy-support does not comprehensively map onto the basic needs of Self-Determination Theory (McDonough & Crocker, 2007; Ng et al., 2011; Ryan & Deci, 2020) and has been shown to be a poor indicator of self-determination and intrinsic motivation (Reeve et al., 2003). Lastly, the experimental design was underpowered for all but unplausibly large effect sizes for motor learning research. With 16 participants per group, the main effect of Group at retention would only be able to detect f of 0.4 (equivalent to d of .8 for a t-test) with 80% power. Such an effect is considerably higher than an estimate of the median effect size in motor learning studies ($d = 0.63$, Lohse et al., 2016). Further, when underpowered designs find significant results, they are prone to be false positives with inflated effect size estimates (Button et al., 2013; Simmons et al., 2011).

Here, we investigated the effects of autonomy-supportive language on motor performance and learning while addressing the methodological limitations of Hooyman et al. (2014). Participants practiced a speed cup stacking task and received instructions with either autonomy-supportive or controlling language. Participants also self-reported their perceptions of autonomy, competence, and intrinsic motivation at multiple time points in the experiment. Based on OPTIMAL theory, we predicted that participants in the autonomy-supportive language group would demonstrate faster stacking times in practice and retention, and report higher perceptions of autonomy, competence, and intrinsic motivation compared to the participants in the controlling language group.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons et al., 2012). Data and code are available at https://github.com/cartermaclab/expt_instructional-language and the pre-registration can be accessed at https://doi.org/10.17605/osf.io/9n46p.

Sample size calculation

To test our primary prediction that autonomy-supportive instructional language would enhance motor skill retention compared to controlling instructional language, we performed a two-stage *a priori* power analysis using the smallest effect size of interest approach (see Lakens, 2022 for a discussion). We specified our smallest effect size of interest as $d = 0.4$. This is a conservative estimate compared to an estimate of the median effect size in the motor learning literature ($d = 0.63$ in Lohse et al., 2016), a meta-analytic estimate of the effect size of autonomy-supportive instructional language ($d = 0.63$, Su & Reeve, 2011), and has been suggested as a reasonable smallest effect size of interest for psychological research (Brysbaert, 2019).

In the first stage, we used a one-sided Welch's t-test with the following parameters: $\alpha = 0.05$, $\beta = 0.20$, and $d = 0.4$, resulting in 78 participants per group for a total of 156 participants. In the second stage, we used a shift function, which is a family of robust statistical techniques for comparing entire distributions of data (Rousselet et al., 2017; Rousselet & Wilcox, 2020; Wilcox, 2021; Wilcox & Rousselet, 2023). It is therefore a useful alternative to comparisons based on means as effects can, and do, occur in the tails of distributions. In other words, the shift function is a powerful tool to determine how, and by how much, two distributions differ (Rousselet et al., 2017; Wilcox, 2021). For this power analysis, we simulated right-skewed distributions with $n = 78$ per group and a mean difference of 0.4. Right-skewed distributions were used because time based quantities are typically asymmetric (see Rousselet & Wilcox, 2020 for a discussion) and our primary outcome variable was stacking time. We performed 10,000 simulated experiments using both a one-sided Welch's t-test and a shift function to determine which statistical analysis should be used to test our primary prediction. The t-test had 80% power (consistent with that from the first stage) whereas the shift function had 88% power. As such, the shift function was chosen as our primary analysis.[1]

Participants

A convenience sample of undergraduate and graduate students at a Canadian university in southwestern Ontario participated in the experiment. Participants were randomly assigned to either the autonomy-supportive instructional language group ($M_{age} = 18.7$ years, $SD = 1.85$, $n = 78$, 54 females) or the controlling language group ($M_{age} = 18.9$ years, $SD = 2.39$, $n = 78$, 52 females). Participants were compensated $15 CAD or with course-credit for their time. All participants gave written informed consent and the experiment was approved by the University's Research Ethics Board.

Task and apparatus

Participants learned the 3-6-3 speed cup stacking sequence in accordance with the rules of the World Sport Stacking Association (https://www.thewssa.com). Official Speed Stack cups (https://www.speedstacks.com) were used and participants performed the task using both of their hands. To successfully complete the 3-6-3 sequence, participants performed an upstack phase and a downstack phase. The cups began in upside down piles consisting of three, six, and three cups from left to right. The upstack phase required participants to create a 3-cup pyramid, followed by a 6-cup pyramid, then another 3-cup pyramid. The down stack phase consisted of collapsing the first 3-cup pyramid from the

[1] We report the t-test as a secondary analysis for the interested reader.

upstack phase, then the 6-cup pyramid, and then the remaining 3-cup pyramid so the cups were in the same configuration as the start of the task. The goal of the task was to perform the upstack and downstack phases as fast as possible.

Procedure

Data collection involved two sessions that occurred on consecutive days. Session 1 consisted of obtaining informed consent, a demographics questionnaire, the pre-test (5 no-feedback trials), an acquisition phase (30 trials with feedback), and questionnaires related to three psychological constructs. Session 2 consisted of the same three questionnaires and the delayed (∼24 hours) retention test (5 no-feedback trials). Participants completed both sessions of the experiment individually. Participants received phase-specific instructions using neutral language at the start of each experimental phase. Group-specific instructions were also provided at the start of acquisition (see below for details). Instructions were displayed on a 22-inch computer monitor (1920x1080 resolution) positioned to the left of the participant.

At the start of each trial, participants stood at a standard height table with their hands on marked locations and the 12 upside down cups arranged in the 3-6-3 configuration on the table in front of them. Participants were shown a "Get Ready!" prompt on the computer monitor for 1 s. After a 1 s constant foreperiod, an audiovisual go-signal (green square and a beep tone) was presented. Participants were instructed to begin the upstack phase as quickly as possible following the go-signal as its presentation initiated the timer. Once the upstack and downstack phases were completed, participants hit the spacebar on a keyboard located in front of them to stop the timer. If an error occurred (e.g., forgot to hit the spacebar to stop timer, only completed the upstack phase then stopped the timer, etc.), the researcher flagged the trial number for later removal.

Prior to the pre-test, participants received neutral language instructions that described the cup stacking task and the pre-test protocol. Included in these instructions were two videos from the Speed Stacks website. The first was a demonstration of how to

Table 1

Cronbach's alpha for each questionnaire at each timepoint.

Questionnaire	After pre-test	After acquisition	Before retention
Perceived autonomy	0.79	0.83	0.85
Perceived competence	0.79	0.86	0.88
Intrinsic motivation	0.88	0.91	0.92

perform the 3-6-3 sequence and the second described what to do if any cups were knocked over during the upstack and/or downstack phases. The pre-test consisted of five trials with no feedback regarding their stacking time. After the pre-test, participants completed three questionnaires related to key psychological variables in OPTIMAL theory (Wulf & Lewthwaite, 2016). Perceived autonomy and perceived competence were assessed using the Basic Needs Satisfaction in Sport Scale, which has been shown to have good reliability and construct validity (Ng et al., 2011). The perceived competence subscale has 5 items, for example "I feel I am good at this task". The perceived autonomy subscale has 10 items to capture choice (4 items, e.g., "In this study, I get opportunities to make choices"), an internal perceived locus of causality (3 items, e.g., "In this study, I feel I am pursuing goals that are my own"), and volition (3 items, e.g., "I choose to participate in this study according to my own free will"). Intrinsic motivation was assessed using the Task Interest and Enjoyment subscale (7 items, e.g., "This cup stacking task was fun to do") of the Intrinsic Motivation Inventory (McAuley et al., 1989). For all questions, participants read a statement on a handout and then verbally reported their answer using a Likert scale ranging from 1 (not true at all) to 7 (very true). Answers were recorded by a trained research assistant and stored for later analysis. Cronbach's alpha values for each questionnaire at each time point are reported in Table 1.

Before the acquisition phase, participants received neutral language instructions that described the cup stacking task and the acquisition protocol. The acquisition phase consisted of 30 trials and feedback about stacking time was displayed for 2 s after every

Table 2

Group specific instructions received during acquisition before trials 1, 11, and 21.

Group	Acquisition instructions
Autonomy-support	Are you ready to learn how to cup stack? Does this sound like an activity you might want to try? It will probably be helpful if you think of the task as a challenge and consider a goal to complete it as quickly as possible. To help, I'll offer some hints here at the beginning. You have probably already noticed that I've put the cups in three stacks. It might be helpful to arrange them in order from left to right, with three cups on the left, six in the middle, and three on the right. You might be thinking that the best way to complete the task is to upstack the cups from left to right, then return to the beginning and also downstack from left to right. If this is what you're thinking, you are right! I understand that you might feel a little hesitant and unsure. Most people feel this way, at least at first. You are free to begin when you wish.
Controlling	Your job is to learn cup stacking – perform it well and do it as quickly as possible. To do so, do what I tell you to do. Don't begin yet, listen carefully to me. Make sure the stacks are in their proper order. I want the stacks in order from left to right, with three cups on the left, six in the middle, and three on the right. Make sure the stacks are in their proper order. If so, good. If not, fix it. When completing the task, I want you to upstack the cups from left to right, then return to the beginning to also downstack from left to right. If you're thinking of doing it differently - don't, that is not what I told you to do. Begin.

trial. Prior to acquisition trials 1, 11, and 21, participants received group-specific instructions based on whether they were randomly assigned to the autonomy-supportive language group or the controlling language group (see Table 2). The group-specific instructions were pre-recorded by a human and played as an audio clip to participants. This was done to ensure that instruction delivery was consistent across participants, and to eliminate potential confounds such as differences in tone, pace, amount of eye contact, etc. Our group-specific instructions were reviewed and revised based on feedback from an expert in autonomy-supportive instructional language.[2] After all acquisition trials were finished, participants completed the three questionnaires a second time.

[2] The reviewed instructions (Dr. Johnmarshall Reeve, personal communication, April 20 2022) referred to a floor curling task rather than the cup stacking task that was ultimately used in the experiment.

Participants returned to the lab for Session 2 approximately 24 hours after finishing Session 1. Upon arrival, participants completed the three questionnaires for a third and final time. Before performing the delayed retention test, participants received neutral language instructions that described the cup stacking task and the retention protocol. The retention test consisted of five trials with no feedback regarding their stacking time.

A custom LabVIEW (National Instruments Inc.) program was created that controlled the presentation of instructions, the timing of experimental protocol, and recorded and stored the data for offline analysis.

Data analysis

Our primary outcome variable was stacking time, which was the interval between the go-signal and the participant hitting the spacebar. Trials recorded as an error (122/6240, 1.96%) during data collection were removed before data analysis. Stacking time was calculated as the mean for blocks of five trials, resulting in one pre-test block, six acquisition blocks, and one delayed retention block. Perceived autonomy and perceived competence scores were respectively calculated as the mean of the 10 autonomy items and six competence items from the Basic Needs Satisfaction in Sport Scale. Intrinsic motivation was calculated as the mean of the seven items of the Task Interest and Enjoyment subscale from the Intrinsic Motivation Inventory.

Alpha was set to .05 for all statistical tests, which are described below. Corrected degrees of freedom using the Greenhouse-Geisser method are always reported when appropriate. Generalized eta squared (η^2_G) is reported as an effect size for all omnibus tests. Post hoc comparisons were adjusted for multiple comparisons using the Holm-Bonferroni correction. Statistical analyses were conducted using `R` (Version 4.3.2; R Core Team, 2021) and the `R`-packages *afex* (Version 1.3.0; Singmann et al., 2023), *computees* (Version 0.2.5; Re, 2013), *cronbach* (Version 0.1; Tsagris & Frangos, 2020), *effsize* (Version 0.8.1; Torchiano, 2020), *emmeans* (Version 1.9.0; Lenth, 2023), *ggpmisc* (Version 0.5.5; Aphalo, 2022), *hmisc* (Version 5.1.0; Harrell Jr, 2023), *kableextra* (Version 1.3.4; Zhu, 2021), *patchwork* (Version 1.2.0; Pedersen, 2022), *pwr* (Version 1.3.0; Champely, 2020), *renv* (Version 1.0.3; Ushey, 2023), *rogme* (Version 0.2.1; Rousselet et al., 2017), *rstatix* (Version 0.7.2; Kassambara, 2023), and *tidyverse* (Version 2.0.0; Wickham et al., 2019) were used in this project.

Results

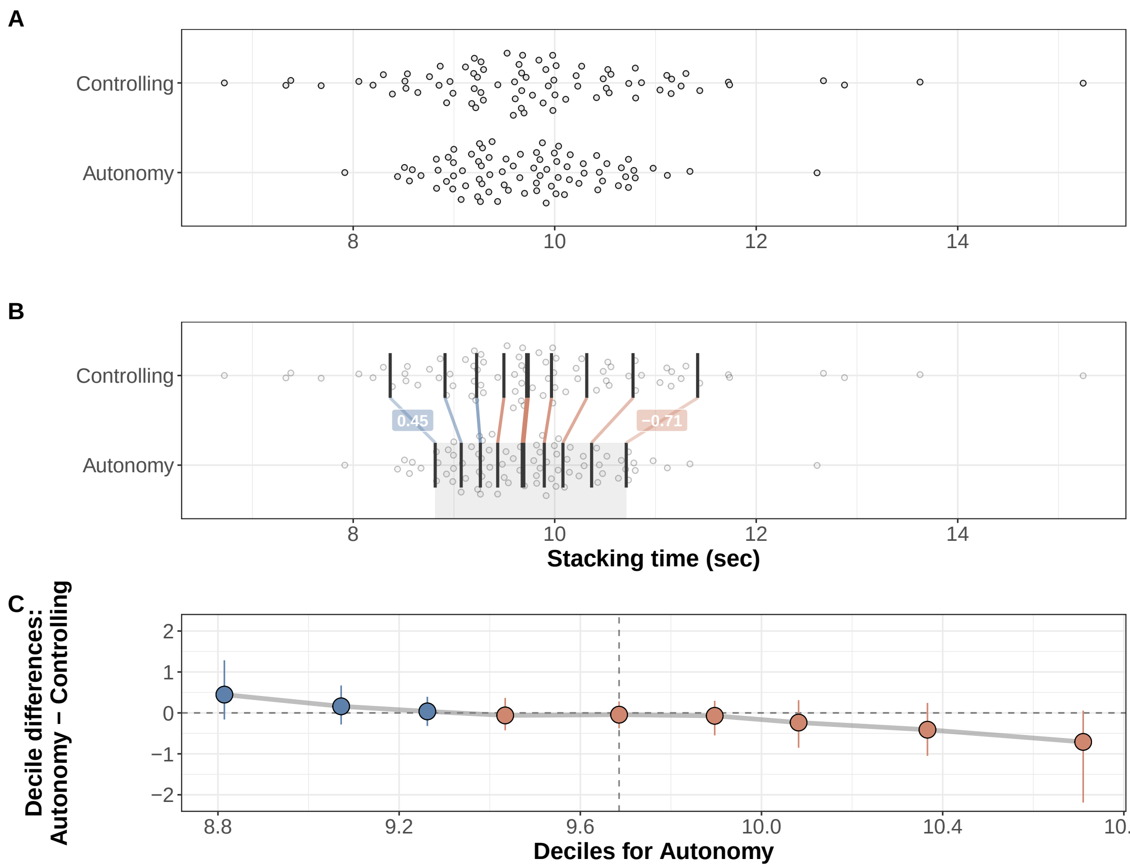
Primary analysis

We performed a shift function on mean stacking time in retention, adjusted for pre-test scores, to test our primary prediction that autonomy-supportive instructional language would enhance learning compared to controlling instructional language. The shift function is a multi-step analysis that first involves calculating the 20% trimmed means for pre-test and retention for each participant. We then regressed retention stacking time onto pre-test stacking time (i.e, using pre-test as a covariate). Next, we computed deciles using the Harrell-Davis estimator (Harrell & Davis, 1982) and 95% confidence intervals around each decile were calculated using percentile bootstraps (Efron & Tibshirani, 1994; Rousselet et al., 2021, 2023). Corrected p-values using Hochberg's method (Hochberg, 1988) were calculated for each decile. The shift function is considered significant if any of the corrected p-values were $\leq .05$.

Pre-test adjusted stacking time data for each participant in the controlling language and autonomy-supportive language groups are shown in Figure 1A. The shift function (Figure 1C) comparing the groups at each decile (Figure 1B) is relatively flat in the middle (deciles 3 through 6), but has a negative slope, indicating that the two distributions differ in their spread (Rousselet et al., 2017). The largest differences between the groups were in the first and last deciles. The controlling language group was faster than the autonomy-supportive group in the first decile whereas the autonomy-supportive language group was faster in the last decile. None of the decile comparisons were significant after p-values were adjusted for multiple comparisons.

Figure 1

Shift function on retention stacking times adjusted for pre-test times.



Note. Scatterplot of stacking time as a function of experimental group **(A)** with each data point representing a 20% trimmed mean for an individual participant. The same scatterplot from the top row with the deciles if each distribution represented by the black lines **(B)**. The thick black line represents the median of each distribution. The difference between groups at each decile are represented by the colored lines. A blue line indicates that the Controlling language group was faster in a decile and an orange line indicates that the Autonomy-supportive language group was faster in a decile. The bottom row illustrates the shift function **(C)**, which focuses on the grey shaded region of the x-axis in the middle row. The deciles for the Autonomy-supportive language group are plotted on the x-axis and the difference in deciles between the two groups are plotted on the y-axis. The vertical dashed line represents the median of the Autonomy-supportive language group. The circles represent the decile differences using the same color coding described above. Error bars represent 95% percentile bootstrapped confidence intervals. All decile comparisons were not significant after p-values were adjusted for multiple comparisons using Hochberg's method. For interpretation of the references to color in this figure, the reader is referred to the online version of this article.

Secondary analyses

Acquisition phase

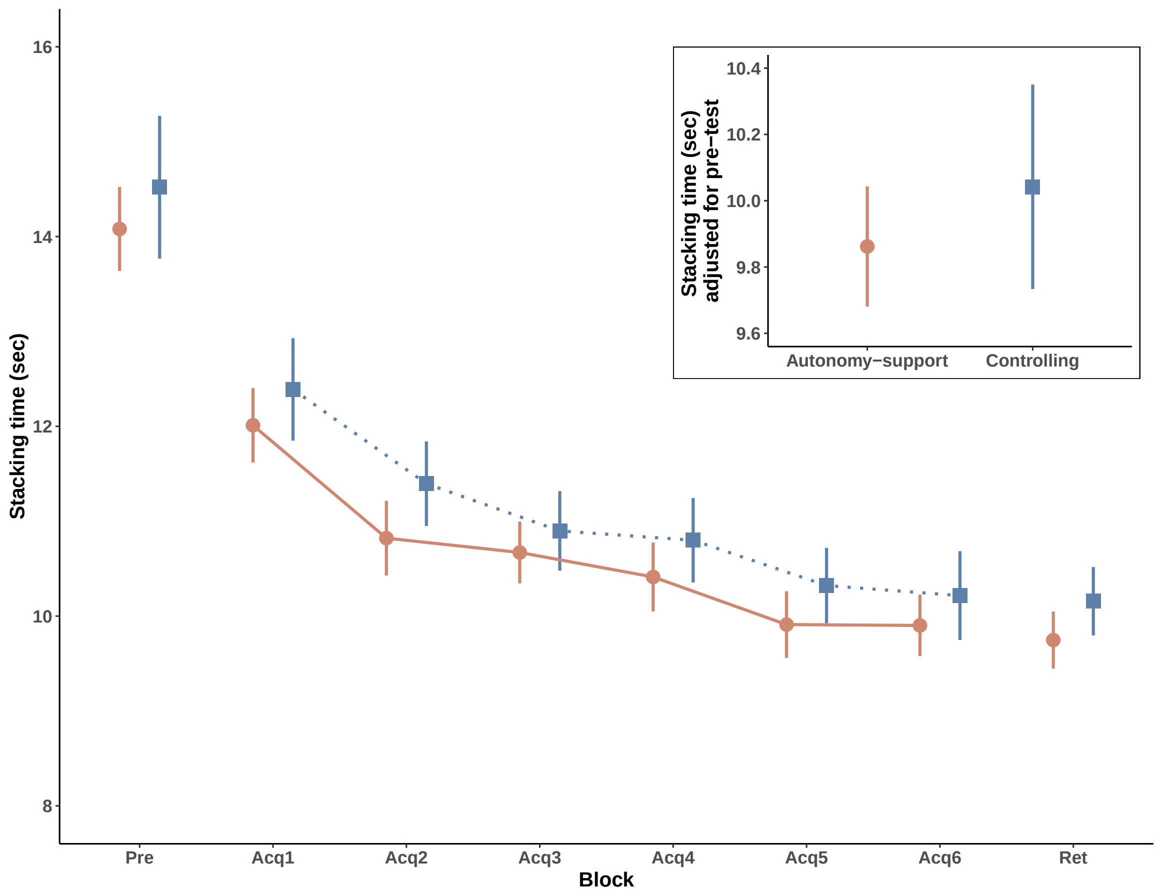
We analyzed mean stacking time during the acquisition period using a 2 Group (Autonomy-support, Controlling) x 6 Block mixed design ANOVA with repeated measures on Block. Participants in the autonomy-supportive language and controlling language groups decreased their stacking time across acquisition blocks (Figure 2). This was supported by a significant main effect of Block, $F(4.40, 677.96) = 111.91$, $p < .001$, $\eta^2_G = .137$. Stacking time in Block 1 was slower than all other blocks (p's $< .001$), Block 2 was slower than Blocks 3 to 6 (p's $\leq .004$), and Blocks 3 and 4 were both slower than Blocks 5 and 6 (p's $< .001$). The main effect of Group, $F(1, 154) = 2.20$, $p = .140$, $\eta^2_G = .011$, and the Group x Block interaction, $F(4.40, 677.96) = 0.59$, $p = .681$, $\eta^2_G < .001$, were not significant.

Retention test

Performance in the delayed retention test was also analyzed using a more familiar approach in motor learning research. A one-tailed Welch's t-test on pre-test adjusted mean stacking times for the Autonomy-supportive language group ($M = 9.86$ s, $SD = 0.80$) and the Controlling language group ($M = 10.04$ s, $SD = 1.37$) was not significant, $t(124.51) = 1.00$, $p = .159$, $d = .16$ $[-.156, .477]$. This finding is consistent with those of our primary analysis using the shift function.

Figure 2

Motor performance data for all experimental phases.



Note. Mean stacking time (s) for the Autonomy-supportive language (orange circles, solid line) and Controlling language (blue squares, dotted line) groups were computed by averaging the data into blocks of five trials. This resulted in one block for pre-test (Pre), six block for acquisition (Acq), and one block for the 24-hr delayed retention test (Ret). The pre-test and acquisition blocks were completed in Session 1 and the retention block was completed in Session 2. Feedback about stacking time (s) was only available during the acquisition blocks and was provided after each trial. Group-specific instructions as a function of experimental group were played as pre-recorded audio clips before trials 1 (start of Block 1), 11 (start of Block 3), and 21 (start of Block 5) in acquisition. The inset figure shows pre-test adjusted retention stacking time (s) for both groups. Error bars in both figures represent 95% confidence intervals. For interpretation of the references to color in this figure, the reader is referred to the online version of this article.

Psychological variables

We assessed the impact of our instructional language manipulation on perceived autonomy, perceived competence, and intrinsic motivation using separate 2 Group (Autonomy-support, Controlling) x 2 Time (After acquisition, Before retention) mixed ANCOVAs with pre-test scores as the covariate and repeated measures on Time.

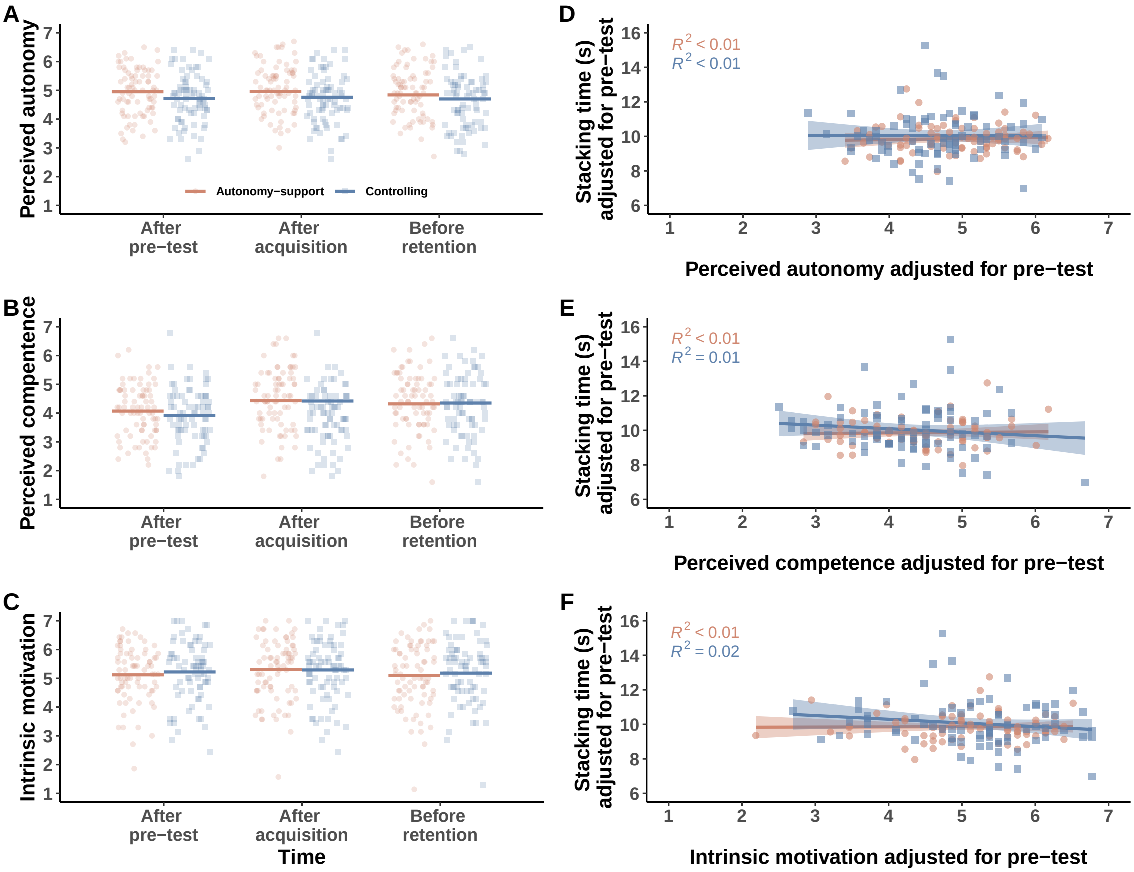
Perceptions of autonomy (adjusted for pre-test) remained consistent within both groups (Figure 3A). Pre-test, the covariate, was a significant predictor of later time points, $F(1, 153) = 247.68$, $p < .001$. Participants in the autonomy-supportive language group self-reported higher scores than the participants in the controlling language group after acquisition and before retention. This was supported by a significant main effect of Group, $F(1, 153) = 3.90$, $p = .05$, $\eta^2_G = .022$. The main effect of Time, $F(1, 153) = 0.01$, $p = .924$, $\eta^2_G < .001$, and the Group x Time interaction, $F(1, 153) = 1.13$, $p = .290$, $\eta^2_G < .001$, were not significant.

Perceptions of competence (adjusted for pre-test) were relatively consistent in both groups (Figure 3B). Pre-test was a significant predictor of later time points, $F(1, 153) = 231.31$, $p < .001$. Groups did not differ in their self-reported perceptions of competence. The main effects for Group, $F(1, 153) = 0.01$, $p = .933$, $\eta^2_G < .001$, and Time, $F(1, 153) = 0.56$, $p = .456$, $\eta^2_G < .001$, as well as the Group x Time interaction, $F(1, 153) = 0.38$, $p = .539$, $\eta^2_G < .001$, were not significant.

Intrinsic motivation (adjusted for pre-test) scores remained consistent within both groups (Figure 3C). Pre-test was a significant predictor of later time points, $F(1, 153) = 387.06$, $p < .001$. Groups did not differ in their self-reported intrinsic motivation. The main effects for Group, $F(1, 153) = 0.11$, $p = .743$, $\eta^2_G < .001$, and Time, $F(1, 153) = 0.21$, $p = .605$, $\eta^2_G < .001$, as well as the Group x Time interaction, $F(1, 153) = 0.13$, $p = .177$, $\eta^2_G = .002$, were not significant.

Figure 3

Questionnaire data.



Note. Self-reported scores for perceived autonomy **(A)**, perceived competence **(B)**, and intrinsic motivation **(C)** after the pre-test, after the acquisition phase, and before the delayed retention test for the Autonomy-supportive language (orange circles) and the Controlling language (blue squares) groups. The horizontal bars represent the group means, with the pre-test adjusted mean shown for after acquisition and before retention. Each data point represents the mean score across subscale items for an individual participant. The relationship between retention stacking time (s) adjusted for pre-test and perceived autonomy **(D)**, perceived competence **(E)**, and intrinsic motivation **(F)** before retention and adjusted for pre-test is shown. Each data point represents the mean score across subscale items for an individual participant in the Autonomy-supportive language (orange circles) and the Controlling language (blue squares) groups. The estimated regression fit (solid lines) for each group is shown. The shaded areas represent the 95% confidence intervals. A negative slope in these plots would suggest faster stacking times were associated with higher self-reported scores on the psychological variable of interest. For interpretation of the references to color in this figure, the reader is referred to the online version of this article.

We also performed some exploratory correlational analyses between our three psychological variables and performance in retention. We plotted pre-test adjusted retention stacking times as a function of perceived autonomy (Figure 3D), perceived competence (Figure 3E), and intrinsic motivation (Figure 3F) scores before retention, adjusted for pre-test for each participant. If there were associations between these psychological variables and performance in retention, we expected to see a negative relationship (i.e., faster stacking times associated with higher self-reported scores). As can be seen, we instead found the relationships between retention performance and each psychological variable to be relatively flat.

Equivalence test

Due to the null findings of the shift function (and t-test) on retention stacking times, we tested for equivalence with a noninferiority test as outlined in our preregistration. We used the two one-sided test procedure (Schuirmann, 1987) and a noninferiority bound of $d = .4$, which was our smallest effect size of interest. The test was not significant, $t(124.5) = 1.50$, $p = .069$. The 90% confidence interval around the effect size in retention was [-.11, .43], indicating that these data are inconsistent with all effects larger than $d = \pm.43$.

Discussion

In their OPTIMAL theory of motor learning, Wulf and Lewthwaite (2016) suggested that motor performance and learning can be enhanced when learners receive task instructions that use autonomy-supportive rather than controlling language. Here, we investigated the effect of autonomy-supportive instructional language on the acquisition and retention of a speed cup stacking task. Based on the OPTIMAL theory, we predicted that participants in the autonomy-supportive language group would demonstrate faster stacking times in acquisition and delayed retention, and would also report higher perceptions of autonomy, competence, and intrinsic motivation compared to those in the controlling language group. Our results did not show a performance benefit from autonomy-supportive language in acquisition or retention compared to controlling language. We found significantly higher perceptions of autonomy in the autonomy-supportive language group compared to the controlling language group, but no significant group differences for perceived competence or intrinsic motivation. Taken together, our findings do not support key predictions of the OPTIMAL theory of motor learning.

We failed to replicate the performance advantage of autonomy-supportive language in acquisition and delayed retention compared to controlling language that was reported by Hooyman et al. (2014). This is also inconsistent with Wulf and Lewthwaite's (2016) OPTIMAL theory wherein task instructions that utilize autonomy-supportive language results in a virtuous cycle that has positive influences on motor performance and learning. Importantly, our failed replication and lack of support for OPTIMAL theory are not the result of participants failing to improve at the motor task or an unsuccessful instructional language manipulation. That is, both the autonomy-supportive language and controlling language groups showed a decrease in stacking times from pre-test to the delayed retention test, suggesting learning occurred (see Figure 2) and the autonomy-supportive language group reported higher perceptions of autonomy (see Figure 3A). These conflicting findings may be due to the previously identified methodological limitation in Hooyman et al. (2014) of a small sample size or potential flexibility in the data analysis as their experiment was not pre-registered. Although such factors may have contributed, we believe the main reason for our discrepant results arise from the confounding analogy included in Hooyman and colleagues' (2014) autonomy-supportive instructions, but excluded from both their controlling and neutral language instructions. It is therefore possible that their "autonomy-supportive language advantage" was *actually* an analogy advantage (e.g., Liao & Masters, 2001; Masters et al., 2020). This possibility clearly highlights the importance of carefully crafting instructions that only differ in terms of the primary predictor variable of interest, instructional language, in future research.

Despite having the largest sample size in an instructional language motor learning experiment to date, the results of our robust shift function, a more traditional t-test, and non-inferiority test were inconclusive. Using our smallest effect size of interest ($d = .4$) as the noninferiority bound, the effect size at delayed retention in the present experiment is inconsistent with all effects larger than $d = \pm.43$. Although this is bigger than our pre-registered smallest effect size of interest, this test would reject the median effect size previously found in motor learning research ($d = .63$ by Lohse et al., 2016). As such, future research investigating the impact of instructional language on motor skill acquisition likely requires larger sample sizes than that used in the present experiment and what is commonly found in motor learning research (see Lohse et al., 2016; McKay, Bacelar, et al., 2023 for discussions). As such, it is critical that motor learning scientists justify their sample sizes in future research (e.g., Lakens, 2022; McKay, Corson, et al., 2023) and when using *a priori* power analyses to ensure all relevant information is reported in a reproducible manner (McKay, Bacelar, et al., 2023).

When examining the stacking time distributions for each group in retention (see Figure 1A), it is clear that the spread of the data in the two distributions is different. Such differences can be masked when researchers only use standard summary statistics such as the mean (see Anscombe, 1973 for the famous Anscombe's quartet example). Although all adjusted decile comparisons in our primary shift function analysis were not significant, there were some interesting trends that could have theoretical and/or practical significance for future work. Specifically, there was a trend for better performance with controlling language for the participants who were in the fastest (i.e., more skilled) stacking time decile (unadjusted $p = .051$) and a trend for better performance with autonomy-supportive language for the participants who were in the slowest (i.e., less skilled) stacking time decile (unadjusted $p = .017$). This pattern suggests that the motor learning benefits of different instructional language wording may potentially interact with skill level; however, a large N experiment would be required to adequately test this hypothesis. If this hypothesis could be empirically supported, it would be incompatible with OPTIMAL theory as Wulf and

Lewthwaite (2016) predicted that autonomy-supportive instructional language is beneficial irrespective of skill level. A possible explanation for why less skilled individuals could benefit from autonomy-supportive instructions compared to more skilled individuals benefiting from controlling language instructions is that the former may act as a buffer against poor performance by allowing learners to persevere and remain engaged in the task during practice. Thus, future work in this area should consider including behavioural, neural, and/or psychological measures related to task engagement (e.g., Fairclough et al., 2009; Leiker et al., 2016; O'Brien & Toms, 2009). Additionally, motor learning scientists may want to consider leveraging modern and robust statistical tools (Wilcox, 2021) in their work as these techniques may provide greater insight and a more nuanced understanding of their data.

Despite the prominent role of autonomy-support facilitating motor performance and learning in OPTIMAL theory, the higher perceptions of autonomy in our autonomy-supportive language group (see Figure 3A) did not translate into superior performance in either acquisition or retention compared to the controlling language group. The higher reported perceived autonomy scores in the autonomy-supportive language group serves as a manipulation check and is also consistent with findings from previous research (e.g., Reeve & Tseng, 2011) and multiple meta-analyses (e.g., Mossman et al., 2022; Ng et al., 2012; Okada, 2021; Su & Reeve, 2011). However, our estimate of this effect on perceptions of autonomy is much smaller than previous estimates. The estimated size in the present experiment is $d = .27$ [.05, .49], which is outside the 95% confidence interval around Su and Reeve's (2011) estimate of $d = .63$ [.43, .83]. A potential explanation for our smaller estimate is that Su and Reeve (2011) identified five components that can make instructions autonomy-supportive: 1) use non-controlling language, 2) acknowledge negative feelings, 3) nurture inner motivational resources, 4) provide meaningful rationales, and 5) offer choices; and many of the experiments in their meta-analysis included either four or all five components. In contrast, our instructions only included the first three components. Although this suggests that not all components may be necessary to have a positive influence on perceptions of autonomy, the strength of the effect may scale with the number of components incorporated in the instructions. This may be important for seeing differences in motor performance and learning. Future research would be needed to test this possibility. Another possibility for the smaller effect size and lack of performance differences in acquisition and retention is that participants received the same pre-recorded, group-specific instructions throughout practice. During skill acquisition outside of a lab, coaches likely alter the wording of their instructions in a more dynamic way to meet an athlete's needs. Thus, future research could test this idea by having slight variations in the instructions each time they are provided to the learners during acquisition.

In OPTIMAL theory, autonomy-support is also predicted to facilitate performance by enhancing expectancies. We did not find support for this prediction in the present experiment as there were no group differences in self-reported perceptions of competence (see Fig. 3B). This differs from Hooyman et al. (2014) who reported enhanced expectancies following autonomy-supportive instructional language compared to their controlling language group. A potential explanation for this discrepancy in findings might relate to measuring different psychological constructs as proxies for enhanced expectancies. Specifically, we measured perceptions of competence whereas Hooyman and colleagues measured self-efficacy. It is worth noting that the positive effect on self-efficacy in Hooyman et al. (2014) was quite transient as this difference between autonomy-supportive and controlling language on Day 1 did not persist on Day 2 in their experiment.[3]Another potential reason for our differences with Hooyman and colleagues might relate to task performance during practice. Hooyman et al. (2014) found superior performance in the autonomy-support group compared to the controlling language group during acquisition, which likely contributed to them reporting higher self-efficacy. In contrast, we did not find a group difference in task performance during practice. When considering this, it is perhaps not too surprising that we did not find differences in perceptions of competence when actual task performance was similar between our autonomy-supportive and controlling language groups. We also did not find higher instrinsic motivation in the autonomy-supportive language group compared to the controlling language group (see Fig. 3C), a finding that is consistent with recent large *N*, and often preregistered, experiments investigating the role of autonomy-supportive manipulations on motor learning (e.g., Bacelar et al., 2022; St. Germain et al., 2023; St. Germain et al., 2022). Lastly, we did not see the expected relationship between self-reported scores for any of the measured psychological constructsbefore retention with performance on the delayed retention test (see Figs. 3D-F). Taken together, these findings are difficult to reconcile with key predictions in OPTIMAL theory (Wulf & Lewthwaite, 2016) where autonomy-supportive practice conditions should enhance expectancies and increase intrinsic motivation relative to autonomy-thwarting practice conditions.

A potential limitation of the current experiment is the lack of a neutral language group. We did not include a neutral language group for several reasons. First, the inclusion of a third group would have substantially increased the required sample size (from $N = 156$ to $N = 246$) to investigate our smallest effect size of interest with adequate power. Second, as such an increase in sample size would have exceeded our resource constraints (Lakens, 2022; Lenth, 2001), we instead decided to conduct a large *N* experiment that focused on the ends of the instructional language continuum, or in other words, the biggest potential difference. Third, in both Reeve and Tseng (2011) and Hooyman et al. (2014), the key differences were between the autonomy-supportive language group and the controlling language group. Lastly, autonomy-supportive and controlling language are often used in real-world settings such as physiotherapy (e.g., Murray et al., 2015) and coaching (e.g., Bartholomew et al., 2009; Carroll & Allen, 2021), with little inclusion of neutral language. For these reasons, we contend that a neutral language group would not have added enough value to offset the costs associated with the substantial increase in sample size. Another potential limitation is that although our autonomy-supportive manipulation was significant, the estimated magnitude of this effect was quite small relative to past research. We attribute this difference to our much larger sample size compared to other motor learning research (e.g., Hooyman et al., 2014), which allowed for greater precision in our estimates. Nevertheless, future research in this area should continue to use large N designs paired with all five components that can make instructions autonomy-supportive as identified by Su and Reeve (2011).

In conclusion, we did not find a motor performance and learning advantage of instructions with autonomy-supportive language compared to controlling language. This finding is inconsistent with past motor learning research (e.g., Hooyman et al., 2014) and the OPTIMAL theory of motor learning (Wulf & Lewthwaite, 2016). Despite no motor performance or learning differences, we did find higher perceptions of autonomy in the participants that received autonomy-supportive instructional language compared to those that received controlling instructional language. While the primary goal of most motor learning interventions is a relatively permanent change in the capability for skill (Schmidt & Lee, 2019), it is worth noting that in some situations autonomy-support in and of itself might be a desired affective outcome (e.g., Ste-Marie et al., 2020). In such situations, autonomy-supportive instructional language could be paired with another form of practice that has more reliable effects on motor learning. Our perceived competence and intrinsic motivation data were also not consistent with OPTIMAL theory. While we do not discount the importance of motivation for motor skill acquisition, based on the current data we suggest that these motivational factors may instead have an *indirect* (e.g., Salmoni et al., 1984) rather than the argued *direct* (e.g., Wulf & Lewthwaite, 2016) influence on motor skill learning.

Author contributions (CRediT Taxonomy) Conceptualization: Laura St. Germain, Brad McKay, Denver M.Y. Brown, Michael J. Carter; Data curation: Laura St. Germain, Michael J. Carter; Formal analysis: Laura St. Germain; Funding acquisition: Laura St. Germain, Michael J. Carter; Investigation: Laura St. Germain, Lidia Barbera, Chitrini Tandon, Jeswende Seedu, Chantal Carrillo; Methodology: Laura St. Germain, Brad McKay, Denver M.Y. Brown, Michael J. Carter; Project administration: Laura St. Germain, Michael J. Carter; Software: Laura St. Germain, Brad McKay, Michael J. Carter; Supervision: Michael J. Carter; Validation: Brad McKay, Denver M.Y. Brown, Michael J. Carter; Visualization: Laura St. Germain, Michael J. Carter; Writing – original draft: Laura St. Germain, Brad McKay, Lidia Barbera, Chitrini Tandon, Jeswende Seedu, Chantal Carrillo, Denver M.Y. Brown, Michael J. Carter; Writing – review & editing: Laura St. Germain, Brad McKay, Lidia Barbera, Chitrini Tandon, Jeswende Seedu, Chantal Carrillo, Denver M.Y. Brown, Michael J. Carter

Funding This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada (RGPIN-2018-05589; MJC) and McMaster University (MJC). LSG was supported by an NSERC Postgraduate Scholarship.

Availability of data and materials Data and R scripts are available at https://github.com/cartermaclab/expt_instructional-language.

Declarations

Conflict of interest All authors have no competing interests to declare.

Ethical approval This study was approved the the Research Ethics Board at McMaster University.

Consent to participate Informed consent was obtained from every participant.

References

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, *27*(1), 17–21. https://doi.org/10.1080/00031305.1973.10478966

Aphalo, P. J. (2022). *Ggpmisc: Miscellaneous extensions to 'ggplot2'* [R package version 0.5.2]. https://CRAN.R-project.org/package=ggpmisc

Bacelar, M. F., Parma, J. O., Cabral, D., Daou, M., Lohse, K. R., & Miller, M. W. (2022). Dissociating the contributions of motivational and information processing factors to the self-controlled feedback learning benefit. *Psychology of Sport and Exercise*, *59*, 102119. https://doi.org/10.1016/j.psychsport.2021.102119

Bartholomew, K. J., Ntoumanis, N., & Th⊘gersen-Ntoumani, C. (2009). A review of controlling motivational strategies from a self-determination theory perspective: Implications for sports coaches. *International Review of Sport and Exercise Psychology*, *2*(2), 215–233. https://doi.org/10.1080/17509840903235330

Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, *14*(1), 99–116. https://doi.org/10.1002/jrsm.1594

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 16. https://doi.org/10.5334/joc.72

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–76.

Carroll, M., & Allen, J. (2021). 'Zooming in' on the antecedents of youth sport coaches' autonomy-supportive and controlling interpersonal behaviours: A multimethod study. *International Journal of Sports Science & Coaching*, *16*(2), 236–248. https://doi.org/10.1177/1747954120958621

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., & Ste-Marie, D. M. (2017). Not all choices are created equal: Task-relevant choices enhance motor learning compared to task-irrelevant choices. *Psychonomic Bulletin & Review*, *24*(6), 1879–1888. https://doi.org/10.3758/s13423-017-1250-7

Champely, S. (2020). *Pwr: Basic functions for power analysis* [R package version 1.3-0]. https://CRAN.R-project.org/package=pwr

Chiviacowsky, S., & Wulf, G. (2002). Self-controlled feedback: Does it enhance learning because performers get feedback when they need it? *Research Quarterly for Exercise and Sport*, *73*(4), 408–415. https://doi.org/10.1080/02701367.2002.10609040

Chiviacowsky, S., & Wulf, G. (2005). Self-controlled feedback is effective if it is based on the learner's performance. *Research Quarterly for Exercise and Sport*, *76*(1), 42–48. https://doi.org/10.1080/02701367.2005.10599260

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Fairclough, S. H., Ewing, K. C., & Roberts, J. (2009). Measuring task engagement as an input to physiological computing. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1–9. https://doi.org/10.1109/ACII.2009.5349483

Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, *69*(3), 635–640. https://doi.org/doi.org/10.2307/2335999

Harrell Jr, F. E. (2023). *Hmisc: Harrell miscellaneous* [R package version 5.1-0]. https://CRAN.R-project.org/package=Hmisc

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–802.

Hooyman, A., Wulf, G., & Lewthwaite, R. (2014). Impacts of autonomy-supportive versus controlling instructional language on motor learning. *Human Movement Science*, *36*, 190–198. https://doi.org/10.1016/j.humov.2014.04.005

Kassambara, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests* [R package version 0.7.2]. https://CRAN.R-project.org/package=rstatix

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Leiker, A. M., Bruzi, A. T., Miller, M. W., Nelson, M., Wegman, R., & Lohse, K. R. (2016). The effects of autonomous difficulty selection on engagement, motivation, and learning in a motion-controlled video game task. *Human Movement Science*, *49*, 326–335. https://doi.org/10.1016/j.humov.2016.08.005

Leiker, A. M., Pathania, A., Miller, M. W., & Lohse, K. R. (2019). Exploring the neurophysiological effects of self-controlled practice in motor skill learning. *Journal of Motor Learning and Development*, *7*(1), 13–34. https://doi.org/10.1123/jmld.2017-0051

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, *55*(3), 187–193. https://doi.org/10.1198/000313001317098149

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.8.5]. https://CRAN.R-project.org/package=emmeans

Lewthwaite, R., Chiviacowsky, S., Drews, R., & Wulf, G. (2015). Choose to move: The motivational impact of autonomy support on motor learning. *Psychonomic Bulletin & Review*, *22*(5), 1383–1388. https://doi.org/10.3758/s13423-015-0814-7

Liao, C.-M., & Masters, R. S. (2001). Analogy learning: A means to implicit motor learning. *Journal of Sports Sciences*, *19*(5), 307–319. https://doi.org/10.1080/02640410152006081

Lohse, K., Buchanan, T., & Miller, M. (2016). Underpowered and overworked: Problems with data analysis in motor learning studies. *Journal of Motor Learning and Development*, *4*(1), 37–58. https://doi.org/10.1123/jmld.2015-0010

Maier, M. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, *28*(1), 107–122.

Masters, R. S., van Duijn, T., & Uiga, L. (2020, November). Advances in implicit motor learning. In N. Hodges & A. Williams (Eds.), *Skill acquisition in sport: Research, theory, and practice* (3rd ed.). Routledge. https://doi.org/10.4324/9781351189750-5

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413

McDonough, M. H., & Crocker, P. R. (2007). Testing self-determined motivation as a mediator of the relationship between psychological needs and affective and behavioral outcomes. *Journal of Sport and Exercise Psychology*, *29*(5), 645–663. https://doi.org/10.1123/jsep.29.5.645

McKay, B., Bacelar, M. F. B., Parma, J. O., Miller, M. W., & Carter, M. J. (2023). The combination of reporting bias and underpowered study designs has substantially exaggerated the motor learning benefits of self-controlled practice and enhanced expectancies: A meta-analysis. *International Review of Sport and Exercise Psychology*, 1–21. https://doi.org/10.1080/1750984X.2023.2207255

McKay, B., Bacelar, M. F., & Carter, M. J. (2023). On the reproducibility of power analyses in motor behavior research. *Journal of Motor Learning and Development*, *11*(1), 29–44.

McKay, B., Corson, A., Vinh, M.-A., Jeyarajan, G., Tandon, C., Brooks, H., Hubley, J., & Carter, M. J. (2023). Low prevalence of a priori power analyses in motor behavior research. *Journal of Motor Learning and Development*, *11*(1), 15–28.

McKay, B., & Ste-Marie, D. M. (2020). Autonomy support and reduced feedback frequency have trivial effects on learning and performance of a golf putting task. *Human Movement Science*, *71*, 102612. https://doi.org/10.1016/j.humov.2020.102612

McKay, B., & Ste-Marie, D. M. (2022). Autonomy support via instructionally irrelevant choice not beneficial for motor performance or learning. *Research Quarterly for Exercise and Sport*, *93*(1), 64–76. https://doi.org/10.1080/02701367.2020.1795056

McKay, B., Yantha, Z. D., Hussien, J., Carter, M. J., & Ste-Marie, D. M. (2022). Meta-analytic findings in the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value. *Meta-Psychology*, *6*. https://doi.org/10.15626/MP.2021.2803

Mossman, L. H., Slemp, G. R., Lewis, K. J., Colla, R. H., & O'Halloran, P. (2022). Autonomy support in sport and exercise settings: A systematic review and meta-analysis. *International Review of Sport and Exercise Psychology*, 1–24. https://doi.org/10.1080/1750984X.2022.2031252

Murray, A., Hall, A. M., Williams, G. C., McDonough, S. M., Ntoumanis, N., Taylor, I. M., Jackson, B., Matthews, J., Hurley, D. A., & Lonsdale, C. (2015). Effect of a self determination theory–based communication skills training program on physiotherapists' psychological support for their patients with chronic low back pain: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*, *96*(5), 809–816. https://doi.org/10.1016/j.apmr.2014.11.007

Ng, J. Y. Y., Ntoumanis, N., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Duda, J. L., & Williams, G. C. (2012). Self-determination theory applied to health contexts: A meta-analysis. *Perspectives on Psychological Science*, *7*(4), 325–340. https://doi.org/10.1177/1745691612447309

Ng, J. Y., Lonsdale, C., & Hodge, K. (2011). The Basic Needs Satisfaction in Sport Scale (BNSSS): Instrument development and initial validity evidence. *Psychology of Sport and Exercise*, *12*(3), 257–264. https://doi.org/10.1016/j.psychsport.2010.10.006

O'Brien, H. L., & Toms, E. G. (2009). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, *61*(1), 50–69. https://doi.org/10.1002/asi.21229

Okada, R. (2021). Effects of perceived autonomy support on academic achievement and motivation among higher education students: A Meta-analysis. *Japanese Psychological Research*, jpr.12380. https://doi.org/10.1111/jpr.12380

Pedersen, T. L. (2022). *Patchwork: The composer of plots* [R package version 1.1.2]. https://CRAN.R-project.org/package=patchwork

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Re, A. C. D. (2013). *Compute.es: Compute effect sizes* [R package version 0.2-5]. https://cran.r-project.org/package=compute.es

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, *44*(3), 159–175. https://doi.org/10.1080/00461520903028990

Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, *95*(2), 375–392. https://doi.org/10.1037/0022-0663.95.2.375

Reeve, J., & Tseng, C.-M. (2011). Cortisol reactivity to a teacher's motivating style: The biology of being controlled versus supporting autonomy. *Motivation and Emotion*, *35*(1), 63–74. https://doi.org/10.1007/s11031-011-9204-2

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. https://doi.org/10.1111/ejn.13610

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2021). The percentile bootstrap: A primer with step-by-step instructions in R. *Advances in Methods and Practices in*

Psychological Science, *4*(1), 2515245920911881. https://doi.org/10.1177/2515245920911881

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2023). An introduction to the bootstrap: A versatile method to make inferences by using data-driven simulations. *Meta-Psychology*, *7*. https://doi.org/10.15626/MP.2019.205

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.

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037110003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, *95*(3), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

Sanli, E. A., Patterson, J. T., Bray, S. R., & Lee, T. D. (2013). Understanding self-controlled motor learning protocols through the self-determination theory. *Frontiers in Psychology*, *3*(JAN), 1–17. https://doi.org/10.3389/fpsyg.2012.00611

Schmidt, R. A., & Lee, T. D. (2019). *Motor learning and performance: From principles to application* (6th ed.). Human Kinetics Publishers.

Schuirmann, D. J. (1987). A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of*

Pharmacokinetics and Biopharmaceutics, *15*(6), 657–680. https://doi.org/10.1007/BF01068419

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 Word Solution* (tech. rep.). Rochester, NY. https://doi.org/10.2139/ssrn.2160588

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *Afex: Analysis of factorial experiments* [R package version 1.3-0]. https://CRAN.R-project.org/package=afex

St. Germain, L., McKay, B., Poskus, A., Williams, A., Leshchyshen, O., Feldman, S., Cashaback, J. G. A., & Carter, M. J. (2023). Exercising choice over feedback schedules during practice is not advantageous for motor learning. *Psychonomic Bulletin & Review*, *30*, 621–633. https://doi.org/10.3758/s13423-022-02170-5

St. Germain, L., Williams, A., Balbaa, N., Poskus, A., Leshchyshen, O., Lohse, K. R., & Carter, M. J. (2022). Increased perceptions of autonomy through choice fail to enhance motor skill retention. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(4), 370–379. https://doi.org/10.1037/xhp0000992

Ste-Marie, D., Carter, M., & Yantha, Z. (2020, November). Self-controlled learning: Current findings, theoretical perspectives, and future directions. In N. Hodges & A. Williams (Eds.), *Skill acquisition in sport: Research, theory, and practice* (3rd ed.). Routledge. https://doi.org/10.4324/9781351189750-7

Ste-Marie, D. M., Lelievre, N., & St. Germain, L. (2020). Revisiting the applied model for the use of observation: A review of articles spanning 2011–2018. *Research Quarterly for Exercise and Sport*, *91*(4), 594–617. https://doi.org/10.1080/02701367.2019.1693489

Su, Y.-L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs designed to support autonomy. *Educational Psychology Review*, *23*(1), 159–188. https://doi.org/10.1007/s10648-010-9142-7

Torchiano, M. (2020). *Effsize: Efficient effect size computation* [R package version 0.8.1]. https://doi.org/10.5281/zenodo.1480624

Tsagris, M., & Frangos, C. (2020). *Cronbach: Cronbach's alpha* [R package version 0.1]. https://CRAN.R-project.org/package=Cronbach

Ushey, K. (2023). *Renv: Project environments* [R package version 0.17.2]. https://CRAN.R-project.org/package=renv

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilcox, R. R. (2021). *Introduction to robust estimation and hypothesis testing* (Fifth). Academic Press.

Wilcox, R. R., & Rousselet, G. A. (2023). An updated guide to robust statistical methods in neuroscience. *Current Protocols*, *3*(3), e719. https://doi.org/10.1002/cpz1.719

Wulf, G., Freitas, H. E., & Tandy, R. D. (2014). Choosing to exercise more: Small choices increase exercise engagement. *Psychology of Sport and Exercise*, *15*(3), 268–271. https://doi.org/10.1016/j.psychsport.2014.01.007

Wulf, G., Iwatsuki, T., Machin, B., Kellogg, J., Copeland, C., & Lewthwaite, R. (2018). Lassoing skill through learner choice. *Journal of Motor Behavior*, *50*(3), 285–292. https://doi.org/10.1080/00222895.2017.1341378

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, *23*(5), 1382–1414. https://doi.org/10.3758/s13423-015-0999-9

Yantha, Z. D., McKay, B., & Ste-Marie, D. M. (2022). The recommendation for learners to be provided with control over their feedback schedule is questioned in a self-controlled learning paradigm. *Journal of Sports Sciences*, *40*(7), 769–782. https://doi.org/10.1080/02640414.2021.2015945

Zhu, H. (2021). *Kableextra: Construct complex table with 'kable' and pipe syntax* [R package version 1.3.4]. https://CRAN.R-project.org/package=kableExtra