Seeing effort: a test of the coach's accuracy rates in predicting the number of repetitions in reserve before reaching task failure

Supplementary materials: https://osf.io/fgycv/

For correspondence: ihalperin@tauex.tau.ac.il

Aviv Emanuel [1, 2, 3], Itai Har-Nir[1, 2], Uri Obolski[4, 5, †] and Israel Halperin[1, 2, *†]

[1]Department of Health Promotion, school of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

[2]Sylvan Adams Sports Institute, Tel Aviv University, Tel-Aviv, Israel

[3]School of Psychological Sciences, Tel-Aviv University, Tel-Aviv, Israel

[4] Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[5] Department of Environmental Studies, Porter School of the Environment and Earth Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

[†] Equal senior authors

Please cite as: Emanuel, A., Har-Nir, I., Obolski, U., Halperin, I. (2022). Seeing effort: a test of the coach's accuracy rates in predicting the number of repetitions in reserve before reaching task failure. *SportRχiv*.

ABSTRACT

Background: A key role of resistance training (RT) coaches is to personalize programs based on their trainees' abilities and goals. Specifically, coaches often assess how many repetitions in reserve (RIR) trainees have until task-failure. Coaches can then modify the number of repetitions assigned per set accordingly. However, coaches' ability to predict the number of RIR is unknown.

All authors have read and approved this version of the manuscript. This article was last modified on Month 6, 2022.

Methods: We recruited 259 certified RT coaches, who were randomly assigned to watch a video of one of eight models. The models performed two sets of barbell squats, followed by two sets of preacher bicep curls, using 70% or 80% of their 1RM, to task-failure. Coaches predicted the models' RIR at 33%, 66% and 90% of the total number of repetitions that the models completed in each set. We fitted a linear mixed model with a range of predictors to the raw and absolute prediction errors as the outcomes (i.e., signed and unsigned predicted minus actual repetitions to task-failure).
Results: The overall average number of repetitions completed by the models was 13.9. The overall average absolute errors were 4.8, 2.0, and 1.2 repetitions, for the 33%, 66% and 90% time-points, respectively. Coaches' absolute prediction error increased in the bicep curl compared to the squat (1.43, 95% CI [1.13, 1.74]), whereas the absolute prediction error decreased for heavier loads (-1.17, 95% CI [-2.16, -0.19]), and in the second set of each exercise (-1.20, 95%CI [-1.38, -1.02]). Surprisingly, coaches' years of experience had a negligible effect on the absolute error (-0.020, 95% CI [-0.039, -0.0007]). Finally, coaches underestimated the RIR of trainees at early prediction time-points, but corrected this bias and even slightly overestimated the RIR at later time-points.
Conclusions: Prior coaching experience does not seem to play a substantial role in RIR predictions. However, even short-term exposures to new trainee's performing different exercises can lead to substantial improvements in coaches' RIR predictions.

Introduction

Prescribing resistance-training (RT) programs is a complex task. It requires coaches to personalize variables such as exercises, loads, number of repetitions and sets. Various predetermined RT programs have been developed over the years, targeting specific populations and outcomes.[1,2] These supply coaches with general RT outlines, simplifying the prescription processes. For example, a RT program composed of 1-3 sets, 8-12 repetitions, and 60–70% of one Repetition Maximum (RM) can be prescribed to improve strength of novice and intermediate trainees.[1] RT programs can be further personalized by modifying training variables based on real-time data.[3] This can be achieved by employing questionnaires, tracking bar velocity, or relying on the coaches' observations (i.e. "coaches eye").[3] The coach's eye can be defined as the coach's ability to monitor trainee's exercise performance for its technical execution and intensity of effort (i.e., distance from task failure). Notably, despite the growing number of programs, questionnaires, and technological tools aimed to assist in RT prescription, the "coach's eye" is still considered an important factor in successful coaching.[4,5]

A prominent RT variable subject to real-time modification is the number of repetitions to be completed per set. The maximal number of repetitions performed for a given exercise while lifting a certain percentage of 1RM varies significantly within (and between) individuals. For example, the maximal number of repetitions trainees can complete is affected by mental fatigue,[6] whether they ingested caffeine,[7] their object of focus when exercising,[8] and even if their preferred music is played in the background.[9] Due to such expected variance in training conditions, prescribing the same number of repetitions on different days may result in inconsistent levels of intensity of effort. Subsequently, this may lead to inconsistent physiological and psychological responses. Given this variance, the coach's ability to accurately estimate the intensity of effort exerted in an ongoing set is an important coaching skill. For example, a coach may notice signs of fatigue during an ongoing set based on the trainee's facial expressions, movement velocity, technique execution, and more. Consequently, a coach may instruct the trainee to terminate a set earlier than planned, or modify the loads or repetitions in subsequent sets. This process can better align the desired intensity of the RT sessions with its goals.

Despite the importance assigned to the "coach's eye" in RT,[4,5] the accuracy with which coaches predict trainees' repetitions in reserve (RIR) before reaching task failure in an ongoing set has never been studied. In this context, all research concerning RIR has focused on trainees, rather than on coaches.[10] In such studies, trainees are instructed to verbally predict the RIR, before or during a set, and their prediction accuracy is examined. If sufficient prediction accuracy is reached, then trainees can use their estimates of RIR to modify their number of repetitions in real-time.[11,12] By doing so, trainees can better account for the variability in their performance and exercise in a more personalized manner. We propose that it is also of interest to conduct analogous study designs that examine the coaches' predictions of trainees' RIR.

In view of the above, the goal of this study was to assess the accuracy of coaches in predicting trainees RIR. To this end, we recruited RT coaches, and presented them with videos of one of eight resistance-trained models performing two sets, of two exercises, with two different loads. At different time-points during the sets, the coaches predicted the models' RIR. We examined whether the following variables influenced prediction accuracy: coaching experience, timing of prediction, exercises, set number, loads, and model.

Methods

Procedures

Participants joined the survey by clicking a link sent via email. The link directed them to the Qualtrics platform (Qualtrics XM Platform, Utah, USA), in which they read and electronically signed an informed consent form. Participants were then asked whether they were certified RT coaches via one of the accredited schools in Israel ("yes/no"). Note that a RT coaching certificate in Israel consists of a yearlong course composed of ~350 hours. In case of a negative response, participants were thanked for their response and notified that the survey has ended. In case of a positive response, participants were directed to the different online platform (www.hapyak.com), in which they first answered a series of demographic questions (Table 1). Participants were then presented with the following instructions:

"You will now watch a video of a trainee performing two sets of the squat exercise and two sets of a bicep-curl exercise using 70% or 80% of the maximal load they can lift once (1RM) to task-failure. Task-failure is defined as an event in which the model terminates the set because s/he cannot complete another repetition or because s/he estimates to be unable to complete another repetition. Please note that the models in the videos have experience in resistance training, they performed all the sets on the same day, rested for about eight minutes between each set, and were to perform the concentric portion of the each repetition as fast as possible, while attempting to maintain a controlled ~2s descend. While watching the videos you will be asked to evaluate several times how many repetitions are left before the model reaches task-failure. In your answer please type the digit itself (for example, 3 and not three)."

Subsequently, each coach watched a video of a single model perform two sets to task-failure in the barbell squat, followed by two sets to task-failure in the biceps curl, using either 70% or 80% of their 1RM. Each coach was randomly assigned to watch one of 15 possible videos as there were eight models who completed two load conditions on separate days (one 70%1RM video of one model was corrupted). The videos stopped at 33%, 66% and 90% of the total repetitions completed in each set during which a question box appeared with the following question “how many repetitions are left before the model reaches task-failure?”. Coaches were required to insert a single number before the video continued. Importantly, coaches were oblivious to how many times and when relative to task-failure they were required to provide their predictions.

Participants

We recruited participants by 1) contacting and asking the accredited RT coaching schools in Israel to distribute our survey link to their alumni, and 2) posting our survey link on various Facebook groups that focus on personal training and RT. The final sample included 259 RT coaches who provided to at least 11 out of the 12 predictions of the RIR. Due to technical errors in the survey platform, complete demographic data was available for only 153 of the participants (Table 1). All procedures were approved by the Institution's Ethics Committee.

Table 1. Coach characteristics (mean±SD) (n = 259)

Age	29.8±7.5
Weight (kg)	74.1±12.7
Height (cm)	174±0.1
Average workouts per week	4.8±3.32
Gender*	46F and 107M
Hours of RT coaching per week (average)*	14.8±12.6
Years of experience in RT*	9.30±5.8

*Data available for 153 participants

Models

The RT coaches were randomly assigned to watch a video of one out of eight models, all of which had experience in RT (Table 2). The models participated in three sessions: A 1RM testing session in the squat and bicep-curl exercises, and two sessions composed of two sets of squats followed by two sets of bicep-curls to task-failure using either 70% or 80% of 1RM, performed on separate days and in a counterbalanced order. The squats were performed within a squat cage and the bicep-curls on a preacher chair. All sessions were performed in the same facility and supervised by the same experimenter at approximately the same hour of the day (±2h). A minimum of three and a maximum of eight days between sessions were allowed. Models were asked to refrain from an intense training session 24h prior to testing days that may lead to performance decrements and muscle soreness, involving the squat and bicep curls. Models were also asked to avoid a heavy meal and caffeinated drinks or supplements at least 2h before sessions and to wear fitting athletic clothing and neutral sports shoes.

At the beginning of each session, the models completed a general warmup consisting structured dynamic stretching and calisthenics, and a five-min individualized self-selected warmup. They then completed an exercise specific warmup consisting of a gradual increase of the lifted loads toward an estimated 1RM, or the target load of the session before each exercise (for a detailed account of the warmup see Emanuel et al.[13]. Models were instructed to perform the concentric portion of each repetition as fast as possible, while attempting to maintain a controlled ~2s descend until lightly touching the box below them (individually set to achieve a knee angle of approximately 60-65 degrees) in the squat, or until fully extending their elbows in the bicep curl exercise, after which they immediately began the concentric portion. In the two last sessions, eight minutes of rest were provided between sets and exercises.

Video Recordings

Video recordings were taken via two Apple iPad Air (Apple, CA, USA), fixated using a designated tripod which angles and heights were determined in the first 1RM session. All videos were recorded with an image size of 720 by 1280, with an added side illumination ~1 meter to the left of the front camera. We recorded the models from their front and from a 90 degrees angle to their left (see Fig. 1). These angles were selected as they provide relevant information on the form of the two exercises. For example, a front view enables detection of facial expressions and asymmetry between the limbs, while a side view enables detection of movement in upper and lower back. The recording setup was fixed per model across the two last sessions via tape marks on the floor, and was set at a distance of 1.5-2m, and 2.5-3m for the side and front views, respectively. The videos were then imported into Final Cut Pro HD (Version 4.5, Apple Cupertino, CA, USA), where they were synchronized, edited further, and combined into a single .MOV file. All models signed informed consent forms approving their videos being published and distributed to participants as part of this study as approved by the Institution's Ethics Committee.

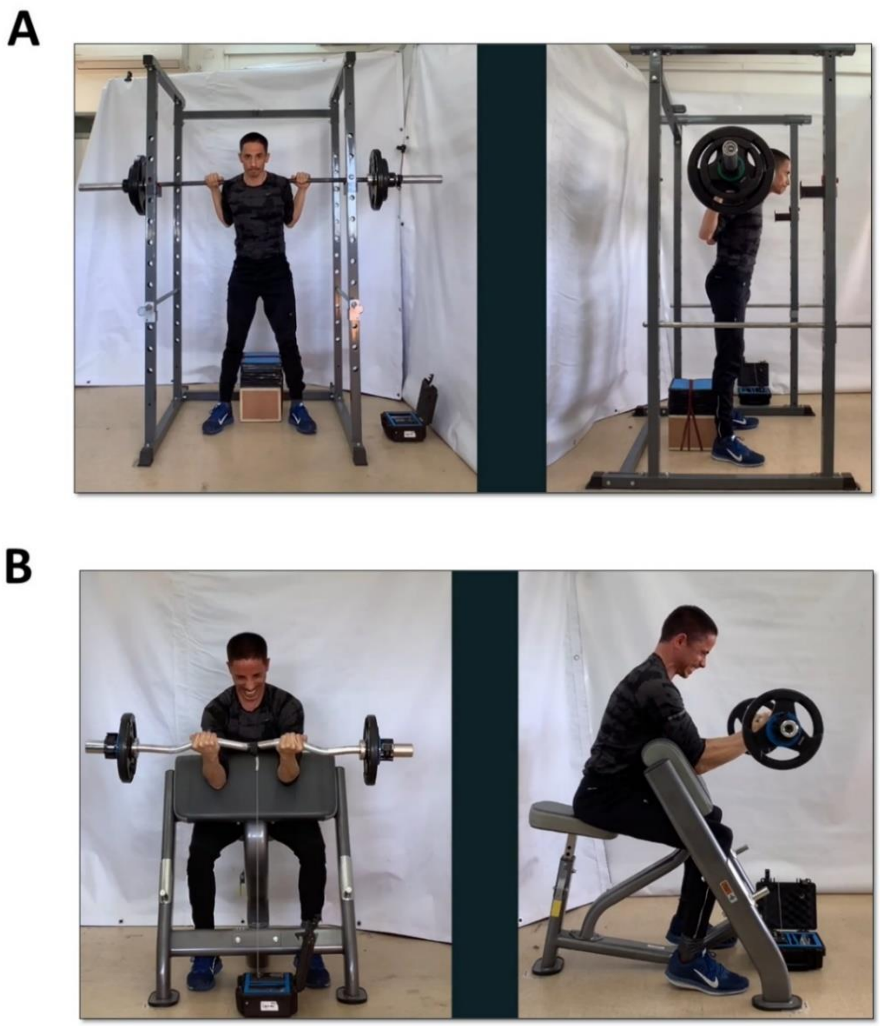

Figure 1. An example of a snapshot from the video interface trainers watched and rated at three time-points.

Table 2. Model characteristics (mean±SD).

	Men (n = 4)	Women (n = 4)
Age	31.7±6.3	29.7±9.0
Years of experience in any workout regime	16.5±5.5	18.0±11.7
Years of experience in RT	9.0±2.6	4.0±2.8
Weight	82.5±12.8	61.7±17.2
Height	180.0±6.6	163.2±8.1

Average workouts per week	4.0±0.8	2.5±0.6

Statistical analysis

Two outcome measures were defined: the predicted and the absolute predicted error made by the trainers (raw error = predicted repetitions – actual repetitions; absolute error = |predicted repetitions – actual repetitions|). The two measures are shown to convey both the direction of the prediction error and its magnitude. These were predicted during 33%, 66%, and 90% of the repetitions performed by the models before reaching task-failure. For example, assuming a model completed 15 repetitions of a given set, a coach was asked to predict the NRLF at 33% of the set – i.e. after 5 repetitions. Consequently, the model had ten repetitions left before reaching task-failure. If the coach predicted that the model had eight repetitions left, then she made a raw error of -2 repetitions and an absolute error of 2 repetitions.

We fitted a linear mixed model with the following predictors: %1RM (70% or 80%), set number (first or second), coaches experience (years as a coach), coaches gender (male or female) exercise (squat or bicep curl), prediction time-point (33%, 66%, or 90% of the repetitions performed), model gender (male or female), and the interaction between the gender of the coach and the model. We added random intercepts to account for dependencies of each coach, as they provided repeated ratings per video, and of each model, as their videos were rated by several coaches. The final regression model, comprising the same independent variables, has been fitted to both raw error and absolute error, where coaches and models are denoted by *p* and *m,* respectively:

$$absolute\ error_{pm}(raw\ error_{pm}) = \beta_{0pm} + \beta_{33\%\ time\ point} \times 33\%\ time\ point_{pm} + \beta_{90\%\ time\ point} \times 90\%\ time\ point_{pm} + \beta_{load} \times 80\%\ 1RM\ load_{pm} + \beta_{squat\ exercise} \times squat\ exercise_{pm} + \beta_{years\ of\ RT\ experience} \times years\ of\ RT\ experience_{p} + \beta_{model\ gender} \times male\ model_{m} + \beta_{coach\ gender} \times \text{male trainer}_{p} + \beta_{coach\ model\ gender\ interaction} \times model\ trainee\ gender \times coach\ gender_{pm} + \beta_{reps} \times number\ of\ repetitions\ performed_{m} + \beta_{set} \times set\ 2_{m}$$

Where the intercept is comprised of overall intercept and the coach (P_{0p}) and model (M_{0m}) random intercepts: $\beta_{0pm} = \gamma_{00} + P_{0p} + M_{0m}$.

Upon inspection of the regression model residuals, heterogeneity of the variance was detected for absolute error. Hence, linear mixed models with robust estimates of the standard errors were used in the model in which absolute error was the dependent variable. The conditional and marginal R^2 for the mixed regression models were calculated to quantify the explained variance.

Due to a technical error in the online platform, we were unable to obtain 106 data points out of the total of 259 for experience and gender variables. Thus, we also ran the same robust linear mixed models without these variables. Both models resulted in similar marginal and conditional R^2 values (Table 4).

Significance was set at p<.05. Statistical analyses and figures were carried out with R (version 4.0.2) using the following packages: robustlmm, ggplot2, Performance. All data collected are available as a Supplemental Materials file at https://osf.io/fgycv/.

Results

We plotted the actual and predicted repetitions left for each model, at each time-point, exercise set, and load in Figs. 2 and 3. The marginal average absolute error across the entire data were 4.8, 2.0, and 1.2 for the 33%, 66% and 90% time-points, whereas the overall average number of repetitions completed by the models was 13.9 (See Table 3 for descriptive data). The marginal average raw error across the entire data was -4.4, -1.0, and 1.0 (where a minus sign indicates underestimation) for the 33%, 66% and 90% time-points, respectively.

To analyze the coaches' prediction patterns, we fitted a regression model to coaches' raw and absolute prediction errors. The results of the statistical models including all n=259 participants, and of the statistical models including only n=153 coaches with complete covariate data, are reported in Tables 4 and 5. The statistical models yielded very similar results for the overlapping covariates, for both prediction errors. This is aligned with the data missingness being attributable to software issues, and consequently a missing completely at random assumption is plausible. Hence, we elaborate on the statistical model results for the n=153 coaches.

The intercept of the model was estimated at 4.52 (95% CI [3.67, 5.38]). In our formulation, the intercept represents the estimated average absolute error, at the average values of the above variables, for the bicep curl exercise, at the first set, at 70% 1RM at the 33% time-point. Progressing to the 66% and the 90% time-points reduced the error further by -1.49 (95% CI [-1.75, -1.22]) and -2.05 (95% CI [-2.40, -1.69]), respectively. When participants observed the second set of each exercise, the absolute error was further reduced by -1.20 (95% CI [-1.38, -1.02]). A significant interaction term between the exercise and load forces us to interpret these two variables slightly differently: Changing the exercise from bicep curl at 70%

1RM to squats at 70% 1RM increased the error by 1.43 (95% CI [1.13, 1.74]); changing the load in bicep curl from 70% 1RM to 80% 1RM reduced the error by -1.17 (95% CI [-2.16, -0.19]); and, changing the load in the squat from 70% 1RM to 80% 1RM reduced the error by -0.80 (95% CI [-1.15, -0.45]). The model also revealed that experience has a negligible but significant effect of reducing of the absolute error by 0.02 per year of experience (95% CI [-0.039, -0.0007]). Finally, the number of repetitions completed prior to a time-point reduced the absolute error by -0.07 (95% CI [-0.11, -0.04]) per repetition performed.

We further modeled the raw prediction error to infer over- and under-estimation of the coaches' predictions, adjusted for other variables. Using the same intercept definition as in the absolute error regression model, we found that coaches underestimated the number of repetitions remaining to failure at the 33% time-point by -4.40 (95% CI [-5.67, -3.12]). The initial underestimation decreased by 2.88 (95% CI [2.50, 3.26]) repetitions at the 66% time-point. Moreover, the initial underestimation changed to a slight overestimation at the 90% time-point, for which an additional raw error of 4.47 (95% CI [2.50, 4.98]) was estimated by the regression model.

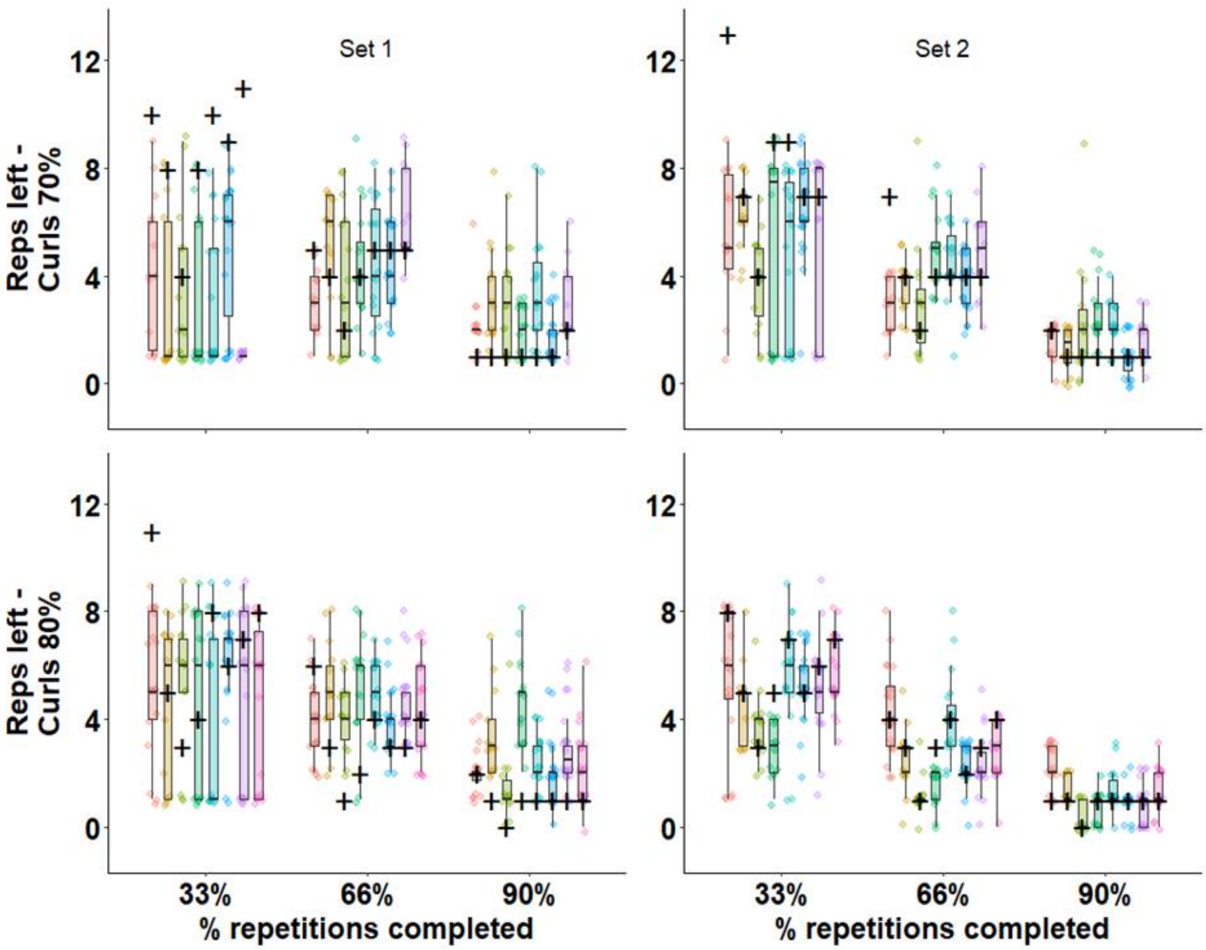

Figure 2: The predicted and actual repetitions to task failure in the squat exercise, stratified by set, load, and model. The actual number of repetitions left for each model, for a given time-point, is represented by a cross. For each model, the estimated number of repetitions given by each coach is represented by a single-colored dot. The distribution of coaches' prediction for each model is given by a consistently ordered and colored boxplot.

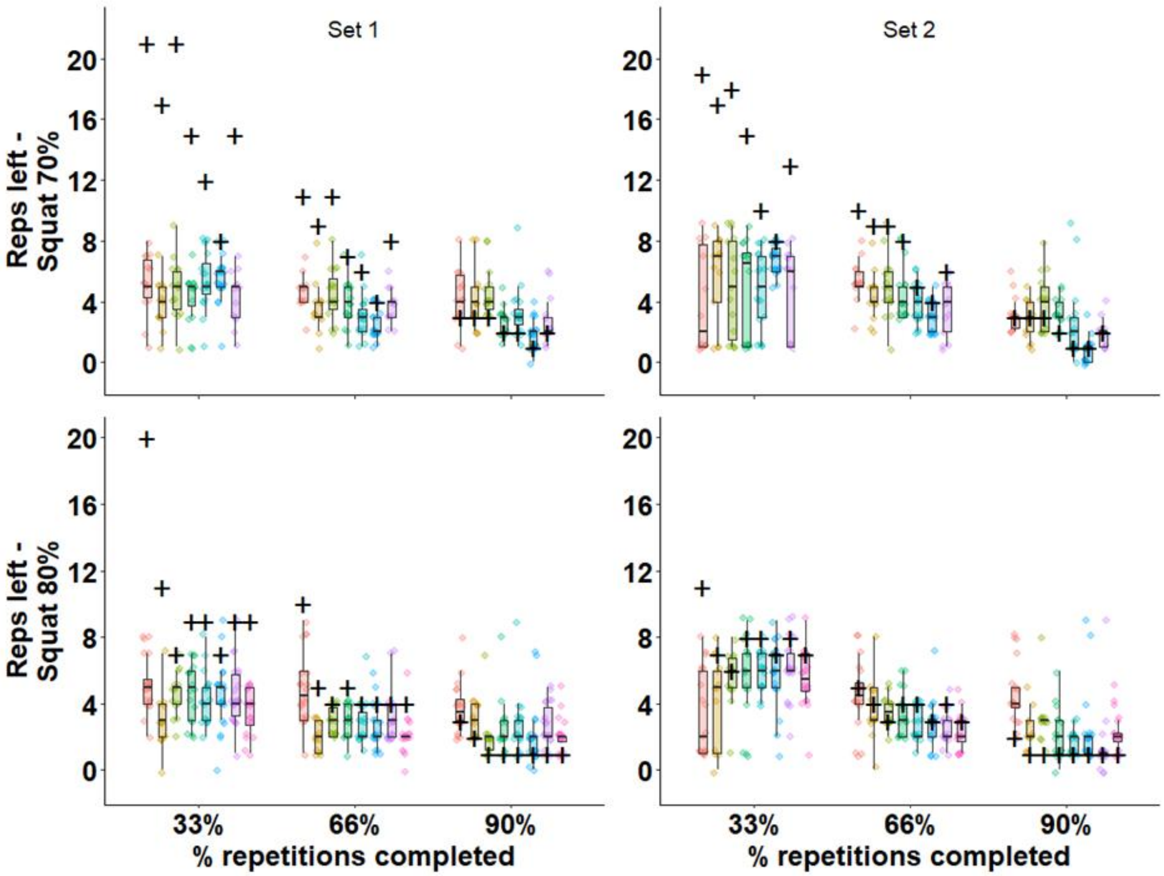

Figure 3: The predicted and actual repetitions to task failure in the bicep curl exercise, stratified by set, load, and model. The actual number of repetitions left for each model, for a given time-point, is represented by a cross. For each model, the estimated number of repetitions given by each coach is represented by a single-colored dot. The distribution of coaches' prediction for each model is given by a consistently ordered and colored boxplot.

Table 3. Descriptive statistics (mean±SD) of coaches' predicted repetitions, and actual repetitions performed by the models.

	Predicted 33%	Actual 33%	Predicted 66%	Actual 66%	Predicted 90%	Actual 90%
Squat 70% set-1	4.8±1.8	14.8±4.6	3.6±1.5	7.6±2.5	3.2±1.7	2.2±0.7
Squat 70% set-2	5.2±2.8	13.6±4.0	4.0±1.6	6.9±2.2	2.4±1.7	2.0±0.8
Squat 80% set-1	4.3±1.7	10.2±4.0	2.9±1.5	5.0±2.0	2.6±1.5	1.3±0.7
Squat 80% set-2	5.3±2.2	7.8±1.4	3.1±1.5	3.7±0.6	2.4±1.8	1.1±0.3
Curl 70% set-1	3.4±2.9	8.6± 2.0	4.4±2.1	4.3±1.0	2.5±1.6	1.1±0.3
Curl 70% set-2	5.9±2.6	7.9±2.4	3.9±1.6	4.1±1.9	1.7±1.2	1.1±0.3
Curl 80% set-1	5.0±2.8	6.7±2.3	4.4±1.5	3.7±1.3	2.6±1.5	1.0±0.4
Curl 80% set-2	4.9±1.9	5.9±1.42	2.6±1.5	3.0±0.9	1.1±0.8	0.9±0.3

Table 4. Mixed regression models predicting absolute estimation error (repetitions completed - centered)

	N = 259; Marginal R^2 = 0.37; Conditional R^2 = 0.48	N = 153; Marginal R^2 = 0.35; Conditional R^2 = 0.46
Variable	Estimate [95%CI], p-value	Estimate [95%CI], p-value
Intercept	4.53 [3.57, 5.46] < 0.001	4.52 [3.67, 5.38] < 0.001
66% vs. 33% failure proximity	-1.60 [-1.82, -1.38] < 0.001	-1.49 [-1.75, -1.22] < 0.001
90% vs. 33% failure proximity	-2.07 [-2.36, -1.78] < 0.001	-2.05 [-2.40, -1.69] < 0.001
Set 2 vs. Set 1	-1.21 [-1.36, -1.07] < 0.001	-1.20 [-1.38, -1.02] < 0.001
Squat vs. bicep-curls	2.23 [1.96, 2.50] < 0.001	1.43 [1.13, 1.74] < 0.001
80% vs. 70%1RM	-1.10 [-2.22, 0.013] 0.052	-1.17 [-2.16, -0.19] 0.022
Repetitions completed (centered)	-0.11 [-0.14, -0.08] < 0.001	-0.07 [-0.11, -0.04] < 0.001

Years of training experience (centered)	-	-0.020 [-0.039, -0.0007] 0.042
Male vs. female model	-0.59 [-1.70, 0.51] 0.266	-0.47 [-1.46, 0.52] 0.324
Male vs. female trainer	-	0.03 [-0.23, 0.30] 0.810
Male vs. female trainer X Male vs. female model	-	0.03 [-0.35, 0.43] 0.843
Squat vs. bicep-curls X 80% vs. 70%1RM	-1.37 [-1.67, -1.06] < 0.001	-0.80 [-1.15, -0.45] < 0.001

CI – confidence interval.

Table 5. Mixed regression models predicting raw estimation error (repetitions completed - centered)

	N = 259; Marginal R^2 = 0.44; Conditional R^2 = 0.54	N = 153; Marginal R^2 = 0.42; Conditional R^2 = 0.52
Variable	Estimate [95%CI], p-value	Estimate [95%CI], p-value
Intercept	-4.26 [-5.54, -2.99] < 0.001	-4.40 [-5.67, -3.12] < 0.001
66% vs. 33% failure proximity	3.02 [2.71, 3.32] < 0.001	2.88 [2.50, 3.26] < 0.001
90% vs. 33% failure proximity	4.81 [4.41, 5.21] < 0.001	4.47 [3.97, 4.98] < 0.001
Set 2 vs. Set 1	0.64 [0.44, 0.84] < 0.001	0.59 [0.34, 0.84] < 0.001
Squat vs. bicep-curls	-3.48 [-3.85, -3.11] < 0.001	-2.75 [-3.19, -2.32] < 0.001
80% vs. 70%1RM	1.34 [-0.15, 2.83]	1.56 [0.09, 3.03]

	0.074	0.038
Repetitions completed (centered)	0.08 [0.04, 0.12] < 0.001	0.06 [0.01, 0.11] 0.0183
Years of training experience (centered)	-	0.002 [-0.025, 0.029] 0.890
Male vs. female model	0.59 [-0.88, 2.07] 0.398	0.58 [-0.90, 2.07] 0.410
Male vs. female trainer	-	0.01 [-0.36, 0.40] 0.927
Male vs. female trainer X Male vs. female model	-	0.02 [-0.53, 0.58] 0.928
Squat vs. bicep-curls X 80% vs. 70%1RM	1.84 [1.43, 2.26] < 0.001	1.29 [0.79, 1.80] < 0.001

CI – confidence interval.

Discussion

In this study, we analyzed coaches' prediction of the RIR of trained models who completed two sets, of two different exercises, using two different loads. We found that the following variables improved coaches' absolute prediction error: later predictions during sets, bicep curl (compared to squat), using heavier loads, the second set, more completed repetitions at the time of predictions, and greater coaching experience, although the latter had a negligible effect. Furthermore, analysis of the raw error showed that coaches tended to underestimate the RIR in the first and second prediction time-points, but reverted to overestimation in the final prediction point.

The higher prediction accuracy observed in the bicep curl compared to the squat can stem from several reasons. First, the dynamic portion of the bicep curl occurs in the elbow joint; whereas in the squat it occurs in the ankle, knee and hip joints. Accordingly, it is possible that coaches directed their attention to a smaller area in which movement occurred, and extracted information that led to better predictions. Second, the extent to which trainees can modify exercise execution, and thus compensate for muscular fatigue, differs between the two exercises. In the biceps curl, the arms are fixated to the preacher curl device, making it difficult to modify exercise execution. Hence, when fatigue of the elbow flexors accumulates, trainees are restricted in their ability to involve other muscle groups to assist in completing further repetitions. Conversely, when squatting, trainees are less restricted in their movements, and can thus alter exercise execution and increase the involvement of different muscle groups.[14,15] For example, to compensate for quadriceps fatigue, trainees may implement greater hip flexion, which leads to increased involvement of the hamstring and gluteal muscles.[16,17] We therefore speculate that exercises that offer fewer opportunities for exercise modifications lead to better RIR predictions.

The better RIR prediction in the second compared to the first set suggests a learning effect. It is likely that coaches were able to collect information about the models' abilities in the first set, leading to improved predictions in the second set. Note that prior to this study, coaches did not observe the models perform the exercises, and received minimal information about their abilities. It is thus likely that predictions would have further improved if coaches had received greater exposure to the models performing the exercises. Surprisingly, the effect of coaching experience on prediction accuracy was negligible, although statistically significant. This result aligns with a meta-analysis inspecting trainees' prediction of the RIR, in which trainees' RT experience was negligibly associated with accurate predictions of the RIR.[10] Collectively, the immediate improvements in predictions that occurred over sets, coupled with the negligible effects of coaching experience, suggests that prediction accuracy of RIR does not generalize well across trainees. Rather, this ability likely depends on coaches' specific knowledge of their trainees' unique abilities.

The prediction accuracy was higher when provided at later time-points, or after more repetitions were completed. This may stem from several reasons: First, coaches inferred that predictions at later time-points, or after more repetitions, are closer to task failure. Second, coaches improved at identifying signs of fatigue exhibited by the models. Third, and complementary to the second reason, the models exhibited greater signs of fatigue, which is associated with alteration in movement execution.[18,19] The latter reason can also explain the improvement in prediction under heavier loads, where models may have exhibited greater signs of fatigue at each repetition. Unfortunately, the current study design cannot disentangle the effects of these proposed reasons.

With respect to the raw prediction error, coaches changed their prediction patterns over successive prediction time-points. Coaches significantly underestimated the RIR in the first time-point. This bias remained during the second time-point, although its magnitude decreased. By the third time-point, coaches were the most accurate, and even slightly overestimated the RIR. While we are uncertain why coaches tend to underestimate the RIR at the beginning of sets, awareness to this bias may be of practical implications. For example, coaches can deliberately add repetitions to their predictions of RIR at early stages of a trainee's set.

This study has several limitations worthy of discussion. First, coaches typically observe trainees complete exercises in person, rather than viewing them on a screen. While we provided coaches with two viewing angles of the exercising models, in-person coaching allows for varying viewing angles and other nuanced information that is absent from a screen. Future research could attempt to assess coach's predictions in a gym environment rather than via videos. Second, coaches observed the exercises and sets in a fixed order: squats preceded the bicep curls, and the first set preceded the second set. Collectively, the structure of this study design may have led to an order effect. Future research can overcome some of these limitations by presenting different segments of the videos in a randomized order. For example, presenting first the set that was performed second, or presenting the last portion of a set before the first one. This will allow testing the proficiency of coaches in predicting the RIR in isolation of the other parts of the video viewed before.

Conclusions

We have shown that the accuracy of coaches' predictions of RIR depends on a number of variables. Mainly, predictions improve when coaches provide them in later stages of a set, when using heavier loads, in later sets, and with the biceps curl. Conversely, coaching experience played a trivial role in improving prediction accuracy. These results are mostly aligned with a recent meta-analysis inspecting trainee's prediction of the RIR.[10] Prediction accuracy improved when trainees provided their predictions closer to task failure, when using heavier loads, in later sets, and was independent of trainees RT experience. In the present study, coaches also tended to underestimate the RIR in the first prediction, but this effect shrunk and eventually turned to an overestimation by the final prediction. Practically, these results suggest that the coach's ability to predict task failure is less accurate in the beginning of a set, but tends to improve as sets progress, with consecutive sets, and for sets composed of heavier loads.

Contributions

AE, IHN, and IH designed the study; AE, IHN, and IH collected the data; AE, UO, and IH analyzed the data; AE, UO, and IH wrote the manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgements

Funding information

This study was supported by a grant from the Israeli Science Foundation (1249/20).

Data and Supplementary Material Accessibility

The full data for the analyses in this work is available at https://osf.io/fgycv/.

References

1. American College of Sports Medicine. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* 2009;**41**:687–708. https://doi.org/10.1249/MSS.0b013e3181915670.
2. Fragala MS, Cadore EL, Dorgo S, Izquierdo M, Kraemer WJ, Peterson MD, et al. Resistance training for older adults: position statement from the National strength and conditioning association. *J Strength Cond Res* 2019;**33**.
3. Scott BR, Duthie GM, Thornton HR, Dascombe BJ. Training Monitoring for Resistance Exercise: Theory and Applications. *Sports Med* 2016;**46**:687–98. https://doi.org/10.1007/s40279-015-0454-0.
4. Sands WA, Wurth JJ, Hewit JK. The National Strength and Conditioning Association's (NSCA) Basics of Strength and Conditioning Manual. *NSCA Ed NSCA* 2012.
5. Suchomel TJ, Nimphius S, Bellon CR, Hornsby WG, Stone MH. Training for Muscular Strength: Methods for Monitoring and Adjusting Training Intensity. *Sports Med* 2021;**51**:2051–66. https://doi.org/10.1007/s40279-021-01488-9.

6. Graham JD, Martin Ginis KA, Bray SR. Exertion of self-control increases fatigue, reduces task self-efficacy, and impairs performance of resistance exercise. *Sport Exerc Perform Psychol* 2017;**6**:70.
7. Green JM, Wickwire PJ, McLester JR, Gendle S, Hudson G, Pritchett RC, et al. Effects of caffeine on repetitions to failure and ratings of perceived exertion during resistance training. *Int J Sports Physiol Perform* 2007;**2**:250–9. https://doi.org/10.1123/ijspp.2.3.250.
8. Marchant DC, Greig M, Bullough J, Hitchen D. Instructions to adopt an external focus enhance muscular endurance. *Res Q Exerc Sport* 2011;**82**:466–73. https://doi.org/10.1080/02701367.2011.10599779.
9. Ballmann CG, McCullum MJ, Rogers RR, Marshall MR, Williams TD. Effects of Preferred vs. Nonpreferred Music on Resistance Exercise Performance. *J Strength Cond Res* 2021;**35**:1650–1655. https://doi.org/10.1519/JSC.0000000000002981.
10. Halperin I, Malleron T, Har-Nir I, Androulakis-Korakakis P, Wolf M, Fisher J, et al. Accuracy in predicting repetitions to task failure in resistance exercise: a scoping review and exploratory meta-analysis. *Sports Med* 2021:1–14.
11. Hackett DA, Cobley SP, Halaki M. Estimation of Repetitions to Failure for Monitoring Resistance Exercise Intensity: Building a Case for Application. *J Strength Cond Res* 2018;**32**:1352–9. https://doi.org/10.1519/JSC.0000000000002419.
12. Zourdos MC, Goldsmith JA, Helms ER, Trepeck C, Halle JL, Mendez KM, et al. Proximity to Failure and Total Repetitions Performed in a Set Influences Accuracy of Intraset Repetitions in Reserve-Based Rating of Perceived Exertion. *J Strength Cond Res* 2021;**35**:S158–65. https://doi.org/10.1519/JSC.0000000000002995.
13. Emanuel A, Smukas IR, Halperin I. How one feels during resistance exercises: A repetition-by-repetition analysis across exercises and loads. *Int J Sports Physiol Perform* 2020;**16**:135–44.
14. Fry AC, Smith JC, Schilling BK. Effect of knee position on hip and knee torques during the barbell squat. *J Strength Cond Res* 2003;**17**:629–33. https://doi.org/10.1519/1533-4287(2003)017<0629:eokpoh>2.0.co;2.
15. List R, Gülay T, Stoop M, Lorenzetti S. Kinematics of the Trunk and the Lower Extremities During Restricted and Unrestricted Squats. *J Strength Cond Res* 2013;**27**:1529–1538. https://doi.org/10.1519/JSC.0b013e3182736034.
16. Hooper DR, Szivak TK, Distefano LJ, Comstock BA, Dunn-Lewis C, Apicella JM, et al. Effects of resistance training fatigue on joint biomechanics. *J Strength Cond Res* 2013;**27**:146–53. https://doi.org/10.1519/JSC.0b013e31825390da.

17. Hooper DR, Szivak TK, Comstock BA, Dunn-Lewis C, Apicella JM, Kelly NA, et al. Effects of fatigue from resistance training on barbell back squat biomechanics. *J Strength Cond Res* 2014;**28**:1127–34. https://doi.org/10.1097/JSC.0000000000000237.
18. Mudie KL, Gupta A, Green S, Clothier PJ. Adaptation of lower limb movement patterns when maintaining performance in the presence of muscle fatigue. *Hum Mov Sci* 2016;**48**:28–36. https://doi.org/10.1016/j.humov.2016.04.003.
19. Yang C, Bouffard J, Srinivasan D, Ghayourmanesh S, Cantú H, Begon M, et al. Changes in movement variability and task performance during a fatiguing repetitive pointing task. *J Biomech* 2018;**76**:212–9. https://doi.org/10.1016/j.jbiomech.2018.05.025.