INTERNATIONAL JOURNAL OF
-EXERCISE SCIENCE-

Original Research

Training Effects and Test-Retest Reliability of Dynavision Mode B in Active Adults

Jessica M. Moon[1,2†], John Pinette[1*], Aubrey Fontenot[1*], Violette Gibbs[1*], Destiny Northcutt[1*], Trevor J. Dufner[1†], and Adam J. Wells[1‡]

[1]Exercise Physiology, Intervention, and Collaboration Lab, School of Kinesiology and Rehabilitation Sciences, University of Central Florida, Orlando, FL, USA; [2]Lindenwood University, College of Science Technology and Health, St. Charles, MO, USA

*Student Investigator, ‡Established Investigator, †Early Career

Abstract

International Journal of Exercise Science 19(4): 4004, 2026. Previous research has demonstrated that the Dynavision D2 Mode A reaction time (RT) assessment requires extensive familiarization to reach peak performance beyond initial training effects. This suggests that other commonly used D2 assessments, such as the Mode B test, may require familiarization periods to accurately assess changes in performance. This distinction is critical to ensure that observed changes reflect true adaptations rather than continued task learning, particularly for practitioners monitoring performance or recovery. This study examined the effects of repeated training and detraining on reliability and performance on Mode B tests. Twenty-six recreationally active adults completed 15 training sessions (two Mode B tests per session), followed by a detraining period of one or two weeks, before completing another 15 training sessions. Total hits (Hits), number of misses, average RT per hit (AvgRT), and the number of numerical T-scope challenges correct (TSC) were used to determine performance. Reliability was assessed using intraclass correlation coefficients (ICC), standard error of measurement (SEM), and minimal detectable change (MDC). Significant improvements in performance (p <.05) were observed across the first 24 tests, with no further changes thereafter. Hits ICC to .968 and standardize spacing. Corrected sentence: "Hits (ICC = .968), AvgRT (ICC = .958), and misses (ICC = .902) demonstrated excellent test-retest reliability, while TSC (ICC = .785) demonstrated good reliability. There were no performance differences based on detraining duration during the retraining period. These findings indicate that at least twenty-four tests are necessary to establish a reliable baseline and minimize training effects when Hits, AvgRT, Misses, and TSC are used as performance outcomes to avoid misinterpretation.

Keywords: Dual-task performance, Sensorimotor training, Visual–motor processing, Performance assessment, Reliability metrics

Introduction

The ability to perceive and respond to external stimuli within a shifting and fluid environment is a key determinant of sporting success. Reaction time (RT), defined as the interval that elapses between the onset of a stimulus and the initiation of a response,[1] is a measurable component of this athletic ability that has garnered interest in sport science due to its simple implementation and ease of quantification. Consequently, it has been suggested that individuals who possess

*Corresponding Author: Adam J. Wells; Email: adam.wells@ucf.edu

greater cognitive reaction time abilities may have an advantage over their counterparts[2] and may likewise have a reduced injury risk.[3] As such, the evaluation of RT may enable coaches and practitioners to characterize athletic ability,[4] implement specific training protocols to improve RT,[5,6] or to identify and rectify weaknesses within their athletes.[3]

To date, several RT assessments and devices have been implemented in laboratory and field settings to quantify and assess changes in RT performance.[7–13] These assessments have included simple reaction time tests, choice reaction time tests, finger tapping tests, and customized sport-specific assessments that are often computer integrated with touch sensitive visual light systems. The Dynavision D2 (D2) is a light source visuomotor reaction device that has been developed to train sensory motor integration through the visual system and is popular among those wishing to assess RT. Among the testing modes that the D2 contains, Mode A and Mode B are the most commonly used within sport and rehabilitation settings. Mode A is a proactive visuomotor assessment during which individuals respond to a single stimulus as quickly as possible for a fixed duration. In contrast, the Mode B assessment introduces a dual-task demand by combining the same motor response requirements as Mode A with an additional cognitive component. During Mode B testing, participants must respond to randomly illuminated targets while simultaneously identifying and verbally reporting numerical sequences presented on a central display (T-scope). This dual-task format increases task complexity and more closely reflects the demands of competitive sport environments, where individuals must respond to dynamic visual cues while simultaneously processing additional information.[3] This added complexity may result in greater learning effects and variability in performance across repeated testing. Despite its growing use in sport science and rehabilitation,[3,14–18] the training effects and the establishment of a stable baseline performance for Mode B remain poorly characterized, particularly in the context of repeated testing. Without establishing this information, it is unclear how many tests are needed to determine a stable baseline, or whether observed deviations in performance reflect true performance improvements or continued learning.

While the overall reliability of the D2 has previously been established,[19] much of the existing research has focused on the Mode A assessment. Moreover, initial improvements in RT have been attributed to a learning effect [20,21] with repeated exposures often producing continued improvements in RT performance.[22,23] For example, Wells et al [22] examined the test-retest reliability, training, and detraining effects that were associated with the Mode A proactive visuomotor RT test and reported that 18 discrete Mode A tests were required to achieve a stable baseline performance. Consequently, the effects of repeated testing on other D2 assessments, such as the Mode B test, should be examined, in order to determine reliable baseline performance scores against which true deviations in score can be assessed.[21] To our knowledge, no previous studies have systematically examined extended training, detraining, and reliability characteristics of the Dynavision Mode B assessment across repeated exposures.

Several studies have previously sought to establish the reliability of the Mode B test.[19-21] However, these attempts remain inconclusive, as improvements in performance during initial trials suggest a learning effect that has not been clearly separated from true test reliability. Early research by Klavora et al[21] reported a significant difference in the number of hits from Test 1 to 3, and a moderate reliability coefficient (R) of 0.73 when three discrete tests were utilized, suggesting the presence of a familiarization effect during testing. In a follow-up investigation, participants completed the Mode B test on five occasions and demonstrated significant differences between Trials 1 and 2 with no further differences observed thereafter.[20] Consequently, the intraclass R coefficient for the final 4 tests improved to 0.92. Despite this improved reliability, participants were required to perform a series of 30-second practice tests prior to formal testing, with progression dependent on achieving stable performance across trials (defined as a trial in which the number of hits was less than or equal to, or more than one hit greater than the number of hits in the preceding trial). Given the extent of additional practice tests, it is difficult to draw conclusions about the specific learning effect or standalone reliability associated with the Mode B test. More recent work by Wells et al[19] examined the reliability of the Mode B test across six distinct trials and observed significant improvements during the initial trials (1-3), with no further differences observed thereafter (trials 4-6). However, participants also completed multiple Dynavision assessments in conjunction with Mode B testing, which may have influenced performance through additional learning or fatigue. Collectively, these findings highlight the need for studies that examine Mode B performance over extended repeated exposures and establish a stable baseline under controlled testing conditions.

Therefore, the purpose of this investigation was to (1) determine how many tests are needed to achieve a reliable baseline (i.e., stabilization in RT) across Mode B performance outcomes (total hits, average reaction time per hit [AvgRT], number of misses, and the number of correctly identified five-digit numbers on the T-scope challenge [TSC]), (2) calculate the minimum detectable change and standard error of measurement against which deviations in scores can be measured, and (3) examine how periods of detraining (1 and 2 weeks) influence frequency requirements for continuous testing. We hypothesize that Mode B RT performance will improve significantly during the initial training phase before stabilizing, and that there will be no impact of detraining or retraining on RT performance.

Methods

Participants

Twenty-seven individuals were enrolled, of whom twenty-six recreationally active participants (10 males and 16 females; Table 1) completed the full protocol. An *a priori* power analysis was conducted using G*Power (v3.1.9.7; Universität Düsseldorf, Germany) for a repeated-measures ANOVA with one group and ten measurement blocks. A large effect size (f = 1.035), derived from prior work using the Dynavision Mode A test (Hits: $\eta^2_p = 0.517$; Wells et al[22]), was used to estimate sample size. The analysis indicated a small minimum sample requirement (n = 2) under these assumptions ($\alpha = .05$, power = 0.80). However, the final sample of twenty-six participants substantially exceeded this threshold and is consistent with prior work in this area. Sample size estimation was based on the primary outcome of performance changes across repeated testing blocks, while reliability analyses were included as a secondary outcome to characterize measurement stability following performance stabilization. Prior to completing any study-related procedures, participants provided written informed consent approved by the Institutional Review Board at the University of Central Florida (Protocol #00004615; approval date: August 25, 2022). Participants then completed a physical activity readiness questionnaire (PAR-Q+), medical history questionnaire (MHQ), and a caffeine consumption questionnaire (CCQ) to determine eligibility. The eligibility criteria included participants completing at least 150 minutes of combined physical activity per week according to the American College of Sports Medicine (ACSM) standards for recreationally active individuals. Participants were also required to be healthy and ready for activity as determined by the PAR-Q+ and MHQ, and to have no prior experience with the D2 Device. Participants were deemed eligible for physical activity based on responses indicating no health concerns requiring medical clearance. Additionally, all participants reported having no significant vision problems; however, the use of prescriptive lenses was permitted and standardized across testing days. An average habitual caffeine consumption of <300 mg·day^{-1} was required, and participants were instructed to maintain their normal caffeine regimens. This research was carried out fully in accordance with the ethical standards of the *International Journal of Exercise Science*[24].

Table 1. Participant Demographics.

		Age (yrs.)	Height (cm)	Body Mass (kg)	Body Fat %
1-wk detraining (n=13)	M (n=6)	21.2 ± 1.2	182.0 ± 2.9	83.2 ± 10.3	18.7 ± 7.3
	F (n=7)	20.4 ± 1.1	164.6 ± 8.0	56.8 ± 12.2	25.7 ± 5.5
2-wk detraining (n=13)	M (n=4)	22.3 ± 1.9	180.9 ± 2.8	81.9 ± 8.0	20.0 ± 2.2
	F (n=9)	22.7 ± 4.4	162.6 ± 6.7	83.4 ± 28.1	36.3 ± 8.6

Data presented as mean ± standard deviation. Wk = week; M = male; F = female; Yrs. = years; cm = centimeters; kg = kilograms; % = percent.

Protocol

Participants reported to the laboratory for a total of 32 visits. Visit one consisted of informed consent and eligibility screening, which included completion of a PAR-Q+, MHQ, and a CCQ. Eligible participants returned for visit two (V2), where anthropometric assessments of height and weight were assessed using a stadiometer and scale (Health-o-meter Professional Patient Weighing Scale, Model 500 KL, Pelstar, Alsip, IL, USA), and body composition was assessed via bioelectrical impedance analysis (BIA; InBody 770, Biospace Co, Ltd. Seoul, Korea). Participants were instructed to be at least three hours fasted, well hydrated, and to void their bladder prior to BIA. All socks, shoes, and jewelry were removed before the test.

During V2, participants also completed familiarization with the Mode B visuomotor reaction time (RT) test using the D2 light board[19] (Dynavision International LLC, West Chester, OH, USA). Participants received standardized instructions delivered from a scripted protocol and were instructed to take an athletic stance with flexed knees, an upright posture, and a low center of gravity. The D2 board was then adjusted to the height of the participant so that the T-scope screen (a central display presenting numerical sequences for the cognitive task) was approximately at eye level, and the outermost target buttons were within hand's reach. Two familiarization tests were performed with postural corrections provided via verbal feedback from a research team member. The height of the D2 board was standardized across all subsequent training and detraining tests for each participant. All testing sessions were conducted in a controlled laboratory environment with consistent lighting and minimal ambient noise across visits.

Following familiarization, participants returned for 15 training sessions (two Mode B tests per session [30 total tests]). The Mode B test required participants to successfully identify and strike as many stimuli as possible in 60 seconds. Each target remained illuminated for 1 second before changing position automatically or following a successful physical hand strike. Participants were required to simultaneously recite a five-digit number presented on the center screen (T-scope). The five-digit number was presented a total of 11 times throughout the test and remained on the T-scope for 0.75 seconds. The timing and presentation of the T-scope challenges were controlled by the Dynavision system and remained consistent across all tests. Performance outcomes recorded included the number of hits, misses (targets not struck within 1 second), average RT per hit (AvgRT), and number of correctly identified five-digit numbers (TSC).

Within each session, the two Mode B tests were separated by approximately one minute, reflecting the time required to record performance outcomes and prepare for the subsequent test. Participants were instructed to maintain their normal daily routines, including caffeine intake, and no additional controls were placed on sleep or other behaviors between visits. Performance outcomes were visible to participants following each test via the Dynavision display.

After completion of the 15 training sessions, participants were randomized to one or two weeks of detraining. During the detraining period, participants did not train on the D2 device. Following detraining, participants returned to complete 15 retraining sessions (30 total tests), conducted in an identical fashion to the initial training sessions. Participants were permitted to complete one testing session per calendar day. The exact timing between sessions was not standardized and varied between participants based on scheduling availability.

Statistical Analysis

To assess training and retraining effects, the 30 tests from each phase were grouped into 10 blocks consisting of three tests per block (10 blocks in training; 10 blocks in retraining). The average of all three tests within each block was used for statistical analysis to reduce intra-session variability, rather than selecting a single best performance. Tests were grouped in sets of three to provide a more stable estimate of performance across repeated exposures while still allowing for detection of changes over time, rather than strictly reflecting individual testing sessions. Normality was assessed via the Shapiro-Wilk test. Non-normal data were $\log_{10}$ transformed and reassessed. If normality was not achieved, raw data were retained, as ANOVA is robust to mild violations.[25] If the assumption of sphericity was violated, a Greenhouse-Geisser correction was applied.

Separate one-way repeated-measures ANOVAs were used to assess differences in Hits, AvgRT, Misses, and TSC during the initial training period. To assess detraining effects, separate two-way (detraining period [one wk vs two wk] × block [retraining blocks 1–10]) repeated-measures ANCOVAs were conducted for each variable with detraining duration and block as factors. Average scores from the final block of the training period (block 10 [tests 28–30]) were used as the covariate to account for individual differences in performance prior to detraining. Least significant difference (LSD) post-hoc analyses were used for pairwise comparisons.

Main effects were evaluated using partial eta squared (η^2p) effect sizes. η^2p values were interpreted as small (0.01–0.058), medium (0.059–0.137), and large (>0.138).[26] Pairwise comparisons were further evaluated using Hedges' g (corrected for small sample size), interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). Training effects were also analyzed using Helmert contrasts (HC), comparing the mean of each level of the factor (i.e., block) to the mean of the subsequent levels.

Test-retest reliability was assessed using intraclass correlation coefficients (ICCs) derived from a two-way mixed-effects model with absolute agreement following the guidelines published by Koo and Li.[27] ICCs were calculated between block 9 (tests 25, 26, and 27) and block 10 (tests 28, 29, and 30) of the training phase, where performance had stabilized based on minimal changes across blocks (i.e., Helmert contrasts, effect sizes, and comparisons relative to SEM and MDC thresholds). Reliability metrics were calculated following the identification of stabilized performance to characterize measurement consistency, rather than to define the point of stabilization. Standard error of measurement (SEM) and minimal detectable change (MDC) values were also calculated for each variable.[28]

Significance for all analyses was accepted at an alpha level of $p \leq .05$. All analyses were conducted using IBM SPSS Statistics for Windows (version 20.0; IBM Corp, Armonk, NY) and are reported as mean (SD).

Results

Overall, performance improved significantly across early and middle training blocks for all variables (Hits, AvgRT, Misses, TSC), with the largest changes occurring prior to block 7. In later blocks, these improvements diminished, with small effect sizes and most changes falling within measurement error thresholds, indicating performance stabilization, as determined using a combination of statistical significance, effect sizes, and comparisons relative to SEM and MDC thresholds rather than relying on statistical significance alone. Following detraining, no meaningful changes were observed during the retraining period.

Mode B Training

The descriptive data for Hits, AvgRT, misses, and TSC during the training period are presented in Table 2. Shapiro-Wilk tests indicated non-normal distributions for misses and TSC. Misses were log-transformed prior to analysis, while TSC was analyzed using raw scores. Overall, significant main effects of block were observed for Hits (p <.001, η^2_p = .618), AvgRT (p <.001, η^2_p = .464), Misses (p <.001, η^2_p = .460), and TSC (p <.001, η^2_p = .269), indicating substantial performance changes across repeated testing blocks. Detailed pairwise comparisons, effect sizes (Hedges g), and comparisons relative to SEM and MDC thresholds are provided in the Supplementary Material (Appendix 1 – 4) to allow for a more comprehensive evaluation of performance changes across blocks.

Hits improved significantly across training blocks (Figure 1), with the largest increases observed during early and middle blocks. Early comparisons demonstrated moderate to very large effect sizes and a high proportion of participants exceeding MDC thresholds, indicating meaningful improvements in performance. In later blocks, changes were smaller in magnitude. Although some statistically significant differences persisted between adjacent blocks, these were characterized by small effect sizes and a low proportion of participants exceeding MDC thresholds (generally < 20%; Supplementary Material – Appendix 1). Notably, the proportion of participants exceeding MDC decreased substantially after approximately blocks 6-7 and remained low through blocks 7-10. Collectively, these findings suggest a progressive stabilization of performance rather than continued systematic improvement.

AvgRT decreased significantly across training blocks (Figure 2), reflecting faster response times with repeated exposure. Improvements were most pronounced during early blocks, with moderate effect sizes and meaningful reductions in reaction time. In the later blocks, changes in AvgRT were minimal with small or trivial effect sizes and limited proportions of participants exceeding MDC thresholds (Supplementary Material – Appendix 2). These findings suggest that improvements in RT diminished over time and stabilized during the later blocks.

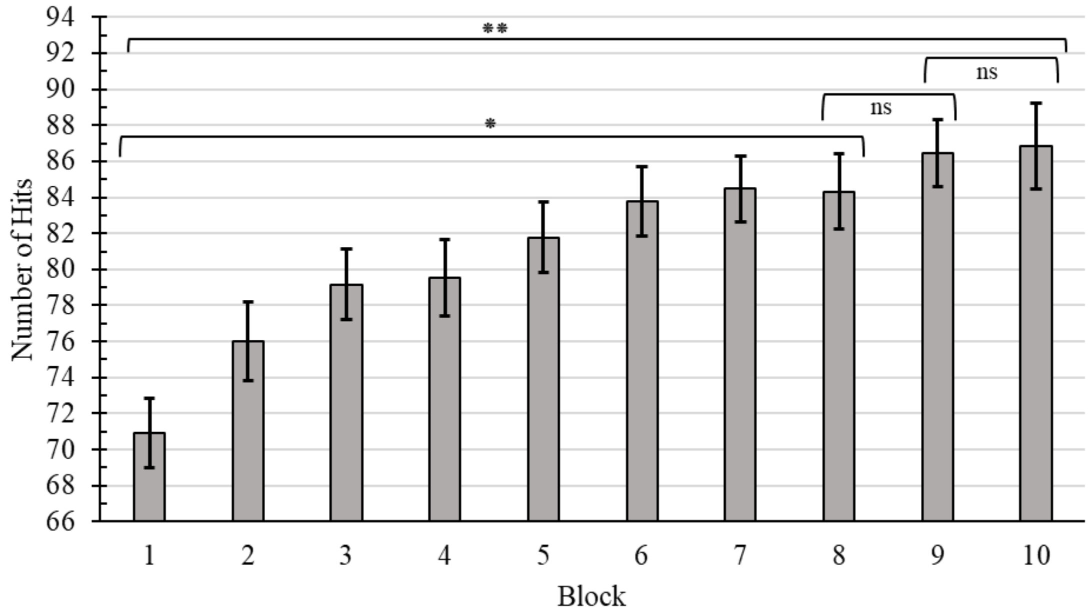

Figure 1. Changes in mean number of Total Hits during Dynavision D2 Mode B training blocks (3 tests per block). Error bars represent ± one standard error of the mean (SEM), reflecting the precision of the group mean across repeated testing blocks (see Table 2 for standard deviations). Statistical differences between blocks were assessed using Helmert contrasts (HC). * p < .05; ** p < .001; p < .01; ns denotes non-significant differences (*p* > .05), reflecting stabilization in performance beginning at Block 8.

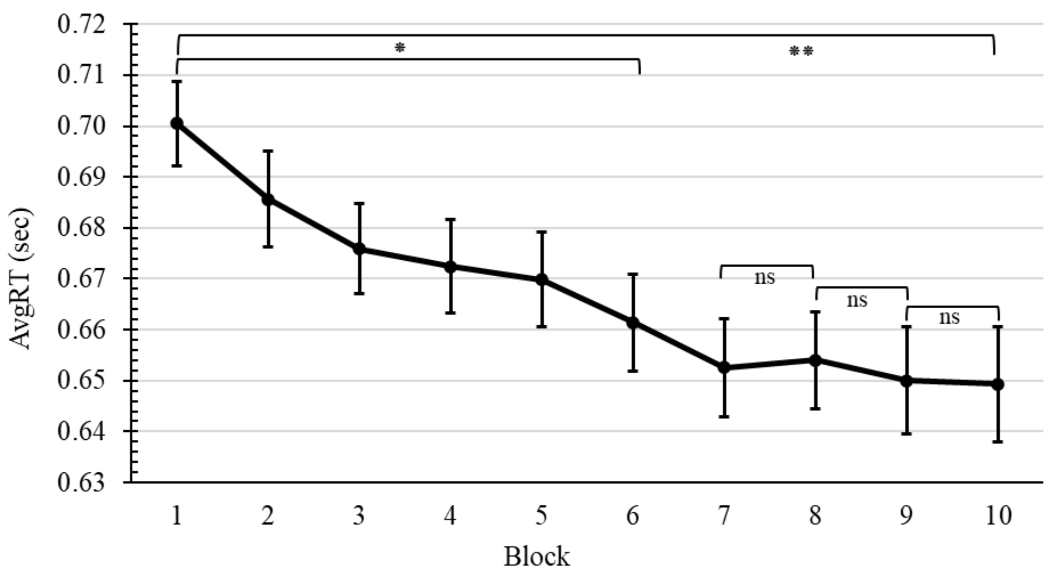

Figure 2. Changes in mean reaction time (AvgRT) during Dynavision D2 Mode B training blocks (3 tests per block). Error bars represent ± one standard error of the mean (SEM), reflecting the precision of the group mean across repeated testing blocks (see Table 2 for standard deviations). Statistical differences between blocks were assessed using Helmert contrasts (HC). * p < .05; ** p< .001; p < .01; ns denotes non-significant differences (p > .05), reflecting stabilization in performance beginning at Block 7; sec = seconds.

Misses decreased significantly across training blocks (Figure 3), with the largest reductions occurring during early testing blocks. Early improvements were associated with moderate effect sizes and a meaningful proportion of participants exceeding MDC thresholds. In contrast, changes observed during later blocks were smaller and less consistent, with many comparisons falling within measurement error thresholds (Supplementary Material – Appendix 3). This pattern suggests reduced variability in performance and stabilization of error rates during later stages of training.

TSC increased significantly across training blocks, although the magnitude of change was smaller relative to other variables. Improvements were observed primarily during early blocks, with fewer consistent differences across later blocks (Supplementary Material – Appendix 4), indicating less stable performance relative to other outcomes.

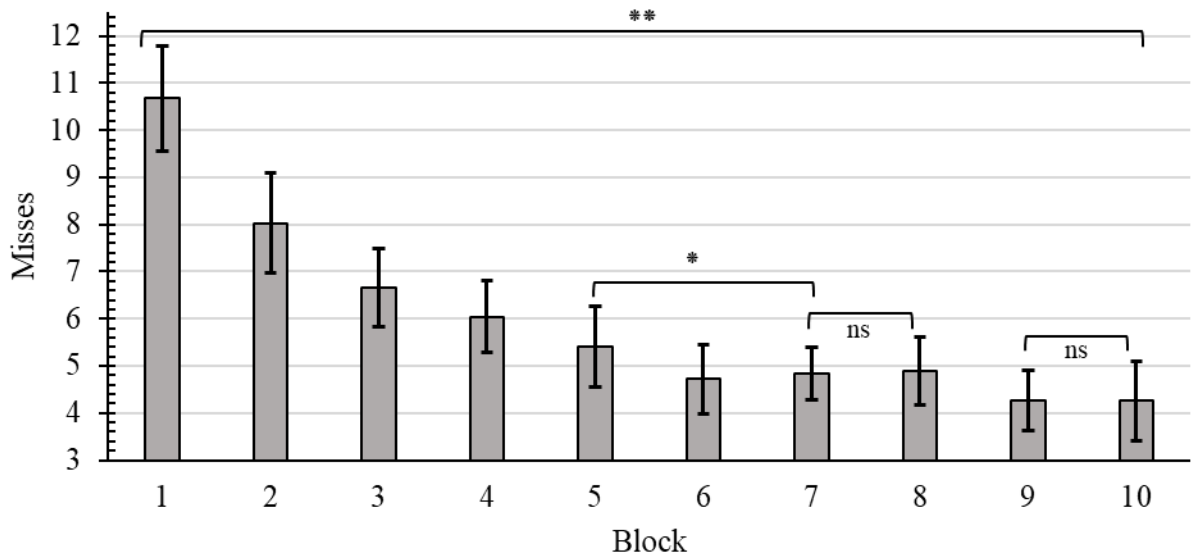

Figure 3. Changes in number of misses during Dynavision D2 Mode B training blocks (3 tests per block). Error bars represent ± one standard error of the mean (SEM), reflecting the precision of the group mean across repeated testing blocks (see Table 2 for standard deviations). Statistical differences between blocks were assessed using Helmert contrasts (HC). * p < .05; ** p< .001; p < .01; ns denotes non-significant differences (p > .05), reflecting stabilization in performance beginning at Block 7.

Table 2. Mean performance across Mode B training blocks for Total Hits, Average Reaction Time per Hit, Number of Misses, and TSC.

		Hits		**AvgRT**		**Misses**		**TSC**	
Block	Test #	Mean (SD)	HC	Mean (SD)	HC	Mean (SD)	HC	Mean (SD)	HC
1	1-3	70.9 (9.7)	*p* <.001	0.70 (.04)	*p* <.001	10.7 (5.7)	*p* <.001	9.1 (1.4)	*p* <.001
2	4-6	76.0 (11.1)	*p* <.001	0.69 (.05)	*p* <.001	8.0 (5.4)	*p* <.001	9.4 (1.2)	*p* <.001
3	7-9	79.2 (10.0)	*p* <.001	0.68 (.04)	*p* <.001	6.7 (4.3)	*p* <.001	9.7 (0.9)	*p* = .039
4	10-12	79.5 (10.9)	*p* <.001	0.67 (.05)	*p* <.001	6.0 (3.9)	*p* <.001	9.9 (1.0)	*p* = .332
5	13-15	81.8 (10.0)	*p* <.001	0.67 (.05)	*p* <.001	5.4 (4.4)	*p* = .019	9.7 (1.3)	*p* = .019
6	16-18	83.8 (9.7)	*p* = .020	0.66 (.05)	*p* = .014	4.7 (3.7)	*p* = .426	10.2 (0.6)	*p* = .293
7	19-21	84.5 (9.2)	*p* = .100	0.65 (.05)	*p* = .719	4.8 (2.9)	*p* = .014	9.8 (1.1)	*p* = .004
8	22-24	84.3 (10.6)	*p* = .005	0.65 (.05)	*p* = .364	4.9 (3.7)	*p* = .137	10.3 (0.7)	*p* = .164
9	25-27	86.4 (9.6)	*p* = .608	0.65 (.05)	*p* = .887	4.3 (3.3)	*p* = .516	10.2 (1.0)	*p* = .560
10	28-30	86.8 (12.0)	-	0.65 (.06)	-	4.3 (4.3)	-	10.1 (0.9)	-

Data presented as mean (standard deviation). Helmert contrast (HC) *p*-values are reported for each block comparison, reflecting the difference between the indicated block and subsequent blocks during the training phase. AvgRT = average reaction time per hit in seconds; TSC = number of correctly identified five-digit numbers on the T-scope challenge.

Mode B Retraining

The descriptive data for Hits, AvgRT, Misses, and TSC during the retraining period are presented in Table 3. No significant main effects of block, group, or group × block interactions were observed for Hits, AvgRT, Misses, or TSC ($p > .05$). Although minor numerical changes were observed across retraining blocks, these did not exceed MDC thresholds and were not consistent across participants, suggesting no meaningful decline in performance following detraining within the timeframes examined (Supplementary Material - Appendix 5-8).

Table 3. Mean performance across Mode B retraining blocks for Total Hits, Average Reaction Time per Hit, Number of Misses, and TSC.

Block	Test #	Hits	AvgRT	Misses	TSC
1	1-3	86.2 (10.5)	0.65 (.05)	3.9 (3.2)	10.2 (0.8)
2	4-6	87.2 (11.7)	0.64 (.06)	4.1 (3.5)	10.1 (1.0)
3	7-9	87.6 (11.6)	0.64 (.06)	3.7 (3.4)	10.3 (0.8)
4	10-12	87.5 (10.6)	0.64 (.06)	3.9 (2.7)	10.3 (0.8)
5	13-15	88.6 (11.9)	0.64 (.06)	3.7 (3.3)	10.3 (0.8)
6	16-18	88.8 (11.4)	0.64 (.06)	3.2 (2.9)	10.0 (1.1)
7	19-21	88.5 (12.5)	0.64 (.06)	3.5 (3.3)	10.4 (0.5)
8	22-24	88.9 (12.0)	0.63 (.06)	3.5 (3.4)	10.2 (0.8)
9	25-27	89.9 (12.2)	0.63 (.06)	3.2 (3.2)	10.4 (0.5)
10	28-30	90.2 (13.5)	0.63 (.07)	3.3 (4.2)	10.3 (0.8)

Data presented as mean (standard deviation). AvgRT = average reaction time per hit in seconds; TSC = number of correctly identified five-digit numbers on the T-scope challenge.

Reliability and Measurement Error

Reliability analyses were conducted between blocks 9 and 10, which were selected to represent a period of stabilized performance based on minimal changes across variables (i.e., Helmert contrasts, effect sizes, and MDC thresholds), consistent with the criteria used to define stabilization in the present study. Hits demonstrated excellent test–retest reliability (ICC = .968; 95% CI: .928 to .985, SEM = 1.940, MDC = 5.377), as did AvgRT (ICC = .958; 95% CI: .906 to .981, SEM = .011, MDC = .032) and Misses (ICC = .902; 95% CI: .780 to .956, SEM = 1.200, MDC = 3.326). TSC demonstrated good reliability (ICC = .785; 95% CI: .519 to .904, SEM = .423, MDC = 1.172). These findings indicate that changes observed during later training blocks were generally within measurement error thresholds, further supporting the interpretation of performance stabilization.

Discussion

The results of this study indicate that there are significant improvements in total hits, AvgRT per hit, number of misses, and the number of TSC correct during the first 24 Mode B assessments on the Dynavision D2 system. Collectively, while individual performance variables stabilized at different time points, no further improvements were observed across any variables from block 9 onward, indicating that a stable baseline across all outcomes was achieved by this stage of testing. Furthermore, the magnitude of the observed effect sizes during earlier blocks (e.g., η^2_p values ranging from moderate to very large, with Hits exceeding 0.60) also supports the presence of substantial learning effects during the initial phases of testing. Additionally, effect sizes for changes in all performance variables between blocks 9 and later were interpreted as trivial while the mean difference in scores did not exceed the SEM. During the retraining period, no significant differences in scores for any variable were observed between blocks, indicating there were no additional improvements in performance and no effect of detraining periods lasting up to two weeks in the present study. Based on the current findings, a stable performance baseline encompassing all test metrics appears to be reached following 24 Mode B tests. Given that participants completed one session per day consisting of two Mode B tests, this corresponds to approximately 12 days of testing under the present protocol. Performance changes beyond this point that exceed the MDC may represent true deviations rather than normal test variability. These findings support our initial hypothesis that Mode B performance would improve during early testing before stabilizing. Additionally, consistent with our second hypothesis, detraining periods of up to two weeks did not meaningfully impact performance upon retesting.

In contrast to previous work, the present study demonstrates that substantially more repeated exposures are required to achieve stable Mode B performance when testing is performed without extensive pre-testing familiarization procedures or concurrent Dynavision assessments. Early investigations by Klavora et al[20,21] reported improvements across initial trials with relatively few exposures required to achieve stable performance. However, these studies incorporated practice trials prior to formal testing, with progression dependent on achieving performance stability, which may have accelerated early learning and reduced the apparent duration of the familiarization period. Additionally, repeated practice trials were administered across testing sessions, further increasing total exposure to the task. Similarly, Wells et al[19] reported improved reliability across repeated Mode B trials; however, participants also completed multiple other Dynavision assessments in conjunction with Mode B testing. This additional exposure may have influenced performance through generalized learning or task familiarity, thereby contributing to earlier stabilization. In contrast, the present study isolated the Mode B assessment and minimized additional practice exposures, allowing for a more direct characterization of training effects across repeated testing. Only two additional trials were permitted during V2, under the direction of the study team, to correct posture or task execution. Collectively, these methodological differences suggest that prior estimates of Mode B reliability may underestimate the number of exposures required to establish a stable baseline when the assessment is implemented in isolation.

To our knowledge, no other studies have examined Mode B performance over an extended period of time; however, a previous investigation has characterized the training and detraining effects associated with the Mode A test.[22] Similarly, the Mode A assessment was found to require significant repeated testing before a true baseline score can be achieved. However, unlike the present investigation, the Mode A test requires only 18 assessments to reach a true baseline while the Mode B test requires at least two familiarization trials, and 24 additional tests before all performance metrics stabilized. The difference in required familiarization lengths may be due in part to the increased complexity of the Mode B test, which requires individuals to balance a cognitive challenge with a visuomotor reaction stimulus. Accordingly, the familiarization for both D2 assessments (Mode A and Mode B) should be considered independently when being considered for use in sports rehabilitation, repeated testing situations, or during acute intervention investigations at this time. Future research should examine whether familiarization with one assessment, may impact or reduce the number of trials needed for adequate familiarization with alternative D2 assessments.

The finding that individual Mode B variables stabilized at different time points likely reflects the distinct performance demands captured by each outcome. Hits and AvgRT primarily reflect visuomotor speed and successful motor execution, which may improve more rapidly as participants become familiar with the task and response requirements. In contrast, misses may remain more variable, as they reflect unsuccessful responses that are more sensitive to lapses in attention, timing, or motor accuracy. Similarly, TSC incorporates an additional cognitive and verbal component beyond target striking, which may contribute to greater variability and a longer familiarization period, as well as slightly lower reliability relative to other performance outcomes. As such, improvements observed in TSC across repeated testing likely reflect task-specific learning and increased familiarity with the dual-task demands of the assessment, rather than generalized enhancements in cognitive function. Collectively, these findings suggest that Mode B performance metrics do not adapt uniformly during repeated testing.

From an applied perspective, these findings have important implications for practitioners using the Dynavision D2 Mode B assessment in sport, rehabilitation, and clinical settings. In athletic populations, insufficient familiarization may lead to misinterpretation of performance changes during return-to-play or training interventions. Similarly, in rehabilitation or aging populations, where reaction time and dual-task performance are often monitored, failure to establish a stable baseline may result in inaccurate assessments of recovery or functional decline. Importantly, these findings also have implications for researchers examining the effects of interventions such as supplementation or training protocols, as improvements observed without adequate familiarization may be incorrectly attributed to the intervention rather than continued task learning. Thus, ensuring adequate familiarization is critical for accurate interpretation of Mode B performance across a range of applied settings. In this context, potential sources of variability such as insufficient familiarization, fluctuations in attention or effort, and the added complexity of the dual-task component should be considered, and standardized testing procedures implemented to minimize measurement error.

Moreover, this study is not without limitations. The present investigation only considered detraining periods of one or two weeks. Consequently, we are unable to fully characterize how additional durations of inactivity might impact performance on the Mode B test. Researchers or practitioners implementing longer periods of absence from repeated testing should consider this limitation and interpret deviations in score with discretion. Further, the implementation of "re-familiarization" trials following longer detraining periods, may be a feasible solution to overcome this limitation. Additionally, this investigation was conducted exclusively in recreationally active, young adults, which may limit the applicability of these findings to other populations, including elite athletes, sedentary individuals, or older adults. No stratified stratified analysis by sex was performed despite the unequal distribution of male and female participants, which may influence variability in performance outcomes between sexes and limit the interpretation of potential sex-specific responses. Previous work has demonstrated differences in the variability of reaction time tasks between older and younger adults.[9] Additionally, athletes tend to exhibit faster reaction times, fewer errors, and greater consistency in reaction time tasks compared to non-athletes.[29] Therefore, the present findings in recreationally active adults may not be fully transferrable to other populations. Accordingly, additional research should examine whether athletes, older adults, and sedentary individuals, require disparate familiarization periods to overcome initial training effects.

In conclusion, substantial training effects and variation in performance outcomes between distinct Mode B tests were observed for all participants during this investigation. These findings highlight the importance of accounting for familiarization and training effects when implementing the Dynavision D2 Mode B test in both applied and research settings. Importantly, these results indicate that insufficient familiarization may lead to misinterpretation of Mode B performance changes as true treatment effects, when they may instead reflect learning. To ensure the reliable use of this assessment, the following guidelines are offered: (1) Prior to utilizing the Mode B test as a clinical or research outcome measure, a familiarization period followed by 24 test trials should be completed to overcome the initial learning effect and establish a stable baseline score across all performance variables, (2) at each additional testing timepoint three distinct Mode B tests should be administered, with the average score used for interpretation, and (3) testing sessions may be separated by upwards of two weeks without negatively impacting performance under the conditions examined.

Acknowledgements

The authors would like to thank the study participants for their time and commitment to this protocol.

References

1. Delmas S, Casamento-Moran A, Park SH, Yacoubi B, Christou EA. Motor planning perturbation: muscle activation and reaction time. *J Neurophysiol*. 2018;120(4):2059-2065. https://doi.org/10.1152/jn.00323.2018

2. Spiteri T, Cochrane JL, Nimphius S. The evaluation of a new lower-body reaction time test. *J Strength Cond Res*. 2013;27(1):174-180. https://doi.org/10.1519/JSC.0b013e318250381f

3. Wilkerson GB, Simpson KA, Clark RA. Assessment and training of visuomotor reaction time for football injury prevention. *J Sport Rehabil*. 2017;26(1):26-34. doi:10.1123/jsr.2015-0068

4. Tønnessen E, Haugen T, Shalfawi SA. Reaction time aspects of elite sprinters in athletic world championships. *J Strength Cond Res*. 2013;27(4):885-892. https://doi.org/10.1519/JSC.0b013e31826520c3

5. Feldhacker DR, Molitor L, Athmann A, et al. Efficacy of high-performance vision training on improving the reaction time of collegiate softball athletes: a randomized trial. *J Sports Med Allied Health Sci*. 2019;4(3):6. https://doi.org/10.25035/jsmahs.04.03.06

6. Balkó Š, Rous M, Balkó I, Hnízdil J, Borysiuk Z. Influence of a 9-week training intervention on the reaction time of fencers aged 15 to 18 years. *Phys Act Rev*. 2017;5:146-154. https://doi.org/10.16926/par.2017.05.19

7. Ando S, Kida N, Oda S. Central and peripheral visual reaction time of soccer players and nonathletes. *Percept Mot Skills*. 2001;92(3 Pt 1):786-794. https://doi.org/10.2466/pms.2001.92.3.786

8. Hoffman JR, Ratamess NA, Gonzalez A, et al. The effects of acute and prolonged CRAM supplementation on reaction time and subjective measures of focus and alertness in healthy college students. *J Int Soc Sports Nutr*. 2010;7:39. https://doi.org/10.1186/1550-2783-7-39

9. Hultsch DF, MacDonald SW, Dixon RA. Variability in reaction time performance of younger and older adults. *J Gerontol B Psychol Sci Soc Sci*. 2002;57(2):P101-P115. https://doi.org/10.1093/geronb/57.2.P101

10. Li XY, Wu XY, Fu C, Shen XF, Yang CB, Wu YH. Effects of acute exposure to mild or moderate hypoxia on human psychomotor performance and visual-reaction time. *Space Med Med Eng (Beijing)*. 2000;13(4):235-239.

11. Mori S, Ohtani Y, Imanaka K. Reaction times and anticipatory skills of karate athletes. *Hum Mov Sci*. 2002;21(2):213-230. https://doi.org/10.1016/s0167-9457(02)00103-3

12. Harper B, Shiraishi M, Soangra R. A portable and reliable tool for on-site physical reaction time (RT) measurement. *Invent Discl*. 2023;3:100013. https://doi.org/10.1016/j.inv.2023.100013

13. Eckner JT, Kutcher JS, Richardson JK. Pilot evaluation of a novel clinical test of reaction time in national collegiate athletic association division I football players. *J Athl Train*. 2010;45(4):327-332. https://doi.org/10.4085/1062-6050-45.4.327

14. Hoffman JR, Markus I, Dubnov-Raz G, Gepner Y. Ergogenic effects of 8 days of sceletium tortuosum supplementation on mood, visual tracking, and reaction in recreationally trained men and women. *J Strength Cond Res*. 2020;34(9):2476-2481. https://doi.org/10.1519/JSC.0000000000003693

15. Moon JM, Dufner TJ, Wells AJ. Evaluating the effects of PeakATP® supplementation on visuomotor reaction time and cognitive function following high-intensity sprint exercise. *Front Nutr*. 2023;10:1237678. https://doi.org/10.3389/fnut.2023.1237678

16. Church DD, Hoffman JR, LaMonica MB, et al. The effect of an acute ingestion of Turkish coffee on reaction time and time trial performance. *J Int Soc Sports Nutr*. 2015;12:37. https://doi.org/10.1186/s12970-015-0098-3

17. Gonzalez AM, Hoffman JR, Wells AJ, et al. Effects of time-release caffeine containing supplement on metabolic rate, glycerol concentration and performance. *J Sports Sci Med*. 2015;14(2):322-332.

18. Pruna GJ, Hoffman JR, McCormack WP, et al. Effect of acute L-Alanyl-L-Glutamine and electrolyte ingestion on cognitive function and reaction time following endurance exercise. *Eur J Sport Sci*. 2016;16(1):72-79. https://doi.org/10.1080/17461391.2014.969325

19. Wells AJ, Hoffman JR, Beyer KS, et al. Reliability of the Dynavision D2 for assessing reaction time performance. *J Sports Sci Med*. 2014;13(1):145-150.

20. Klavora P, Gaskovski P, Forsyth RD. Test-retest reliability of three Dynavision tasks. *Percept Mot Skills*. 1995;80(2):607-610. https://doi.org/10.2466/pms.1995.80.2.607

21. Klavora P, Gaskovski P, Forsyth R. Test-retest reliability of the Dynavision apparatus. *Percept Mot Skills*. 1994;79(1 Pt 2):448-450. https://doi.org/10.2466/pms.1994.79.1.448

22. Wells AJ, Johnson BDI. Test-retest reliability, training, and detraining effects associated with the Dynavision D2 Mode A visuomotor reaction time test. *J Sport Rehabil*. 2022;31(2):253-261. https://doi.org/10.1123/jsr.2020-0550

23. Ong NCH. The use of Dynavision in sport and exercise research: a review. *Int J Sport Exerc Psychol*. 2020;18(5):561-580. https://doi.org/10.1080/1612197X.2018.1549582

24. Navalta JW, Stone WJ, Lyons TS. Ethical issues relating to scientific discovery in exercise science. *Int J Exerc Sci*. 2020;12(1):1-8. https://doi.org/10.70252/EYCD6235

25. Blanca MJ, Alarcón R, Arnau J, Bono R, Bendayan R. Non-normal data: Is ANOVA still a valid option? *Psicothema*. 2017;29(4):552-557. https://doi.org/10.7334/psicothema2016.383

26. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Routledge; 2013.

27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. https://doi.org/10.1016/j.jcm.2016.02.012

28. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240. https://doi.org/10.1519/15184.1

29. Kuan YM, Zuhairi NA, Manan FA, Knight VF, Omar R. Visual reaction time and visual anticipation time between athletes and non-athletes. *Malays J Public Health Med*. 2018;1(1):135-141.

Supplementary Materials

Appendix 1. Between-block differences in Hits during training. Abbreviations: CI, confidence interval; ES, Hedges' g (corrected for small sample size); SEM, standard error of measurement; MDC, minimal detectable change. Note: Values are presented as mean difference (SD), with negative values indicating higher performance in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Helmert contrast results are presented separately (see Table 2). Effect sizes (ES) are reported as Hedges' g (corrected for small sample size) and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates whether the group mean difference exceeded measurement error (±SEM), reflecting group-level meaningful change. "% Exceeding MDC" represents the proportion of participants whose individual change exceeded ±MDC, reflecting individual-level meaningful change. SEM = ±1.940; MDC = ±5.377.

Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC
B1 vs B2	-5.077 (5.2)	<.001	-7.164 to -2.990	0.648 (Moderate)	Yes	46.2
B1 vs B3	-8.244 (6.6)	<.001	-10.904 to -5.583	1.117 (Moderate)	Yes	65.4
B1 vs B4	-8.603 (7.9)	<.001	-11.814 to -5.391	1.110 (Moderate)	Yes	76.9
B1 vs B5	-10.846 (6.9)	<.001	-13.653 to -8.039	1.468 (Large)	Yes	80.8
B1 vs B6	-12.846 (7.0)	<.001	-15.674 to -10.019	1.762 (Large)	Yes	92.3
B1 vs B7	-13.551 (6.5)	<.001	-16.188 to -10.914	1.907 (Large)	Yes	92.3
B1 vs B8	-13.385 (7.8)	<.001	-16.552 to -10.217	1.753 (Large)	Yes	84.6
B1 vs B9	-15.526 (7.0)	<.001	-18.356 to -12.696	2.148 (Very Large)	Yes	100.0
B1 vs B10	-15.923 (8.0)	<.001	-19.17 to -12.676	1.942 (Large)	Yes	96.2
B2 vs B3	-3.167 (4.9)	.003	-5.129 to -1.205	0.399 (Small)	Yes	34.6
B2 vs B4	-3.526 (6.1)	.007	-5.984 to -1.068	0.426 (Small)	Yes	46.2
B2 vs B5	-5.769 (5.4)	<.001	-7.932 to -3.607	0.727 (Moderate)	Yes	57.7
B2 vs B6	-7.769 (6.5)	<.001	-10.408 to -5.130	0.990 (Moderate)	Yes	65.4
B2 vs B7	-8.474 (5.4)	<.001	-10.647 to -6.302	1.104 (Moderate)	Yes	65.4
B2 vs B8	-8.308 (6.4)	<.001	-10.882 to -5.733	1.017 (Moderate)	Yes	80.8
B2 vs B9	-10.449 (6.1)	<.001	-12.923 to -7.974	1.342 (Large)	Yes	76.9
B2 vs B10	-10.846 (6.4)	<.001	-13.431 to -8.261	1.247 (Large)	Yes	80.8
B3 vs B4	-0.359 (5.8)	.753	-2.683 to 1.965	0.046 (Trivial)	No	38.5
B3 vs B5	-2.603 (4.7)	.009	-4.489 to -0.716	0.348 (Small)	Yes	30.8
B3 vs B6	-4.603 (5.5)	<.001	-6.819 to -2.386	0.623 (Moderate)	Yes	38.5
B3 vs B7	-5.308 (4.9)	<.001	-7.272 to -3.343	0.736 (Moderate)	Yes	42.3
B3 vs B8	-5.141 (5.7)	<.001	-7.462 to -2.820	0.665 (Moderate)	Yes	46.2
B3 vs B9	-7.282 (6.4)	<.001	-9.858 to -4.706	0.994 (Moderate)	Yes	61.5
B3 vs B10	-7.679 (6.6)	<.001	-10.326 to -5.033	0.926 (Moderate)	Yes	76.9
B4 vs B5	-2.244 (5.5)	.047	-4.452 to -0.035	0.286 (Small)	Yes	11.5
B4 vs B6	-4.244 (6.0)	.001	-6.683 to -1.804	0.547 (Small)	Yes	38.5
B4 vs B7	-4.949 (4.8)	<.001	-6.904 to -2.993	0.652 (Moderate)	Yes	26.9
B4 vs B8	-4.782 (5.6)	<.001	-7.029 to -2.535	0.592 (Small)	Yes	50.0
B4 vs B9	-6.923 (6.1)	<.001	-9.397 to -4.449	0.899 (Moderate)	Yes	65.4
B4 vs B10	-7.321 (6.6)	<.001	-9.992 to -4.649	0.849 (Moderate)	Yes	73.1
B5 vs B6	-2.000 (3.4)	.005	-3.352 to -0.648	0.270 (Small)	Yes	15.4
B5 vs B7	-2.705 (4.4)	.004	-4.464 to -0.947	0.375 (Small)	Yes	26.9
B5 vs B8	-2.538 (4.7)	.010	-4.428 to -0.649	0.328 (Small)	Yes	26.9
B5 vs B9	-4.679 (4.7)	<.001	-6.565 to -2.794	0.638 (Moderate)	Yes	38.5
B5 vs B10	-5.077 (4.7)	<.001	-6.981 to -3.173	0.612 (Moderate)	Yes	50.0
B6 vs B7	-0.705 (4.8)	.463	-2.653 to 1.243	0.099 (Trivial)	No	15.4
B6 vs B8	-0.538 (4.4)	.537	-2.310 to 1.233	0.070 (Trivial)	No	11.5
B6 vs B9	-2.679 (4.2)	.004	-4.394 to -0.965	0.370 (Small)	Yes	15.4
B6 vs B10	-3.077 (4.6)	.002	-4.948 to -1.205	0.375 (Small)	Yes	38.5

B7 vs B8	0.167 (4.3)	.846	-1.584 to 1.917	-0.022 (Trivial)	No	19.2
B7 vs B9	-1.974 (4.5)	.035	-3.798 to -0.151	0.280 (Small)	Yes	26.9
B7 vs B10	-2.372 (5.5)	.038	-4.605 to -0.139	0.294 (Small)	Yes	26.9
B8 vs B9	-2.141 (4.6)	.027	-4.016 to -0.266	0.282 (Small)	Yes	19.2
B8 vs B10	-2.538 (4.1)	.004	-4.186 to -0.891	0.298 (Small)	Yes	15.4
B9 vs B10	-0.397 (3.9)	.608	-1.973 to 1.178	0.049 (Trivial)	No	7.7

Appendix 2. Between-block differences in average reaction time (AvgRT) during training. Abbreviations: CI, confidence interval; ES, Hedges' g; AvgRT, average reaction time; SEM, standard error of measurement; MDC, minimal detectable change. Note: Values are presented as mean difference (SD), with positive values indicating faster reaction time (improvement) in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Helmert contrast results are presented separately (see Table 2). Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" reflects whether the group mean difference exceeded measurement error (±SEM), while "% Exceeding MDC" reflects the proportion of participants demonstrating meaningful individual change. SEM = ±0.011; MDC = ±0.032.

Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC
B1 vs B2	0.015 (.025)	.004	0.005 to 0.025	-0.447 (Small)	Yes	38.5
B1 vs B3	0.025 (.025)	<.001	0.015 to 0.035	-0.790 (Moderate)	Yes	50.0
B1 vs B4	0.028 (.025)	<.001	0.017 to 0.039	-0.885 (Moderate)	Yes	50.0
B1 vs B5	0.031 (.025)	<.001	0.020 to 0.042	-0.926 (Moderate)	Yes	38.5
B1 vs B6	0.039 (.031)	<.001	0.027 to 0.051	-1.160 (Moderate)	Yes	50.0
B1 vs B7	0.048 (.031)	<.001	0.035 to 0.061	-1.422 (Large)	Yes	69.2
B1 vs B8	0.047 (.036)	<.001	0.033 to 0.060	-1.404 (Large)	Yes	61.5
B1 vs B9	0.051 (.036)	<.001	0.036 to 0.066	-1.417 (Large)	Yes	73.1
B1 vs B10	0.051 (.046)	<.001	0.034 to 0.069	-1.329 (Large)	Yes	69.2
B2 vs B3	0.010 (.015)	.003	0.004 to 0.016	-0.300 (Small)	No	11.5
B2 vs B4	0.013 (.020)	.002	0.005 to 0.021	-0.395 (Small)	Yes	11.5
B2 vs B5	0.016 (.025)	.005	0.005 to 0.026	-0.442 (Small)	Yes	26.9
B2 vs B6	0.024 (.025)	<.001	0.013 to 0.035	-0.674 (Moderate)	Yes	50.0
B2 vs B7	0.033 (.025)	<.001	0.022 to 0.044	-0.918 (Moderate)	Yes	46.2
B2 vs B8	0.032 (.031)	<.001	0.020 to 0.044	-0.900 (Moderate)	Yes	42.3
B2 vs B9	0.036 (.036)	<.001	0.022 to 0.049	-0.939 (Moderate)	Yes	50.0
B2 vs B10	0.036 (.041)	<.001	0.020 to 0.052	-0.884 (Moderate)	Yes	73.1
B3 vs B4	0.003 (.010)	.164	-0.002 to 0.008	-0.103 (Trivial)	No	0.0
B3 vs B5	0.006 (.020)	.128	-0.002 to 0.014	-0.157 (Trivial)	No	3.8
B3 vs B6	0.014 (.020)	.001	0.006 to 0.023	-0.404 (Small)	Yes	23.1
B3 vs B7	0.023 (.020)	<.001	0.015 to 0.031	-0.656 (Moderate)	Yes	30.8
B3 vs B8	0.022 (.020)	<.001	0.013 to 0.031	-0.636 (Moderate)	Yes	26.9
B3 vs B9	0.026 (.025)	<.001	0.015 to 0.036	-0.688 (Moderate)	Yes	34.6
B3 vs B10	0.027 (.036)	<.001	0.013 to 0.040	-0.643 (Moderate)	Yes	46.2
B4 vs B5	0.003 (.015)	.436	-0.004 to 0.009	-0.055 (Trivial)	No	7.7
B4 vs B6	0.011 (.020)	.006	0.003 to 0.019	-0.302 (Small)	No	7.7
B4 vs B7	0.020 (.020)	<.001	0.012 to 0.028	-0.551 (Small)	Yes	19.2
B4 vs B8	0.018 (.020)	<.001	0.011 to 0.026	-0.532 (Small)	Yes	26.9
B4 vs B9	0.022 (.025)	<.001	0.012 to 0.032	-0.588 (Small)	Yes	42.3
B4 vs B10	0.023 (.036)	.002	0.010 to 0.037	-0.549 (Small)	Yes	50.0
B5 vs B6	0.008 (.015)	.007	0.003 to 0.014	-0.244 (Small)	No	11.5
B5 vs B7	0.017 (.020)	<.001	0.009 to 0.026	-0.489 (Small)	Yes	19.2
B5 vs B8	0.016 (.020)	<.001	0.008 to 0.024	-0.469 (Small)	Yes	19.2
B5 vs B9	0.020 (.025)	<.001	0.009 to 0.030	-0.529 (Small)	Yes	30.8
B5 vs B10	0.021 (.036)	.004	0.007 to 0.034	-0.493 (Small)	Yes	38.5

B6 vs B7	0.009 (.020)	.055	0.000 to 0.018	-0.239 (Small)	No	15.4
B6 vs B8	0.007 (.020)	.062	0.000 to 0.015	-0.219 (Small)	No	15.4
B6 vs B9	0.011 (.025)	.026	0.001 to 0.021	-0.289 (Small)	No	23.1
B6 vs B10	0.012 (.031)	.056	0.000 to 0.024	-0.266 (Small)	Yes	34.6
B7 vs B8	-0.001 (.015)	.669	-0.008 to 0.005	0.021 (Trivial)	No	7.7
B7 vs B9	0.003 (.025)	.611	-0.008 to 0.013	-0.060 (Trivial)	No	23.1
B7 vs B10	0.003 (.031)	.602	-0.009 to 0.016	-0.047 (Trivial)	No	26.9
B8 vs B9	0.004 (.025)	.403	-0.006 to 0.014	-0.080 (Trivial)	No	26.9
B8 vs B10	0.005 (.031)	.417	-0.007 to 0.016	-0.067 (Trivial)	No	19.2
B9 vs B10	0.001 (.020)	.887	-0.009 to 0.010	0.009 (Trivial)	No	7.7

Appendix 3: Between-block differences in misses during training. Abbreviations: CI, confidence interval; ES, Hedges' g; SEM, standard error of measurement; MDC, minimal detectable change. Note: Values are presented as mean difference (SD), with negative values indicating fewer misses (improvement) in later blocks. Data are presented as raw values for interpretability. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Helmert contrast results are presented separately (see Table 2). Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates group-level change beyond measurement error, whereas "% Exceeding MDC" reflects the proportion of participants exceeding individual-level meaningful change. SEM = ±1.200; MDC = ±3.326.

Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC
B1 vs B2	2.654 (2.1)	<.001	1.790 to 3.518	-0.639 (Moderate)	Yes	38.5
B1 vs B3	4.013 (3.1)	<.001	2.746 to 5.280	-1.064 (Moderate)	Yes	53.8
B1 vs B4	4.641 (3.1)	<.001	3.406 to 5.876	-1.270 (Large)	Yes	65.4
B1 vs B5	5.282 (3.3)	<.001	3.955 to 6.609	-1.389 (Large)	Yes	73.1
B1 vs B6	5.962 (3.8)	<.001	4.428 to 7.495	-1.649 (Large)	Yes	76.9
B1 vs B7	5.846 (3.9)	<.001	4.287 to 7.406	-1.728 (Large)	Yes	76.9
B1 vs B8	5.782 (3.7)	<.001	4.289 to 7.275	-1.611 (Large)	Yes	76.9
B1 vs B9	6.410 (3.8)	<.001	4.893 to 7.928	-1.836 (Large)	Yes	76.9
B1 vs B10	6.423 (3.9)	<.001	4.84 to 8.006	-1.697 (Large)	Yes	84.6
B2 vs B3	1.359 (2.3)	.006	0.431 to 2.287	-0.374 (Small)	Yes	26.9
B2 vs B4	1.987 (2.3)	<.001	1.070 to 2.905	-0.565 (Small)	Yes	26.9
B2 vs B5	2.628 (2.4)	<.001	1.657 to 3.599	-0.716 (Moderate)	Yes	38.5
B2 vs B6	3.308 (3.1)	<.001	2.046 to 4.569	-0.952 (Moderate)	Yes	46.2
B2 vs B7	3.192 (3.2)	<.001	1.895 to 4.489	-0.987 (Moderate)	Yes	38.5
B2 vs B8	3.128 (3.0)	<.001	1.922 to 4.335	-0.907 (Moderate)	Yes	50.0
B2 vs B9	3.756 (3.3	<.001	2.440 to 5.073	-1.122 (Moderate)	Yes	46.2
B2 vs B10	3.769 (3.3)	<.001	2.450 to 5.088	-1.032 (Moderate)	Yes	46.2
B3 vs B4	0.628 (1.8)	.087	-0.099 to 1.355	-0.206 (Small)	No	7.7
B3 vs B5	1.269 (2.0)	.004	0.452 to 2.087	-0.393 (Small)	Yes	26.9
B3 vs B6	1.949 (2.2)	<.001	1.041 to 2.857	-0.649 (Moderate)	Yes	30.8
B3 vs B7	1.833 (2.6)	.001	0.785 to 2.881	-0.674 (Moderate)	Yes	26.9
B3 vs B8	1.769 (2.3)	<.001	0.827 to 2.711	-0.595 (Small)	Yes	19.2
B3 vs B9	2.397 (2.6)	<.001	1.341 to 3.453	-0.840 (Moderate)	Yes	38.5
B3 vs B10	2.410 (3.2)	<.001	1.128 to 3.692	-0.752 (Moderate)	Yes	34.6
B4 vs B5	0.641 (1.7)	.067	-0.05 to 1.332	-0.207 (Small)	No	3.8
B4 vs B6	1.321 (1.9)	.002	0.547 to 2.094	-0.462 (Small)	Yes	19.2
B4 vs B7	1.205 (2.0)	.004	0.412 to 1.999	-0.471 (Small)	Yes	15.4
B4 vs B8	1.141 (1.8)	.004	0.399 to 1.884	-0.404 (Small)	No	15.4
B4 vs B9	1.769 (2.2)	<.001	0.865 to 2.673	-0.655 (Moderate)	Yes	19.2
B4 vs B10	1.782 (2.7)	.003	0.671 to 2.893	-0.581 (Small)	Yes	30.8
B5 vs B6	0.679 (2.2)	.122	-0.194 to 1.553	-0.223 (Small)	No	19.2

B5 vs B7	0.564 (2.8)	.306	-0.547 to 1.675	-0.204 (Small)	No	19.2
B5 vs B8	0.500 (2.4)	.295	-0.462 to 1.462	-0.166 (Trivial)	No	19.2
B5 vs B9	1.128 (2.7)	.040	0.056 to 2.200	-0.389 (Small)	No	23.1
B5 vs B10	1.141 (2.9)	.060	-0.049 to 2.331	-0.352 (Small)	No	23.1
B6 vs B7	-0.115 (2.5)	.819	-1.142 to 0.911	0.046 (Trivial)	No	11.5
B6 vs B8	-0.179 (2.2)	.678	-1.058 to 0.699	0.065 (Trivial)	No	3.8
B6 vs B9	0.449 (2.2)	.316	-0.455 to 1.352	-0.170 (Trivial)	No	19.2
B6 vs B10	0.462 (2.7)	.390	-0.626 to 1.549	-0.153 (Trivial)	No	15.4
B7 vs B8	-0.064 (2.0)	.872	-0.876 to 0.747	0.026 (Trivial)	No	7.7
B7 vs B9	0.564 (2.0)	.163	-0.244 to 1.373	-0.243 (Small)	No	15.4
B7 vs B10	0.577 (2.5)	.260	-0.454 to 1.608	-0.210 (Small)	No	15.4
B8 vs B9	0.628 (2.2)	.157	-0.259 to 1.515	-0.241 (Small)	No	11.5
B8 vs B10	0.641 (2.1)	.127	-0.196 to 1.478	-0.214 (Small)	No	11.5
B9 vs B10	0.013 (2.3)	.978	-0.929 to 0.954	-0.004 (Trivial)	No	7.7

Appendix 4: Between-block differences in TSC responses during training. Abbreviations: CI, confidence interval; ES, Hedges' g; SEM, standard error of measurement; MDC, minimal detectable change.
Note: Values are presented as mean difference (SD), with negative values indicating improved performance in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Helmert contrast results are presented separately (see Table 2). Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates whether the group mean difference exceeded measurement error. "% Exceeding MDC" reflects the proportion of participants demonstrating meaningful individual change. SEM = ±0.423; MDC = ±1.172.

Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC
B1 vs B2	-0.256 (0.9)	.146	-0.608 to 0.095	0.260 (Small)	No	15.4
B1 vs B3	-0.615 (0.9)	.002	-0.984 to -0.247	0.681 (Moderate)	Yes	19.2
B1 vs B4	-0.833 (1.0)	<.001	-1.256 to -0.411	0.908 (Moderate)	Yes	30.8
B1 vs B5	-0.564 (1.2)	.025	-1.005 to -0.078	0.563 (Small)	Yes	30.8
B1 vs B6	-1.141 (1.2)	<.001	-1.626 to -0.656	1.406 (Large)	Yes	38.5
B1 vs B7	-0.705 (1.0)	.002	-1.119 to -0.291	0.754 (Moderate)	Yes	26.9
B1 vs B8	-1.179 (1.1)	<.001	-1.621 to -0.738	1.391 (Large)	Yes	38.5
B1 vs B9	-1.064 (0.9)	<.001	-1.423 to -0.705	1.168 (Moderate)	Yes	30.8
B1 vs B10	-0.974 (1.2)	<.001	-1.465 to -0.484	1.112 (Moderate)	Yes	42.3
B2 vs B3	-0.359 (0.8)	.035	-0.691 to -0.027	0.443 (Small)	No	19.2
B2 vs B4	-0.577 (0.9)	.003	-0.945 to -0.209	0.698 (Moderate)	Yes	30.8
B2 vs B5	-0.308 (1.1)	.160	-0.745 to 0.130	0.335 (Small)	No	30.8
B2 vs B6	-0.885 (1.1)	<.001	-1.333 to -0.436	1.253 (Large)	Yes	34.6
B2 vs B7	-0.449 (0.9)	.013	-0.796 to -0.102	0.531 (Small)	Yes	19.2
B2 vs B8	-0.923 (1.0)	<.001	-1.327 to -0.520	1.235 (Large)	Yes	26.9
B2 vs B9	-0.808 (0.8)	<.001	-1.131 to -0.484	0.986 (Moderate)	Yes	34.6
B2 vs B10	-0.718 (0.8)	<.001	-1.036 to -0.400	0.921 (Moderate)	Yes	30.8
B3 vs B4	-0.218 (0.9)	.205	-0.563 to 0.127	0.300 (Small)	No	19.2
B3 vs B5	0.051 (1.2)	.831	-0.437 to 0.540	-0.062 (Trivial)	No	26.9
B3 vs B6	-0.526 (0.9)	.005	-0.879 to -0.172	0.894 (Moderate)	Yes	19.2
B3 vs B7	-0.09 (0.8)	.574	-0.414 to 0.235	0.120 (Trivial)	No	11.5
B3 vs B8	-0.564 (0.9)	.003	-0.923 to -0.205	0.885 (Moderate)	Yes	23.1
B3 vs B9	-0.449 (0.8)	.012	-0.791 to -0.106	0.624 (Moderate)	Yes	23.1
B3 vs B10	-0.359 (0.8)	.040	-0.699 to -0.019	0.532 (Small)	No	23.1
B4 vs B5	0.269 (0.8)	.094	-0.049 to 0.588	-0.318 (Small)	No	11.5
B4 vs B6	-0.308 (0.8)	.063	-0.633 to 0.017	0.505 (Small)	No	11.5
B4 vs B7	0.128 (0.8)	.434	-0.204 to 0.460	-0.167 (Trivial)	No	15.4
B4 vs B8	-0.346 (0.7)	.021	-0.635 to -0.057	0.527 (Small)	No	11.5

B4 vs B9	-0.231 (0.8)	.145	-0.547 to 0.085	0.313 (Small)	No	15.4
B4 vs B10	-0.141 (0.8)	.368	-0.458 to 0.176	0.203 (Small)	No	15.4
B5 vs B6	-0.577 (1.1)	.014	-1.025 to -0.129	0.790 (Moderate)	Yes	19.2
B5 vs B7	-0.141 (1.0)	.474	-0.541 to 0.259	0.163 (Trivial)	No	26.9
B5 vs B8	-0.615 (1.0)	.005	-1.027 to -0.204	0.798 (Moderate)	Yes	38.5
B5 vs B9	-0.5 (1.0)	.019	-0.909 to -0.091	0.595 (Small)	Yes	30.8
B5 vs B10	-0.41 (1.1)	.062	-0.843 to 0.023	0.512 (Small)	No	23.1
B6 vs B7	0.436 (1.1)	.047	0.005 to 0.866	-0.686 (Moderate)	Yes	23.1
B6 vs B8	-0.038 (0.7)	.767	-0.303 to 0.226	0.077 (Trivial)	No	3.8
B6 vs B9	0.077 (0.9)	.665	-0.285 to 0.439	-0.128 (Trivial)	No	19.2
B6 vs B10	0.167 (0.9)	.327	-0.177 to 0.510	-0.306 (Small)	No	11.5
B7 vs B8	-0.474 (0.7)	.003	-0.767 to -0.181	0.696 (Moderate)	Yes	15.4
B7 vs B9	-0.359 (0.7)	.012	-0.633 to -0.085	0.473 (Small)	No	19.2
B7 vs B10	-0.269 (0.7)	.076	-0.569 to 0.031	0.376 (Small)	No	11.5
B8 vs B9	0.115 (0.6)	.294	-0.106 to 0.337	-0.178 (Trivial)	No	3.8
B8 vs B10	0.205 (0.8)	.206	-0.120 to 0.531	-0.343 (Small)	No	11.5
B9 vs B10	0.09 (0.8)	.560	-0.223 to 0.403	-0.131 (Trivial)	No	11.5

Appendix 5: Between-block differences in hits during retraining. Abbreviations: CI, confidence interval; ES, Hedges' g; SEM, standard error of measurement; MDC, minimal detectable change. **Note:** Values are presented as mean difference (SD), with negative values indicating higher performance in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" reflects group-level change beyond measurement error, while "% Exceeding MDC" reflects individual-level meaningful change. SEM = ±1.940; MDC = ±5.377.

	All Participants (n=26)				1-wk Detraining (n=13)		2-wk Detraining (n=13)	
Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC	Exceeded SEM	% Exceeding MDC
B1 vs B2	-0.987 (3.2)	.130	-2.287 to 0.312	0.119 (Trivial)	No	23.1	No	0.0
B1 vs B3	-1.372 (3.4)	.052	-2.757 to 0.014	0.165 (Trivial)	No	23.1	No	15.4
B1 vs B4	-1.269 (4.9)	.201	-3.262 to 0.723	0.160 (Trivial)	No	15.4	No	15.4
B1 vs B5	-2.359 (3.8)	.004	-3.896 to -0.821	0.280 (Small)	Yes	23.1	No	23.1
B1 vs B6	-2.590 (4.0)	.003	-4.202 to -0.977	0.316 (Small)	Yes	23.1	Yes	38.5
B1 vs B7	-2.244 (4.4)	.016	-4.023 to -0.464	0.260 (Small)	No	30.8	Yes	23.1
B1 vs B8	-2.679 (4.3)	.004	-4.437 to -0.922	0.317 (Small)	Yes	15.4	Yes	23.1
B1 vs B9	-3.538 (4.6)	<.001	-5.413 to -1.664	0.414 (Small)	Yes	23.1	Yes	46.2
B1 vs B10	-3.910 (5.7)	.002	-6.213 to -1.608	0.431 (Small)	Yes	46.2	Yes	69.2
B2 vs B3	-0.385 (3.1)	.534	-1.645 to 0.876	0.044 (Trivial)	No	7.7	No	0.0
B2 vs B4	-0.282 (4.7)	.761	-2.179 to 1.615	0.034 (Trivial)	No	23.1	No	15.4
B2 vs B5	-1.372 (4.1)	.104	-3.046 to 0.303	0.155 (Trivial)	No	23.1	No	15.4

B2 vs B6	-1.603 (5.2)	.128	-3.700 to 0.495	0.185 (Trivial)	No	23.1	Yes	30.8
B2 vs B7	-1.256 (4.8)	.195	-3.205 to 0.692	0.139 (Trivial)	No	30.8	Yes	23.1
B2 vs B8	-1.692 (5.0)	.099	-3.732 to 0.347	0.191 (Trivial)	No	23.1	Yes	23.1
B2 vs B9	-2.551 (5.4)	.024	-4.738 to -0.364	0.285 (Small)	Yes	30.8	Yes	46.2
B2 vs B10	-2.923 (6.0)	.021	-5.370 to -0.476	0.309 (Small)	Yes	30.8	Yes	69.2
B3 vs B4	0.103 (4.8)	.915	-1.856 to 2.061	-0.012 (Trivial)	No	7.7	No	23.1
B3 vs B5	-0.987 (4.0)	.224	-2.621 to 0.646	0.112 (Trivial)	Yes	7.7	No	15.4
B3 vs B6	-1.218 (5.5)	.271	-3.451 to 1.015	0.141 (Trivial)	No	23.1	No	30.8
B3 vs B7	-0.872 (4.1)	.284	-2.515 to 0.772	0.097 (Trivial)	No	7.7	No	23.1
B3 vs B8	-1.308 (4.6)	.162	-3.179 to 0.564	0.148 (Trivial)	No	15.4	No	15.4
B3 vs B9	-2.167 (5.1)	.040	-4.229 to -0.104	0.242 (Small)	Yes	30.8	No	38.5
B3 vs B10	-2.538 (5.4)	.024	-4.719 to -0.358	0.269 (Small)	Yes	30.8	No	53.8
B4 vs B5	-1.090 (3.8)	.152	-2.612 to 0.433	0.128 (Trivial)	Yes	23.1	No	15.4
B4 vs B6	-1.321 (5.1)	.202	-3.400 to 0.759	0.160 (Trivial)	No	23.1	No	30.8
B4 vs B7	-0.974 (4.8)	.309	-2.912 to 0.964	0.112 (Trivial)	No	15.4	No	15.4
B4 vs B8	-1.410 (5.8)	.228	-3.764 to 0.943	0.166 (Trivial)	No	15.4	No	38.5
B4 vs B9	-2.269 (5.5)	.046	-4.496 to -0.042	0.264 (Small)	Yes	23.1	No	30.8
B4 vs B10	-2.641 (6.7)	.056	-5.352 to 0.070	0.290 (Small)	Yes	38.5	No	61.5
B5 vs B6	-0.231 (3.3)	.725	-1.573 to 1.111	0.0226 (Trivial)	No	0.0	No	23.1
B5 vs B7	0.115 (3.0)	.848	-1.114 to 1.345	-0.013 (Trivial)	No	15.4	No	0.0
B5 vs B8	-0.321 (4.0)	.687	-1.944 to 1.303	0.036 (Trivial)	No	7.7	No	30.8
B5 vs B9	-1.179 (3.8)	.132	-2.741 to 0.382	0.130 (Trivial)	No	7.7	No	7.7
B5 vs B10	-1.551 (5.6)	.169	-3.813 to 0.711	0.162 (Trivial)	No	7.7	No	53.8
B6 vs B7	0.346 (4.0)	.659	-1.257 to 1.950	-0.039 (Trivial)	No	7.7	No	15.4
B6 vs B8	-0.090 (4.5)	.920	-1.921 to 1.742	0.010 (Trivial)	No	0.0	No	23.1
B6 vs B9	-0.949 (4.6)	.305	-2.819 to 0.922	0.107 (Trivial)	No	0.0	No	30.8
B6 vs B10	-1.321 (5.6)	.240	-3.585 to 0.944	0.141 (Trivial)	Yes	23.1	No	46.2

B7 vs B8	-0.436 (3.4)	.519	-1.812 to 0.940	0.048 (Trivial)	No	7.7	No	15.4
B7 vs B9	-1.295 (3.7)	.090	-2.806 to 0.216	0.140 (Trivial)	Yes	7.7	No	7.7
B7 vs B10	-1.667 (4.3)	.060	-3.411 to 0.077	0.171 (Trivial)	Yes	30.8	No	30.8
B8 vs B9	-0.859 (3.2)	.183	-2.154 to 0.436	0.095 (Trivial)	No	0.0	No	15.4
B8 vs B10	-1.231 (4.4)	.171	-3.032 to 0.570	0.128 (Trivial)	Yes	7.7	No	30.8
B9 vs B10	-0.372 (4.0)	.641	-1.999 to 1.255	0.038 (Trivial)	No	0.0	No	23.1

Appendix 6. Between-block differences in average reaction time (AvgRT) during retraining.
Abbreviations: CI, confidence interval; ES, Hedges' g; AvgRT, average reaction time; SEM, standard error of measurement; MDC, minimal detectable change. Note: Values are presented as mean difference (SD), with positive values indicating faster reaction time (improvement) in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates group-level change beyond measurement error, whereas "% Exceeding MDC" reflects the proportion of participants demonstrating meaningful individual change. SEM = ±0.011; MDC = ±0.032

	All Participants (n=26)				1-wk Detraining (n=13)		2-wk Detraining (n=13)	
Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC	Exceeded SEM	% Exceeding MDC
B1 vs B2	0.008 (.015)	.013	0.002 to 0.015	-0.167 (Trivial)	No	0.0	No	15.4
B1 vs B3	0.006 (.020)	.126	-0.002 to 0.013	-0.138 (Trivial)	No	7.7	No	0.0
B1 vs B4	0.007 (.020)	.116	-0.002 to 0.016	-0.146 (Trivial)	No	7.7	No	15.4
B1 vs B5	0.014 (.025)	.014	0.003 to 0.025	-0.333 (Small)	Yes	30.8	Yes	30.8
B1 vs B6	0.01 (.025)	.073	-0.001 to 0.020	-0.223 (Small)	No	7.7	Yes	30.8
B1 vs B7	0.01 (.025)	.013	0.000 to 0.020	-0.225 (Small)	No	23.1	Yes	23.1
B1 vs B8	0.015 (.025)	.006	0.005 to 0.025	-0.345 (Small)	Yes	30.8	Yes	15.4
B1 vs B9	0.017 (.025)	.004	0.006 to 0.027	-0.372 (Small)	Yes	38.5	Yes	30.8
B1 vs B10	0.018 (.031)	.011	0.004 to 0.031	-0.355 (Small)	Yes	38.5	Yes	61.5
B2 vs B3	-0.003 (.015)	.418	-0.009 to 0.004	0.027 (Trivial)	No	7.7	No	0.0
B2 vs B4	-0.001 (.025)	.796	-0.011 to 0.009	0.018 (Trivial)	No	15.4	No	15.4
B2 vs B5	0.006 (.025)	.260	-0.005 to 0.016	-0.168 (Trivial)	No	15.4	No	15.4
B2 vs B6	0.001 (.031)	.807	-0.01 to 0.013	-0.061 (Trivial)	No	15.4	No	23.1
B2 vs B7	0.002 (.020)	.634	-0.006 to 0.010	-0.062 (Trivial)	No	0.0	No	23.1

B2 vs B8	0.007 (.025)	.156	-0.003 to 0.016	-0.178 (Trivial)	No	23.1	No	15.4
B2 vs B9	0.008 (.025)	.117	-0.002 to 0.019	-0.209 (Small)	No	23.1	No	30.8
B2 vs B10	0.009 (.031)	.104	-0.002 to 0.021	-0.203 (Small)	No	23.1	No	53.8
B3 vs B4	0.001 (.020)	.745	-0.007 to 0.010	-0.009 (Trivial)	No	15.4	No	23.1
B3 vs B5	0.009 (.025)	.100	-0.002 to 0.019	-0.192 (Trivial)	No	15.4	No	23.1
B3 vs B6	0.004 (.031)	.479	-0.008 to 0.016	-0.087 (Trivial)	No	15.4	No	15.4
B3 vs B7	0.005 (.020)	.295	-0.004 to 0.014	-0.087 (Trivial)	No	7.7	No	23.1
B3 vs B8	0.009 (.009)	.066	-0.001 to 0.019	-0.203 (Small)	Yes	15.4	No	15.4
B3 vs B9	0.011 (.025)	.059	0.000 to 0.022	-0.232 (Small)	Yes	23.1	No	23.1
B3 vs B10	0.012 (.031)	.050	-2.669E-5 to 0.024	-0.226 (Small)	Yes	23.1	No	38.5
B4 vs B5	0.007 (.020)	.085	-0.001 to 0.015	-0.182 (Trivial)	No	15.4	Yes	7.7
B4 vs B6	0.003 (.020)	.509	-0.006 to 0.011	-0.077 (Trivial)	No	7.7	No	7.7
B4 vs B7	0.003 (.020)	.407	-0.005 to 0.011	-0.078 (Trivial)	No	23.1	No	7.7
B4 vs B8	0.008 (.020)	.084	-0.001 to 0.017	-0.193 (Trivial)	No	15.4	Yes	30.8
B4 vs B9	0.009 (.025)	.049	4.725E-5 to 0.019	-0.222 (Small)	No	23.1	Yes	30.8
B4 vs B10	0.011 (.025)	.060	0.000 to 0.022	-0.216 (Small)	No	23.1	Yes	46.2
B5 vs B6	-0.004 (.020)	.256	-0.012 to 0.003	0.102 (Trivial)	No	7.7	No	23.1
B5 vs B7	-0.004 (.015)	.219	-0.010 to 0.003	0.103 (Trivial)	No	7.7	No	7.7
B5 vs B8	0.001 (.015)	.819	-0.006 to 0.008	-0.009 (Trivial)	No	0.0	No	7.7
B5 vs B9	0.002 (.020)	.524	-0.005 to 0.010	-0.042 (Trivial)	No	15.4	No	7.7
B5 vs B10	0.003 (.025)	.511	-0.007 to 0.014	-0.048 (Trivial)	No	0.0	No	30.8
B6 vs B7	0.001 (.020)	.888	-0.007 to 0.008	0.000 (Trivial)	No	0.0	No	15.4
B6 vs B8	0.005 (.020)	.174	-0.002 to 0.013	-0.111 (Trivial)	No	0.0	No	23.1
B6 vs B9	0.007 (.020)	.129	-0.002 to 0.016	-0.142 (Trivial)	No	0.0	No	30.8
B6 vs B10	0.008 (.031)	.207	-0.005 to 0.021	-0.141 (Trivial)	No	15.4	No	38.5
B7 vs B8	0.005 (.015)	.100	-0.001 to 0.010	-0.112 (Trivial)	No	0.0	No	0.0
B7 vs B9	0.006 (.020)	.104	-0.001 to 0.014	-0.143 (Trivial)	No	7.7	No	7.7

B7 vs B10	0.007 (.020)	.104	-0.002 to 0.017	-0.142 (Trivial)	No	0.0	No	7.7
B8 vs B9	0.002 (.015)	.612	-0.005 to 0.008	-0.034 (Trivial)	No	0.0	No	7.7
B8 vs B10	0.003 (.025)	.588	-0.007 to 0.013	-0.040 (Trivial)	No	0.0	No	23.1
B9 vs B10	0.001 (.020)	.789	-0.008 to 0.010	-0.008 (Trivial)	No	0.0	No	15.4

Appendix 7. Between-block differences in Misses during retraining. Abbreviations: CI, confidence interval; ES, Hedges' g; SEM, standard error of measurement; MDC, minimal detectable change. Note: Values are presented as mean difference (SD), with negative values indicating fewer misses (improvement) in later blocks. Data are presented as raw values for interpretability. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates group-level change beyond measurement error, whereas "% Exceeding MDC" reflects the proportion of participants exceeding individual-level meaningful change. SEM = ±1.200; MDC = ±3.326.

	All Participants (n=26)				1-wk Detraining (n=13)		2-wk Detraining (n=13)	
Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC	Exceeded SEM	% Exceeding MDC
B1 vs B2	-0.205 (1.8)	.568	-0.938 to 0.528	0.082 (Trivial)	No	15.4	No	0.0
B1 vs B3	0.179 (1.7)	.592	-0.503 to 0.862	-0.073 (Trivial)	No	7.7	No	15.4
B1 vs B4	-0.038 (2.3)	.934	-0.985 to 0.909	0.017 (Trivial)	No	15.4	No	7.7
B1 vs B5	0.167 (1.1)	.466	-0.299 to 0.632	-0.069 (Trivial)	No	0.0	No	0.0
B1 vs B6	0.692 (1.6)	.033	0.062 to 1.323	-0.302 (Small)	No	0.0	No	7.7
B1 vs B7	0.359 (1.7)	.295	-0.333 to 1.051	-0.148 (Trivial)	No	7.7	No	0.0
B1 vs B8	0.385 (1.6)	.228	-0.258 to 1.027	-0.157 (Trivial)	No	0.0	No	7.7
B1 vs B9	0.667 (1.8)	.070	-0.058 to 1.392	-0.280 (Small)	No	7.7	No	15.4
B1 vs B10	0.590 (2.2)	.179	-0.290 to 1.469	-0.212 (Small)	No	0.0	No	15.4
B2 vs B3	0.385 (1.9)	.301	-0.368 to 1.137	-0.150 (Trivial)	No	7.7	No	0.0
B2 vs B4	0.167 (2.6)	.745	-0.881 to 1.215	-0.072 (Trivial)	No	23.1	Yes	23.1
B2 vs B5	0.372 (1.7)	.272	-0.312 to 1.055	-0.146 (Trivial)	No	7.7	No	7.7
B2 vs B6	0.897 (2.0)	.033	0.079 to 1.716	-0.373 (Small)	No	7.7	No	15.4
B2 vs B7	0.564 (2.3)	.223	-0.367 to 1.495	-0.223 (Small)	No	15.4	No	15.4
B2 vs B8	0.590 (2.2)	.186	-0.305 to 1.485	-0.230 (Small)	No	7.7	No	15.4
B2 vs B9	0.872 (2.3)	.062	-0.046 to 1.789	-0.349 (Small)	No	7.7	No	30.8

B2 vs B10	0.795 (3.0)	.187	-0.414 to 2.003	-0.276 (Small)	No	15.4	No	23.1
B3 vs B4	-0.218 (2.5)	.662	-1.237 to 0.801	0.094 (Trivial)	No	7.7	No	23.1
B3 vs B5	-0.013 (1.7)	.969	-0.697 to 0.671	0.005 (Trivial)	No	0.0	No	7.7
B3 vs B6	0.513 (2.3)	.262	-0.409 to 1.435	-0.215 (Small)	No	7.7	No	30.8
B3 vs B7	0.179 (1.9)	.631	-0.584 to 0.943	-0.071 (Trivial)	No	0.0	No	15.4
B3 vs B8	0.205 (2.0)	.607	-0.609 to 1.019	-0.081 (Trivial)	No	7.7	No	15.4
B3 vs B9	0.487 (2.1)	.252	-0.370 to 1.345	-0.197 (Trivial)	No	0.0	No	30.8
B3 vs B10	0.410 (2.4)	.390	-0.559 to 1.379	-0.143 (Trivial)	No	7.7	No	23.1
B4 vs B5	0.205 (2.1)	.626	-0.654 to 1.064	-0.090 (Trivial)	Yes	15.4	No	7.7
B4 vs B6	0.731 (2.5)	.155	-0.298 to 1.760	-0.345 (Small)	Yes	15.4	No	15.4
B4 vs B7	0.397 (2.7)	.458	-0.693 to 1.488	-0.176 (Trivial)	No	7.7	No	15.4
B4 vs B8	0.423 (3.0)	.475	-0.782 to 1.628	-0.184 (Trivial)	No	23.1	No	7.7
B4 vs B9	0.705 (2.9)	.232	-0.482 to 1.893	-0.318 (Small)	Yes	15.4	No	15.4
B4 vs B10	0.628 (3.4)	.361	-0.766 to 2.022	-0.237 (Small)	Yes	30.8	No	7.7
B5 vs B6	0.526 (1.2)	.038	0.031 to 1.021	-0.223 (Small)	No	0.0	No	7.7
B5 vs B7	0.192 (1.5)	.516	-0.411 to 0.796	-0.077 (Trivial)	No	0.0	No	0.0
B5 vs B8	0.218 (1.6)	.504	-0.446 to 0.882	-0.087 (Trivial)	No	0.0	No	15.4
B5 vs B9	0.500 (1.5)	.106	-0.115 to 1.115	-0.205 (Small)	No	0.0	No	15.4
B5 vs B10	0.423 (2.2)	.342	-0.480 to 1.326	-0.149 (Trivial)	No	7.7	No	15.4
B6 vs B7	-0.333 (1.7)	.322	-1.015 to 0.349	0.143 (Trivial)	No	15.4	No	0.0
B6 vs B8	-0.308 (1.8)	.381	-1.020 to 0.405	0.129 (Trivial)	No	7.7	No	7.7
B6 vs B9	-0.026 (1.7)	.939	-0.711 to 0.660	0.011 (Trivial)	No	0.0	No	7.7
B6 vs B10	-0.103 (2.0)	.798	-0.920 to 0.715	0.038 (Trivial)	No	7.7	No	7.7
B7 vs B8	0.026 (1.3)	.923	-0.519 to 0.570	-0.011 (Trivial)	No	7.7	No	0.0
B7 vs B9	0.308 (1.6)	.339	-0.344 to 0.959	-0.127 (Trivial)	No	7.7	No	0.0
B7 vs B10	0.231 (2.0)	.554	-0.564 to 1.025	-0.082 (Trivial)	No	15.4	No	7.7
B8 vs B9	0.282 (1.5)	.355	-0.335 to 0.900	-0.114 (Trivial)	No	0.0	No	0.0

B8 vs B10	0.205 (1.7)	.539	-0.475 to 0.886	-0.071 (Trivial)	No	15.4	No	7.7
B9 vs B10	-0.077 (1.7)	.816	-0.751 to 0.597	0.028 (Trivial)	No	7.7	No	7.7

Appendix 8. Between-block differences in TSC responses during retraining. Abbreviations: CI, confidence interval; ES, Hedges' g; SEM, standard error of measurement; MDC, minimal detectable change.
Note: Values are presented as mean difference (SD), with negative values indicating improved performance in later blocks. P-values reflect pairwise comparisons derived from the repeated-measures ANOVA using least significant difference (LSD) post-hoc tests. Effect sizes (ES) are reported as Hedges' g and interpreted as trivial (<0.2), small (0.2–0.49), moderate (0.5–0.79), large (0.8–1.19), and very large (≥1.2). "Exceeded SEM" indicates whether the group mean difference exceeded measurement error (±SEM). "% Exceeding MDC" reflects the proportion of participants demonstrating meaningful individual change. SEM = ±0.423; MDC = ±1.172.

	All Participants (n=26)				1-wk Detraining (n=13)		2-wk Detraining (n=13)	
Block	MD (SD)	P value	95% CI	ES	Exceeded SEM	% Exceeding MDC	Exceeded SEM	% Exceeding MDC
B1 vs B2	0.103 (0.8)	0.539	-0.238 to 0.443	-0.149 (Trivial)	No	7.7	No	7.7
B1 vs B3	-0.064 (0.5)	0.531	-0.272 to 0.144	0.105 (Trivial)	No	0.0	No	7.7
B1 vs B4	-0.115 (0.6)	0.318	-0.349 to 0.119	0.189 (Trivial)	No	0.0	No	7.7
B1 vs B5	-0.077 (0.6)	0.529	-0.326 to 0.172	0.121 (Trivial)	No	0.0	No	15.4
B1 vs B6	0.154 (1.0)	0.437	-0.248 to 0.556	-0.208 (Small)	No	0.0	No	7.7
B1 vs B7	-0.244 (0.5)	0.016	-0.438 to -0.049	0.463 (Small)	No	0.0	No	15.4
B1 vs B8	-0.051 (0.5)	0.622	-0.264 to 0.161	0.085 (Trivial)	No	0.0	No	0.0
B1 vs B9	-0.205 (0.6)	0.101	-0.454 to 0.043	0.403 (Small)	No	0.0	No	15.4
B1 vs B10	-0.128 (0.7)	0.337	-0.399 to 0.143	0.214 (Small)	No	7.7	No	15.4
B2 vs B3	-0.167 (1.0)	0.397	-0.566 to 0.233	0.249 (Small)	No	23.1	No	15.4
B2 vs B4	-0.218 (1.0)	0.265	-0.612 to 0.176	0.323 (Small)	No	15.4	No	15.4
B2 vs B5	-0.179 (1.0)	0.381	-0.595 to 0.236	0.259 (Small)	No	23.1	No	7.7
B2 vs B6	0.051 (1.1)	0.810	-0.385 to 0.488	-0.065 (Trivial)	No	15.4	No	15.4
B2 vs B7	-0.346 (0.9)	0.056	-0.701 to 0.009	0.577 (Small)	No	7.7	Yes	30.8
B2 vs B8	-0.154 (0.9)	0.375	-0.506 to 0.198	0.231 (Small)	No	7.7	No	7.7
B2 vs B9	-0.308 (1.0)	0.126	-0.709 to 0.093	0.526 (Small)	No	7.7	Yes	15.4
B2 vs B10	-0.231 (1.1)	0.279	-0.661 to 0.199	0.348 (Small)	No	23.1	No	23.1
B3 vs B4	-0.051 (0.6)	0.658	-0.288 to 0.185	0.089 (Trivial)	No	7.7	No	0.0

B3 vs B5	-0.013 (0.5)	0.893	-0.208 to 0.183	0.021 (Trivial)	No	0.0	No	0.0
B3 vs B6	0.218 (1.0)	0.271	-0.182 to 0.618	-0.301 (Small)	No	7.7	No	7.7
B3 vs B7	-0.179 (0.5)	0.092	-0.391 to 0.032	0.363 (Small)	No	0.0	No	7.7
B3 vs B8	0.013 (0.4)	0.885	-0.169 to 0.195	-0.022 (Trivial)	No	0.0	No	0.0
B3 vs B9	-0.141 (0.7)	0.330	-0.434 to 0.152	0.297 (Small)	No	7.7	No	7.7
B3 vs B10	-0.064 (0.6)	0.605	-0.317 to 0.189	0.114 (Trivial)	No	7.7	No	0.0
B4 vs B5	0.038 (0.6)	0.762	-0.221 to 0.298	-0.065 (Trivial)	No	7.7	No	15.4
B4 vs B6	0.269 (0.9)	0.158	-0.112 to 0.651	-0.369 (Small)	No	7.7	No	23.1
B4 vs B7	-0.128 (0.6)	0.288	-0.372 to 0.116	0.256 (Small)	No	0.0	No	7.7
B4 vs B8	0.064 (0.7)	0.623	-0.202 to 0.33	-0.110 (Trivial)	No	7.7	No	7.7
B4 vs B9	-0.09 (0.8)	0.587	-0.426 to 0.247	0.186 (Trivial)	No	0.0	No	23.1
B4 vs B10	-0.013 (0.7)	0.931	-0.314 to 0.288	0.023 (Trivial)	No	7.7	No	23.1
B5 vs B6	0.231 (0.9)	0.221	-0.149 to 0.61	-0.310 (Small)	No	15.4	No	7.7
B5 vs B7	-0.167 (0.7)	0.212	-0.435 to 0.102	0.321 (Small)	No	0.0	No	15.4
B5 vs B8	0.026 (0.5)	0.813	-0.196 to 0.248	-0.041 (Trivial)	No	0.0	No	0.0
B5 vs B9	-0.128 (0.7)	0.343	-0.402 to 0.145	0.257 (Small)	No	7.7	No	7.7
B5 vs B10	-0.051 (0.5)	0.610	-0.256 to 0.154	0.089 (Trivial)	No	0.0	No	7.7
B6 vs B7	-0.397 (1.0)	0.057	-0.807 to 0.012	0.601 (Moderate)	No	15.4	Yes	30.8
B6 vs B8	-0.205 (1.0)	0.299	-0.605 to 0.194	0.285 (Small)	No	0.0	No	7.7
B6 vs B9	-0.359 (1.0)	0.087	-0.774 to 0.056	0.554 (Small)	No	15.4	Yes	15.4
B6 vs B10	-0.282 (0.9)	0.143	-0.667 to 0.103	0.392 (Small)	No	15.4	No	7.7
B7 vs B8	0.192 (0.5)	0.048	0.002 to 0.383	-0.391 (Small)	No	0.0	No	0.0
B7 vs B9	0.038 (0.5)	0.720	-0.181 to 0.258	-0.101 (Trivial)	No	0.0	No	7.7
B7 vs B10	0.115 (0.7)	0.432	-0.183 to 0.414	-0.234 (Small)	No	0.0	No	23.1
B8 vs B9	-0.154 (0.6)	0.210	-0.401 to 0.093	0.326 (Small)	No	7.7	No	7.7
B8 vs B10	-0.077 (0.7)	0.557	-0.344 to 0.19	0.136 (Trivial)	No	7.7	No	7.7
B9 vs B10	0.077 (0.7)	0.587	-0.212 to 0.366	-0.163 (Trivial)	No	0.0	No	15.4