ACT Golf: A Preliminary Investigation into Effects on Sport Performance with Acceptance and Commitment Training Consistent Instructions.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Psychology

By

Patrick Smith

Dr. Steven C Hayes/Thesis Advisor

August, 2021
THE GRADUATE SCHOOL

We recommend that the thesis prepared under our supervision by

PATRICK SMITH

Entitled

ACT Golf: A Preliminary Investigation into Effects on Sport Performance with Acceptance and Commitment Training Consistent Instructions.

be accepted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Steven C Hayes, Ph.D.

Advisor

W. Larry Williams, Ph.D.

Committee Member

James Sundali, Ph.D.

Graduate School Representative

David W. Zeh, Ph.D., Dean

Graduate School

August, 2021
Abstract

Sport is an important area of human life dominated by rule governed skilled performances. Implications of Relational Frame Theory (RFT) and Acceptance and Commitment Training (ACT) in human sport performance suggest that status quo methods of training skilled tasks may be counterproductive. Previous behavioral research has not tested this implication directly nor compared status quo Control type instructions to ACT informed Functional Awareness instructions. Additional implications of behavioral variation as a moderator to long term performance suggest novel measurement and analysis of a skilled performance task such as golf putting.

In a multiple baseline across participants experiment, participants completed half of their assigned putt attempts from a fixed distance before reading one of two instructional interventions and then completing their remaining assigned putts. Putt outcomes were categorized as hit or one of 8 miss types based on combinations of relative speed and hole alignment. Overall accuracy and putting outcome variation were analyzed for intervention effects both within subjects and across intervention groups.

Results suggest both interventions had no effect on accuracy or variability of putting behavior. Implications of the lack of an effect by popular status quo putting instructions, the novel preparation, variability analysis, as well as study strengths and limitations discussed in light of the data.
Acknowledgments

Tara and Lucy, this is as much yours as it is mine. Steve, thank you for letting me
play, explore, grow and for always being in my corner.

There are so many who have touched this project. All in a meaningful way.
Sufficient acknowledgement of my deep appreciation would outnumber the pages of the
manuscript itself. While there is only one name on the byline, I am, and by extension this
is, a reflection of my history with each of you. A history I cherish and appreciate making
every day. Thank you.
Table of Contents

ABSTRACT .................................................................................................................................I

ACKNOWLEDGMENTS ................................................................................................................II

LIST OF TABLES .........................................................................................................................IV

LIST OF FIGURES .....................................................................................................................V

INTRODUCTION .............................................................................................................................1

METHOD ........................................................................................................................................18

RESULTS ........................................................................................................................................29

DISCUSSION ..................................................................................................................................41

REFERENCES ..................................................................................................................................50
List of Tables

Table 1: Participant Descriptive Statistics
Table 2: Participant Demographics by Various Groupings
Table 3: Participant Psychometric Measures Within Group Across Time
List of Figures

Figure 1: Experimental Structure
Figure 2: Putting Setup Diagram (Top) and During Session (Bottom)
Figure 3: Cumulative Record of Hits 34 Putt Attempt Group
Figure 4: Cumulative Record of Hits 50 Putt Attempt Group
Figure 5: Cumulative Record of Hits 67 Putt Attempt Group
**Introduction**

Sport is an important area of human life. In 2019, in the US alone, each day one in five individuals 15 and older engaged in sport activities, and averaged more time participating than their leisure time on the computer, reading for personal interest, or playing games (Bureau of Labor Statistics, 2020). While these statistics include activities like walking for exercise (~30% of respondents) and use of cardiovascular equipment (~9%), 6 of the top ten activities reported had competitive variants (e.g. weightlifting, running, cycling, golf, soccer, etc.) and accounted for over 30% of respondents. This suggests that on any given day in the US, roughly 8 million individuals engage in, possibly competitive, sport activity (U.S. Census Bureau, 2011).

The same systems that train athletes to compete can, unintentionally, result in dysfunctional levels of rule following behaviors. The competitor that is the best at maximizing their behavior within the bounds of their competition’s set of rules has a significant advantage. To be the best competitor, the athlete must experience regular reinforcement for following sport rules with a heightened level of insensitivity toward consequences operating on them and those around them. Within many competition communities, this culture of tolerating sustained physical and mental discomfort is even turned into borderline pyrrhic competitions. The CrossFit athlete that trains to the point of their cells self-destructing wears the diagnosis of rhabdomyolysis as a badge of honor (Glassman, 2005). The cyclist that endures hours of discomfort and dehydration gets to tell ‘war stories’ to others of their ability to suffer the best. In endurance communities, willful and intentional suffering even has a name; “Type II Fun” (Dunfee, 2018). Once
sport is viewed through the lens of rule governed behavior it is unsurprising that mental health clinical frequencies are higher than normal in these populations (Schaal et al., 2011).

But elimination of sport to address adverse mental health outcomes is not likely and would be a pyrrhic victory itself. Physical inactivity is literally killing millions and sport is an enormous source of life extending, society strengthening, activity. So, a solution must shape contextually sensitive human flourishing behaviors while enabling individuals to be active and compete at whatever level they choose.

**Growth of Sport Performance Science**

The science of sport performance has grown enormously, especially in areas of biometrics, nutrition, training physiology, and the like. Analysis of sport science publications between 2001 and 2018 found that over 50% of those have occurred in the most recent 5-year period and were dominated by discussions of performance (Belfiore et al., 2019). The largest North American based professional organization in sport science, the American College of Sports Medicine (ACMS), boasts over 50,000 members representing 90 nationalities and claims to be “the only organization that offers a 360-degree view of the profession” (American College of Sports Medicine, 2021).

Within ACSM, mental and behavioral performance goes nearly unrepresented and sport performance is approached from a perspective of the body as a machine. This perspective emphasizes performance problem solving through a comprehensive mechanical understanding of tissues, transmitters, and the myriad bodily systems. And while this focus can fill in understanding questions like when is the musculature most
likely to benefit from weight training (American College of Sports Medicine, 2009), or what the dose effects of such training might be (Garber et al., 2011), it can also result in interventions that target those same bodily systems in a decontextualized way that may proximally reduce symptoms but otherwise make no functional difference for the individual within the environment they operate in.

Mental and behavioral performance is a much smaller area of research, but it too has been growing. The Association for Applied Sport Psychology (AASP) is the leading North American professional organization dedicated to the topic and reports over 3,000 members (Association for Applied Sport Psychology, n.d.). According to Web of Science, citations of their flagship publication, the Journal of Applied Sports Psychology, has an annual growth rate of ~16% since 2004. Over 30% of those citations have been in the last five years.

**Performance as a Control Effort**

Similar to the medical approach, traditional coaching methods for sport performance have emphasized the individual in a decontextualized and mechanical manner. Reflected in Dr. Vealey’s repeated assertions that sport psychology is “a distinct subdiscipline of kinesiology” (Vealey, 2006), this mechanical human system focus has resulted in behavioral and mental interventions that emphasize the athlete controlling their forms of motions (Davids et al., 2003), emotions (Jones, 2003), confidence (Woodman & Hardy, 2003), self-efficacy (Vancouver et al., 2002), and attention (Wulf & Su, 2007).
This emphasis on an optimal form to achieve desired sport outcomes is collected in the widely adopted sport psychology model of Psychological Skills Training (PST; Boutcher & Rotella, 1987; Vealey, 1994). PST, as a set of specified skills and/or interventions, varies widely (Beckmann, 2001; Lane, 2008), possibly due to the generic and descriptive nature of the name, but generally teaches self-regulation. Self-regulation is divided into efforts of describing a preferred outcome (Goal Setting), repeatedly imagining that preferred outcome (Visualization), altering one’s physiological arousal level (Arousal Control or Relaxation), altering the focal point of one’s attention (Concentration), and suppressing unwanted cognitions (Thought Stopping) (Beckmann, 2001). Through these and similar efforts, PST purports to control the formal state of the individual such that they are most likely to perform at their peak.

Research with PST demonstrated positive effects on performance, psychometric measures, or both across sports including soccer (Slimani et al., 2016; Thelwell et al., 2006), swimming (Sheard & Golby, 2006), endurance events (McCormick et al., 2015), and golf (Bernier et al., 2009). Meta-analysis of PST in sport using single case experimental design found evidence of a positive effect publication bias within the literature as well as a negative correlation between reported effect sizes and design quality (Barker, 2020). This finding suggests that measured and reported effects of PST may be overestimated and have limited generalizability due to procedural quality confounds. While these findings were limited to experimental data from single case designs, authors noted that between-group designs historically further inflate effect sizes and within-subject over time, idiographic, experimental design has long been considered the gold standard for identifying effective performance interventions (Smith, 1989).
**Paradox of Control & Insensitivity**

In addition to effect reporting and methodological criticisms within the literature, attempts to control experiences is contraindicated by one of the most well documented clinical paradoxes; the White Bear Effect (Wegner et al., 1987). In brief, to be told to ‘not think of a white bear’ is to elicit the thought of a white bear. Not only does this effort to eliminate a thought result in further strengthening and elaboration of the thought, it can bring the experience to such dominance that the individual becomes less responsive to other experiences that may be more useful. In a sports context, telling someone to stop having undesirable thoughts and feelings is a sure way for them to experience those thoughts and feelings and be less sensitive to the moment. Relatedly, instructing someone to bring to mind what they most want primes them to readily recall the exact opposite with only the slightest of suggestions. Thus, attempts at intervention based on direct control efforts has a likely outcome of entrenching the targeted undesired experiences and rendering the individual insensitive to the present moment.

**Acceptance & Verbal Instruction**

Modern behavioral methods address the White Bear paradox and insensitivity through training skills including acceptance and awareness. Acceptance paradigm methods developed within the clinical behavioral science wing of psychology and the two most visible examples are Acceptance and Commitment Therapy (or Training in non-clinical applications) (ACT pronounced as the word; Hayes et al., 2012) and Mindful-Acceptance-Commitment (MAC; Gardner & Moore, 2004). Both train skills around open and willing noticing, experiencing, and engaging with the present moment in ways that
result in outcomes most aligned with personally valued directions. Where ACT encourages clinicians to “dance” between six contributing processes (skills) as the interaction with clients surface opportunities, MAC uses a manualized set sequence of didactic and experiential steps in order to effect functionally similar outcomes for clients.

A core issue in the acceptance paradigm methods is how verbal instruction influences behavior, sensitivity to outcomes, and the impact of experiences. The way an individual is influenced by instructions they generate for themselves as well as are provided by others has been studied by behavior analysts through the lens of Rule Governed Behavior and rule-based insensitivity. The most studied behavioral theory encompassing these concepts is Relational Frame Theory (RFT; Hayes et al., 2001).

**Relating as Language Behavior**

Relational Frame Theory is a behavioral account of cognition (thoughts and feelings) based on operant and respondent principles of learning. It describes how deriving a mutual relation between two events when only one relation is directly trained in a given instance can be a product of reinforcing mutual relations among multiple sets of arbitrary and non-arbitrary stimuli. How those derived and trained responses are brought under the control of arbitrary stimulus cues. How the function (antecedent, consequential, or respondent) of those cues become contextually discriminated and transform the functions of other elements related by training or derivation history of the organism. In somewhat simpler terms, RFT is an account of how language, and thus most cognition, is a learned behavioral repertoire influenced by experiences. The premise of that account being that the arbitrarily applicable derived relational response (AADRR) is
the keystone behavior. The literature commonly drops the “D” and uses the acronym AARR synonymously.

Derivation combined with arbitrary application make AARRing behavioral repertoires extremely generative. Derivation is an operant behavior that has two variants. Consistent responding to the reversal of a trained relationship without a history of reinforcement for that specific response is Mutual Entailment. For example, from a trained comparative relation such as A>B, a verbally competent person will derive that B<A. Symmetry, as found in stimulus equivalence, is usually the first and simplest form of Mutual Entailment to emerge in children -- in this case derivation and is evidenced by responding to B as Same-as A when only A Same-as B has been trained (Sidman & Tailby, 1982). Consistently responding to both novel mutually entailed relations between two stimuli, never directly reinforced, through combining two previously trained relations among three stimuli is Combinatorial Entailment. (i.e., training A>B & A<C, leads to derived B<C and C>B). There are many types of arbitrarily applicable derived relational responses (opposition, difference, hierarchy, and so on). Derivation of the mutually and combinatorially entailed relations of a small number (n) of trained relations results in an exponential number of possible derived responses (n^2) and an overall repertoire of n+n^2 probable responses. The common saying about this exponential additive outcome of derivation is “Train one, get two, train two, get six.”

Arbitrary applicability is a characteristic of relational responses where learned and derived relations between stimuli are limitless. Specifically, not limited to non-arbitrary physical (aka formal) properties of those stimuli. Relational responding under the influence of non-arbitrary features of stimuli (i.e. picking the taller stimuli based in
relative height or the heavier stimuli based on relative mass) has been well documented in both human and non-human species (Galizio & Bruce, 2018), but the absence of AARR has been consistent in non-human species (Dugdale & Lowe, 2000; Hayes, 1989; Lionello-DeNolf & Urcuioli, 2002; Sidman et al., 1982). This is in contrast to decades of successful experimental and applied demonstrations of AARR with humans (Hayes et al., 2021). As such, while still somewhat debated (Hughes & Barnes-Holmes, 2014), arbitrary applicability is considered the defining characteristic of human language and the key to many orders of magnitude more generativity and flexibility than most other behavioral repertoires. In simple terms, arbitrary applicability means that, given a particular history of reinforcement, a language-able human can respond to any stimuli as if it is related to any other stimuli in any way imaginable. And the term stimuli here is inclusive of both relations functioning as eliciting stimuli themselves such that responses can be under the influence of relations among relations (Stewart & Barnes-Holmes, 2001) as well as relations functioning as consequential (Bordieri et al., 2016), establishing (Ju & Hayes, 2008), and respondent stimuli (Dougher et al., 1994).

An outcome of AARR is a detectable change in elicited response to stimuli called Transformation of Stimulus Function. This phenomenon is when the stimulus function(s) of one or more related stimuli changes once the individual has been reinforced for, or derives, a relation between a novel stimulus and some aspect of their established relational repertoire. Learning a second language is dense with this phenomenon as new words are related to established language repertoires. For example, an English only speaker doesn’t react to “gendarme” as related to law enforcement until learning it is French for “police”. The moment when “gendarme” exhibits police related stimulus
functions and/or vice versa for that individual, the stimulus function amongst related stimuli has transformed to cohere with the established relational repertoires.

**Relating in a Network**

Established repertoires of AARR are metaphorically called relational networks suggesting a web of connected nodes (stimuli) and branches (relations). The nodal distance and branch lengths represent the individual’s history of derivation and history respectively. Nodal distance, number of nodes between two stimuli in a network, represents the degrees of abstraction from one to the other (Barnes-Holmes et al., 2017). Branch length, proximity of two directly connected nodes, represents the probability of that relational response being elicited given interaction with either of the connected nodes (Belisle & Clayton, 2021). In this metaphor, branch length is negatively correlated with response probability and is relative to other branches off the same node. Relational network elements (stimuli and relations) and characteristics (arrangement, nodal distance, and branch length) for an individual are contextually bound such that one in context, like competition, competition specific stimuli and relational responses are more frequent and probable whereas in a social context, those same stimuli and responses are less frequent and less probable. The limit to this metaphor is that no relational response disappears entirely from an individual’s history, only the relative probability of occurrence of any particular response changes given an individual’s context and history.
**Rules & Insensitivity**

From an RFT point of view, insensitivity to experience may come from histories of reinforcement that shaped a relational network with extensive relating of stimuli (many branches, low nodal distance) high response probabilities amongst a few core responses such that nearly any experience is only one degree away from eliciting only one or a few extremely high probability responses that compete, at times destructively, with more nuanced sensitivity to changes in the moment. Rule Governed Behavior (RGB) is a subcategory of AARR repertoires described by RFT. For brevity, this paper will use the acronym of “RGB” or “rule following” interchangeably. RGB encompasses three specific histories of reinforcement that shape AARR behaviors that are functionally unique but formally similar: Pliance, Tracking, and Augmenting.

Pliance, derived from the word “compliance,” is behavior coherent with a given rule due to a history of the behaving individual being reinforced by observers for this coherence. In this case, given the right history, a behaving individual may also serve as the source of the rule and the observer reinforcing coherence (or punishing incoherence). Additionally, the consequence of reinforcement mediated by the observer is only contingent upon following the rule per se and does not account for the natural outcome of the behavior described by the rule itself. Pliant rule following is more probable when the provider of a rule has previously reinforced the individual for sticking to their rules regardless of other outcomes. A blindfolded individual may continue to follow the directions from friends if following results in social praise, even if they repeatedly guide the blindfolded person into otherwise humiliating situations. Insensitivity in the case of
pliance is relative to the context and consequences not otherwise connected to the source of primary reinforcement. The individual complying with peer pressure is often caricatured as being insensitive to extreme risk (i.e., death) in the name of doing whatever thing their peers do (i.e., drink and drive, attend large parties during COVID, own firearms, etc.)

Tracking is behavior coherent with a given rule due to a history of the behaving individual being reinforced by the natural outcome of the contingency described by the rule. Tracking rule following is more probable when the provider of a rule has previously provided rules that described contingencies with reinforcing outcomes regardless of if the provider can observe any coincident adherence to said rule. When arriving at a destination is a reinforcing consequence, a historically consistent provider of accurate directions (directions being the rule(s) to be followed in this case) is more likely to elicit tracking behavior than an inconsistent provider. Similar to pliance, a behaving individual can be their own provider of rules that elicit tracking rule following. Insensitivity in the case of tracking is relative to the source of provided rules as well as available reinforcement for rule following per se. A person who always tells others exactly what they are thinking because that has historically resulted in reinforcing outcomes for them may be considered insensitive to others providing more nuanced social interaction rules or the availability of socially mediated reinforcement for complying with social niceties.

Augmenting is behavior coherent with a given rule due to an alteration of the consequent function (i.e. strengthens or weakens the reinforcing value) of a contingency, either via establishing the function of a consequence (formative augmenting) or temporarily changing the impact of a given consequence (motivative augmenting).
Telling a race car driver to drive a specific way such as “wait until the very last second and then brake hard coming into the chicane” may elicit tracking or pliance depending on their history with the source of the rule. Adding “… like the hard charging future champion you are” to the end can change (augment) the reinforcing value of following the rule such that following it the first time results in not only any naturally occurring consequences that may reinforce or punish the rule following, but future adherence may also elicit responses related to the reinforcing value of being a champion. Insensitivity in the case of augmentals is a relative change in probability of eliciting a pliant or tracking response when the rule includes an augmental. A person who wants to avoid becoming their parent may engage in a normally low probability response, say engage in tracking of a rule provided by a historically inaccurate provider, if that rule includes a suggestion that the parent would hate such behavior. In this case, the temporary insensitivity is to the history with respect to the source of the rule, but in pliance it could temporarily desensitize the individual to socially mediated consequences or sensitize them to naturally occurring consequences.

This introduction of RGB is not to be interpreted as “all rule following is bad.” RGBs are very efficient and beneficial repertoires for evolving particularly complex behavior within our eusocial society (Hayes & Sanford, 2014). It is only when one or more of these repertoires results in persistent or extreme insensitivity to critical details, that an individual is likely to contact increasingly detrimental consequences and shape maladaptive repertoires to cope with said consequences.
**Evolution, Behavioral Variation, and ACT**

This theory of how language, especially rules, can shape and maintain limited variability, insensitivities, and maladaptive repertoires is actually the underlying theory of ACT itself. ACT is nested within the evolutionary framework of Variation, Selection, and Retention within Domains and across Levels. ACT targets six core processes influencing an individual’s repertoires for engaging in contextually effective behavioral variation with sensitivity toward valued outcomes such that experiences function to select and retain behaviors aligned with said values within domains of import to that individual. The overarching repertoire scaffolded by ACT is called Psychological Flexibility and the core processes are acceptance, defusion, self as context, present moment awareness, values, and committed action. Significant weakness in one or more of these components can result in behaviors that lack an effective degree of variation, demonstrate insensitivity to key experiences, and/or exhibit behavior dominated by immediate reinforcing consequences (impulsivity). Collectively, such repertoires are referred to as Psychological Inflexibility (PF) or Experiential Avoidance. These two terms are used somewhat interchangeably within the literature.

Because ACT emphasizes a set of skills for human flourishing, it has also found success in nonclinical settings such as organizations (Bond & Bunce, 2000) and leadership (Moran, 2011). When applied outside clinical settings, ACT is commonly referred to as Acceptance and Commitment Training.
RFT & ACT Performance Implications

Based on that line of thinking we might improve sport performance by emphasizing an athlete’s efforts to accept and experience the present moment with an emphasis on coming into contact with experiences that orient the individual toward current conditions influencing their ability to move in a valued direction. Within this, normalizing behavioral variation and emphasizing function more so than form (process over outcome) are likely to result in a reduction in frequency of pliant behavior and a concomitant increase in tracking and natural contingency shaping such that the individual’s sport performance comes under significant influence of the conditions of the here and now. This is in contrast to behavior largely under the influence of rules anchored in there and then and thus taking the athlete out of the moment where they can be most effective.

This process for performance improvement has a counter intuitive implication to this model of behavior change. Performance may degrade prior to measurable improvement. In the discussion above regarding RGB, all three elements are brought to high probability in the individual by reinforcing consequences. That means that behaviors we exhibit today are probable for the exact reason that at some point in our history, or our species’ history, those behaviors were repeatedly effective to some degree. When RGB becomes so dominant as to be dysfunctional, aka Rigid Rule Governed Behavior, it is because the individual’s repertoire has been extensively shaped by the reinforcing consequences of that behavior. This can result in a condition called local maximization. In this condition, any alternative responses deviating from behaviors that have
historically resulted in reinforcement are relatively much less probable and have a much higher likelihood of resulting in aversive experiences. Colloquially, it’s all downhill from there. Depending on the severity of the dysfunctional repertoire, the process of performance improvement may include significant declines related to addressing one or more of the core process deficits and shaping enough behavioral variation for new sources of reinforcement to begin to be contacted. The silver lining in this process of change is that same arbitrary applicability of relational responses. Just as a small set of learned relations can result in elaborate networks that transform the stimulus functions across those networks toward dysfunctional outcomes, a similarly small set of learned relations can just as rapidly transform the stimulus function of same or similar elaborate networks just as rapidly toward sustainable outcomes. And this learning can largely be accomplished through language functioning as rules.

**Literature Review**

This has been examined in a few studies on motivation and task persistence using instructional or augmental interventions. In an experiment with college students enrolled in a spin bike exercise program, participants demonstrated significantly increased levels of effort when reminded of why they were exercising compared to their own effort levels with reminders about proper cycling form or when no reminders were provided (Jackson et al., 2016). In two more experiments, participants demonstrated differences in task persistence during an increasingly aversive isometric hold using various control consistent or awareness consistent instructions. When Crossfit athletes had to stand with one arm extended forward and parallel to the ground holding a ~2lb (~0.9 kg) weight
until exhaustion, repeated instructions describing how they are to hold the weight (control of form) or to think of something else (control of thoughts) resulted in significantly shorter times to exhaustion compared to repeated prompts asking “are you willing to continue?” (Leeming, 2016) In a systematic replication of the previous study, endurance runners held a static plank position longest during sessions with a similar repeated willingness prompt. Sessions with a defusion targeted instruction, “sing your thoughts to the tune of a silly song,” resulted in persistence greater than baseline sessions where participants were only instructed to hold the plank position for as long as they could (Encinias, 2021).

In the Jackson and Leeming studies, both included interventions representative of the common control paradigm and in both cases, measured behavior was outperformed by the functional awareness-based interventions. While ACT or MAC interventions show positive results in softball (Little & Simpson, 2000), rowing (García et al., 2004), chess (F. J. J. Ruiz, 2006; F. J. Ruiz & Luciano, 2009, 2012), ice hockey (Lundgren et al., 2020), swimming (Gardner & Moore, 2004; Schwanhausser, 2009), darts throwing (Zhang et al., 2016), golf (Bernier et al., 2009), and Paralympic athletes (Macdougall et al., 2019), these are the only two studies to compare RFT coherent interventions directly against the status quo practices of either PST or control paradigm coaching instructions. For that reason alone, conclusions about the suitability of such interventions as an alternative to the status quo are severely limited.
**Effort, Persistence, & Skill**

Additionally, while sport requires task persistence and peak efforts, much of it also includes elements of skill. To date, a comparison of functional awareness-based instruction to control of form-based instruction on skill-based performances has not been tested. Questions such as, “does an ACT informed functional awareness intervention have a measurable proximal effect on a sport skill task?” and, “does that effect differ from commonly employed control methods for the same task?” require a preparation not yet fully described by the available literature.

A preparation for such investigation would be constrained by a number of expected issues including being able to detect small behavioral effects, being able to take many repeated behavioral measures within an individual and being able to run many participants in a reasonable amount of time. Additionally, if a goal of the investigation is broad generalizability to sport populations, the task must be reasonably sensitive across a wide range of participant skill levels.

For these requirements, golf putting might work. It is a skill-based sport task that at a short distance is reasonably challenging to a wide range of participants. It can be set up in a controlled environment such as a laboratory or office. It has a clearly discriminable hit/miss behavioral outcome as well as misses may be measured for characteristics such as speed, angle, and shortest distance to the center of the cup. A single participant may attempt many putts in a very short period of time. And it allows for experimental designs that can isolate intervention effects from expected temporal confounds such as learning and fatigue.
Such a preparation may be subject to weaknesses including limits to external validity and measurement challenges. The controlled laboratory and contrived task remove much of the environmental stimuli that would otherwise influence a participant’s skilled performance behavior. This limits the possibility of an effect that is functionally tied to participant’s normal golf repertoire as well as the generalizability of any conclusions brought by the data. The preparation’s emphasis on measuring many incidences of behavior over time for each participant constrains the ability to measure any particular instance with precision without introducing invasive and time intensive steps after each instance or investing in the development of automated non-invasive measurement systems that are not currently available.

With these strengths and weaknesses considered, this proposed preparation was explored in a comparison of brief status quo control versus functional awareness instructions using golf putting in a multiple baseline across individuals design in this study.

Method

IRB Approval

The following experiment was conducted under approval from the institutional review board of the University of Nevada Reno.

Participants

Participants were recruited from the University of Nevada and surrounding community as well as by word of mouth and social media. Students of the university
were compensated for their participation via extra credit in their classes. Non-student participants were not compensated for participation. Due to restrictions imposed because of COVID-19, the study design was modified to enable remote participation. Some participants set up, recorded, and uploaded video data without interacting with the researchers. Where possible, the first author shipped or hand delivered sanitized equipment so participants would be using the same putter, green, and balls. Where possible, participants came to a university laboratory space and their participation was overseen by a member of the research team under strict adherence to all university, state, and federally mandated safety measures. Those measures included sanitization, covid symptom prescreening, mandatory use of face coverings, and social distancing.

**Structure**

102 participants completed a single session experiment consisting of a questionnaire packet, baseline putting attempts, an intervention packet, post intervention putting attempts, and a final questionnaire packet. Participants were randomly assigned to one of two intervention groups (Control or Functional Awareness). Within intervention groups each was further randomly assigned one of three possible number of putt attempts (34, 50, or 67 putts). The number of putt attempts applied to both pre and post intervention putting rounds (i.e. 34 attempts baseline and 34 attempts post-intervention). This resulted in a 2x3 intervention by multiple baseline by putt attempts structure (Figure 1). This multiple baseline element was adopted to discriminate effects due to intervention versus expected confounding effects due to time (learning or fatigue). Balanced randomization was done by the survey software Qualtrics.
Questionnaires

Questionnaires, informed consent, and intervention were all delivered, and responses recorded, online with Qualtrics survey software. Participants used their own web capable smartphone or were provided with a tablet to complete all online content. Pre-task questionnaires consisted of Demographic, Golf History, and psychometric surveys. The final questionnaire consisted of the same psychometric surveys. Demographics questions collected age, gender, height, and dominant hand. Golf history questions collected self-reported level of competitiveness (Likert: Avoids Competition (0) - Extremely Competitive (7)), self-reported perception of golf putting ability (Likert: Novice (0) – Professional (7)), time range since their most recent golfing, type of their most recent golf experience (Traditional or Miniature Golf) and a brief description of their experience putting.

Psychometric measures used were the Acceptance and Action Questionnaire Version 2 (AAQ-II: Bond et al., 2011), Mindfulness Attention Awareness Scale (MAAS;
Brown & Ryan, 2003), Cognitive Fusion Questionnaire (CFQ; Gillanders et al., 2014), and the Open and Engaged State Questionnaire (OESQ; Benoy et al., 2019). All four of these measures have been developed under the theoretical framework of Psychological Flexibility.

The AAQ-II is the most widely adopted of the measures and is used as a general measure of psychological inflexibility which may moderate effects of the interventions. While there are variants of the AAQ specific to some sport activities, since this experiment recruited from a more general population, it was decided to stick with the non-specific variant. The measure uses seven Likert response questions on a scale of Never True (1) to Always True (7). The scale is scored by summing the seven items. Higher scores equal greater levels of psychological inflexibility.

The MAAS targets frequency of the participant engaging in the present moment and does not overlap with the construct of Acceptance to the degree of other commonly used mindfulness measures. It was chosen because the interventions in this experiment are theorized to directly influence mindful awareness. The measure uses fifteen Likert response questions on a frequency scale of Almost Always (1) to Almost Never (6) of behavior relevant to mindful attention. The scale is scored by summing the fifteen items. Higher scores represent lower frequencies of the participant engaging in present moment awareness behaviors.

The CFQ targets problematic rule governance. It was chosen because the interventions are theorized to directly influence the dominance of rule following behaviors. The measure uses the same number of questions, structure, and Likert response scales as the AAQ, but the response prompts are specific to struggles with
thoughts, rumination, and over-analyzing. The scale is scored by summing the seven items. Higher scores represent increasingly rigid rule following repertoires.

The OESQ is a 4-item alternative measure of psychological flexibility that utilizes temporally specific and behavior referenced questions and a Not At All (0) to Very Much (10) slider response scale. Psychometric validation of the OESQ with clinical populations demonstrated added predictive utility above and beyond common clinical assessments as well as addressed reports that the AAQ-II may be measuring neuroticism more so than psychological flexibility (Wolgast, 2014). While early work with the OESQ demonstrates that it may replace the AAQ-II as a measure of psychological flexibility, Benoy et al. note a current limitation of the measure is it has only been applied to clinical populations. This limits generalizations of findings with non-clinical participants such as the students and community members in this experiment. For this reason, both the AAQ-II and OESQ measures were used and OESQ data was analyzed for any additional predictive utility of behavioral outcomes.

**Task**

Participants attempted their baseline number of putts on a ~7-foot-long indoor putting green featuring a 2.6-inch-tall and ~15-degree slope ramp leading to the elevated hole. The green was also decorated with a number of white lines. See Figure 2 for details. All putts were attempted from the same tee-point regardless of outcome. After baseline putts, participants read a one-page instructional intervention titled “Focus Points for Putting.” Following intervention, participants completed an attentional check quiz and then repeated the same number of putt attempts as they attempted in baseline. Putting
rounds were recorded via smartphone or tablet such that the entire putting green and participants, waist down, were visible in frame.

Figure 2: Putting Setup Diagram (Top) and During Session (Bottom)

**Intervention**

The five focus points read between putting rounds was the intervention. Control participants read:

Next you need to attempt the same number of putts as you just did. Please incorporate the following five focus points into this round of putting.

- Notice your body. Align your body parallel to the line of the putt. Place your feet roughly shoulder width apart pointing straight toward the putting green. Tilt forward from the hips, keeping your spine in a neutral posture.
Hold the putter so that your hands mirror each other with your trailing hand lower on the grip than your leading hand and your thumbs making a straight line pointing toward the putter head. Grip firm enough that the putter is an extension of your arms but you aren’t driving the blood out of your hands.

- Notice your efforts to control each swing. Be consistent and swing through each putt. Try to keep your swing path as straight to the hole as possible and your putter face contacting the ball perfectly square to the swing path.

- Notice the putting surface. There are lines to mark distance to the hole and the center lane. Focus on keeping your putt between the center lane markers to keep it headed towards the hole. Putting the ball outside the center lane markers is likely to miss.

- Notice the outcome of each putt. Some will go in. Some won’t. Try to make more in more often.

- Regulate your emotions. Keep calm. Getting upset will interfere with your putting and make it harder to complete the session. If you find yourself getting nervous, focus on pushing yourself to be calm instead so you can make the most putts.

Read these as many times as you like before you go on to the next page which is a quiz about this information.
The Control version was derived from popular putting technique instructions found online and included directions such as how the participants should be gripping the club, orienting the club face to the ball, and where the ball should roll on the green. Wording emphasized the importance of not missing, controlling the ball, and controlling their thoughts and feelings.

Functional Awareness participants read:

Next you need to attempt the same number of putts as you just did. Please incorporate the following five focus points into this round of putting.

- Notice your body. Your grip, your stance, the bend of your knees, and how you breathe. You may find you are very relaxed or very tight. These variations in your body are normal responses. Take note of these differences as you putt and see how they influence your putting.

- Notice your efforts to control each swing. You may find that you want to be robotic and precise with each putt. This is a common reaction. By allowing for greater variation of your swing, in the face of urges to control, you are more likely to find your best method of putting through variety. Practice willingness to putt differently each time in order to find many ways that work for you. Take note of how differing motions and timing influence your putting.

- Notice the putting surface. There are lines for distance and guidance, a ramp, and the hole. There may be slight surface imperfections. As you try different ways of putting, notice how the ball travels along the putting
surface in relation to each of the surface elements. You may find yourself focusing more on particular elements of the surface. Take note of how focusing on different elements of the surface influences your putting.

- Notice the outcome of each putt. Some will go in, some will not. Find what you can learn from each putt, including the unsuccessful attempts. This is how putting works.

- Notice your emotions. You may notice emotional responses such as happiness, frustration, or boredom at different times. You may notice a desire to speed up or go slow. Acknowledge those emotional responses as you would a neighbor or a classmate. Say hello, notice they are there, and then come back to the putting task at hand. Take note of how acknowledging your emotional reactions influences your putting.

Read these as many times as you like before you go on to the next page which is a quiz about this information.

The Functional Awareness version was derived from the psychological flexibility processes of openness to experience, awareness of the present moment, and defusion from judgements and frustrations that may arise in the course of putting. The text prompted participants to observe how experiences and putt outcomes interacted, no instructions specified a formal technique or encouraged participants to engage in any effort to control the ball or their thoughts and feelings.

Participants completed a four-question attentional check asking how many putts they would need to attempt in the upcoming round, what five things they should notice,
what they should be willing to do while putting, and what impressions they had about the focus points. Incorrect responses to the first three questions were used to screen out the data of participants that did not sufficiently attend to the intervention.

**Putting Data Handling & Coding & Reliability**

Video recordings of putting rounds were uploaded to cloud storage by participants, or the observing researcher in the case of sessions conducted in person. Full length recordings were then separated into 4 second clips of individual putt attempts by PS and saved using a unique file name that was a concatenation of the anonymous participant ID number, round number, and putt attempt number. This resulted in 9,458 individual putt attempt clips.

PS and two trained undergraduate research assistants watched and coded the individual putt instances remotely using personal computers. Due to the sheer number of putt instances, each instance was watched and coded by one of the above mentioned three.

Coding of putt outcomes was completed using a categorical scoring method that evaluated speed (too fast, too slow, likely to go into the hole) and angle (left, right, or aligned with the hole) of the ball relative to the hole. This method results in 9 possible outcomes arrayed in a 3x3 matrix where the center combination (likely to go in and in line with the hole) was synonymous with a make (or hit). Seen in the image of the actual putting setup in Figure 1, the hole was surrounded on three sides by a raised boundary and gutter in order to capture and return balls. On rare occasions, a ball would rebound off this boundary and land in the hole. Because these were not a-priori excluded, were a
possible skilled and effective outcome given the current context and were theorized to have possible reinforcing effects on the participant, they were coded as a make.

A random sample of at least 10% of coded sessions were rated by at least one other rater for inter rater reliability. In the case where 10% of one rater’s sessions coded was not a whole number, the number of sessions to be reviewed by a second rater was rounded up. This resulted in 21 of 186 coded sessions, 1407 of 9458 putt instances (~15%), being sampled for reliability.

Inter-rater reliability was calculated using trial-by-trial total agreements divided by the sum of agreements and disagreements. Percentage agreement ranged from 88% to 100% based on the complexity of the discrimination. All raters agreed 100% on if the ball went into the hole or not. For left/right/inline discrimination, agreement dropped to 95%. When the speed discrimination was added in, agreement was the most varied at 88%. The sources of agreement variation were dominated by putts that were literally and figuratively edge cases. For balls that interacted with the left or right edge of the cup, colloquially referred to as a ‘lip out’, coders were more likely to disagree on if it was over the hole or missed to the side, as well as if the speed was too fast or just right. Early attempts at coding were unusable due to this issue impacting inter-rater reliability. Additional training was implemented using putts from participants who failed the attentional check. Once coders successfully coded multiple sets of hard to distinguish putts with greater than 90% agreement, they re-coded the main data set with the above IRR results.
Results

Participants

102 participants completed the experiment. 9 were excluded for failing the attentional check. 43 identified as Female and 49 as Male. Age was bimodally distributed with modes at 18 and 65. Mean height was 67.5 inches (sd=3.9). Handedness was primarily right-handed (80) with 8 participants reporting left hand dominance. Time since their last golfing experience was bimodal with modes at “1-2 months” (n=17) and “Greater than 1 year” (n=35) 50 participants reported that last session being miniature golf. Self-reported competitiveness (M=3.9, sd=1.9) was skewed toward more competitive with only 31% reporting within the lower half of the options though only 5 participants selected the “Extremely Competitive” option. Self-reported ability level (M=3.0, sd=1.4) was more evenly distributed among the middle options of the scale. Only two participants reported “never putted before” (option 0) or near-professional ability (option 6) and zero reported “Professional golfer” (option 7). Psychometric baseline measures across all four scales were normally distributed and means for each scale fell well within non-clinical ranges. Table 1 details participant demographic and baseline psychometric scores.
<table>
<thead>
<tr>
<th>Demographics</th>
<th>Age (median [IQR])</th>
<th>30.50 [18.00, 56.25]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>43 (46.2)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49 (52.7)</td>
<td></td>
</tr>
<tr>
<td>LGBTQ+</td>
<td>1 (1.1)</td>
<td></td>
</tr>
<tr>
<td>Height (inches) (median [IQR])</td>
<td>68.00 [64.00, 70.00]</td>
<td></td>
</tr>
<tr>
<td>Competitiveness (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoid Competition (8)</td>
<td>6 (6.5)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11 (11.8)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7 (7.3)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5 (5.4)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>24 (25.8)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>21 (22.6)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>14 (15.1)</td>
<td></td>
</tr>
<tr>
<td>Extremely Competitive (7)</td>
<td>5 (5.4)</td>
<td></td>
</tr>
<tr>
<td>Ability (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Putted Before (8)</td>
<td>2 (2.2)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13 (14.0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>22 (23.7)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17 (18.3)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25 (26.9)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12 (12.9)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Professional Golfer (7)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Handedness (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>80 (86.0)</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>8 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>5 (5.4)</td>
<td></td>
</tr>
<tr>
<td>Time Since Last Golf (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-7 Days</td>
<td>10 (10.8)</td>
<td></td>
</tr>
<tr>
<td>1-2 Weeks</td>
<td>1 (1.1)</td>
<td></td>
</tr>
<tr>
<td>2-4 Weeks</td>
<td>8 (8.6)</td>
<td></td>
</tr>
<tr>
<td>1-2 Months</td>
<td>17 (18.3)</td>
<td></td>
</tr>
<tr>
<td>2-3 Months</td>
<td>3 (3.2)</td>
<td></td>
</tr>
<tr>
<td>3-6 Months</td>
<td>2 (2.2)</td>
<td></td>
</tr>
<tr>
<td>6-12 Months</td>
<td>17 (18.3)</td>
<td></td>
</tr>
<tr>
<td>&gt;12 months</td>
<td>35 (37.6)</td>
<td></td>
</tr>
<tr>
<td>Last Golf Type (Count [%])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golf</td>
<td>43 (46.2)</td>
<td></td>
</tr>
<tr>
<td>Miniature Golf</td>
<td>50 (53.8)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: Participant Descriptive Statistics**

*Non-normally distributed described by Mean and Interquartile Range

| n = 93 |
|-----------------|-----------------|-----------------|-----------------|
| *Non-normalled described by Mean and Interquartile Range* |

**Pre-Experimental Checks**

Intervention groups were checked for pre-experimental differences across demographic, psychometrics, golf specific self-report variables, and individual baseline...
accuracy. Table 2 breaks out demographic, golf specific self-report, and baseline psychometric data across intervention group assignment and gender. With only one participant identifying as other than CIS gendered, their data was excluded from analysis across gender groups. Significant gender differences were noted for age, height, ability, and last golf type. Participants identifying as female were younger, shorter, reported their ability as lower, and were more likely to have last played miniature golf. Despite these differences, gender distribution was not significantly different across intervention group assignments or putt attempts assignments. All other variables tested as not significantly different between the two intervention groups.
Table 2: Participant Descriptive Statistics by Various Groupings

*Non-normally distributed described by Mean and Interquartile Range

<table>
<thead>
<tr>
<th>n</th>
<th>45</th>
<th>48</th>
<th>43</th>
<th>40</th>
<th>43</th>
<th>40</th>
<th>0.231</th>
<th>Female</th>
<th>Male</th>
<th>0.072</th>
<th>0.972</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (median[IQR])</td>
<td>36.50[18.75, 53.75]</td>
<td>27.50[18.00, 53.75]</td>
<td>46.00[18.00, 60.00]</td>
<td>36.00[20.00, 60.00]</td>
<td>0.231</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>21 (46.7)</td>
<td>22 (45.8)</td>
<td>23 (51.1)</td>
<td>26 (54.2)</td>
<td>0.574</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24 (53.3)</td>
<td>26 (54.2)</td>
<td>26 (54.2)</td>
<td>32 (65.8)</td>
<td>0.072</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGBTQIA+</td>
<td>1 (2.2)</td>
<td>0 (0.0)</td>
<td>1 (2.2)</td>
<td>0 (0.0)</td>
<td>0.806</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (inches) (median[IQR]):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.00[46.00, 70.00]</td>
<td>68.00[63.00, 71.00]</td>
<td>68.50[63.88, 71.00]</td>
<td>68.50[63.00, 71.00]</td>
<td>0.480</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitiveness (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoid Competition (0)</td>
<td>4 (8.9)</td>
<td>2 (4.2)</td>
<td>2 (4.7)</td>
<td>4 (8.2)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 (11.1)</td>
<td>6 (12.5)</td>
<td>7 (16.3)</td>
<td>4 (8.2)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 (6.7)</td>
<td>4 (8.3)</td>
<td>6 (14.0)</td>
<td>1 (2.0)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 (2.2)</td>
<td>4 (8.3)</td>
<td>3 (7.0)</td>
<td>2 (4.1)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15 (33.3)</td>
<td>9 (18.8)</td>
<td>11 (25.6)</td>
<td>12 (24.5)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8 (17.8)</td>
<td>13 (27.1)</td>
<td>10 (23.3)</td>
<td>11 (22.4)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7 (15.6)</td>
<td>7 (14.6)</td>
<td>3 (7.0)</td>
<td>11 (22.4)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely Competitive (7)</td>
<td>2 (4.4)</td>
<td>3 (6.2)</td>
<td>1 (2.3)</td>
<td>4 (8.2)</td>
<td>0.999</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Putted Before (0)</td>
<td>0 (0.0)</td>
<td>2 (4.2)</td>
<td>1 (2.3)</td>
<td>1 (2.0)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8 (17.8)</td>
<td>5 (10.4)</td>
<td>9 (20.8)</td>
<td>4 (8.2)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9 (20.0)</td>
<td>13 (27.1)</td>
<td>13 (28.2)</td>
<td>9 (18.4)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10 (22.2)</td>
<td>7 (14.6)</td>
<td>8 (18.6)</td>
<td>8 (16.3)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9 (20.0)</td>
<td>18 (33.3)</td>
<td>11 (25.6)</td>
<td>14 (28.6)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8 (17.8)</td>
<td>4 (8.3)</td>
<td>1 (2.3)</td>
<td>11 (22.4)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 (2.2)</td>
<td>1 (2.1)</td>
<td>0 (0.0)</td>
<td>2 (4.1)</td>
<td>0.307</td>
<td>0.843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Golfer (7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.349</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handedness (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>40 (88.9)</td>
<td>40 (83.3)</td>
<td>36 (83.7)</td>
<td>43 (87.8)</td>
<td>0.349</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>2 (4.4)</td>
<td>6 (12.5)</td>
<td>4 (9.3)</td>
<td>4 (8.2)</td>
<td>0.349</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>3 (6.7)</td>
<td>2 (4.2)</td>
<td>3 (7.0)</td>
<td>2 (4.1)</td>
<td>0.349</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Since Last Golf (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-7 Days</td>
<td>7 (15.6)</td>
<td>3 (6.2)</td>
<td>1 (2.3)</td>
<td>9 (18.4)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 Weeks</td>
<td>1 (2.2)</td>
<td>0 (0.0)</td>
<td>1 (2.3)</td>
<td>0 (0.0)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4 Weeks</td>
<td>3 (6.7)</td>
<td>5 (10.4)</td>
<td>3 (7.0)</td>
<td>5 (10.2)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 Months</td>
<td>8 (17.8)</td>
<td>9 (18.8)</td>
<td>8 (18.6)</td>
<td>8 (16.3)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-3 Months</td>
<td>1 (2.2)</td>
<td>2 (4.2)</td>
<td>2 (4.7)</td>
<td>1 (2.0)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 Months</td>
<td>2 (4.4)</td>
<td>0 (0.0)</td>
<td>1 (2.3)</td>
<td>1 (2.0)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-12 Months</td>
<td>7 (15.6)</td>
<td>10 (20.8)</td>
<td>8 (18.6)</td>
<td>9 (18.4)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;12 months</td>
<td>16 (35.6)</td>
<td>19 (39.6)</td>
<td>19 (44.2)</td>
<td>16 (32.7)</td>
<td>0.518</td>
<td>0.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last Golf Type (Count[%])</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golf</td>
<td>21 (46.7)</td>
<td>22 (45.8)</td>
<td>11 (25.6)</td>
<td>32 (65.8)</td>
<td>1.000</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miniature Golf</td>
<td>24 (53.3)</td>
<td>26 (54.2)</td>
<td>32 (74.4)</td>
<td>17 (34.7)</td>
<td>1.000</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Baseline accuracy combined across putting assignments approached significance differences between intervention groups (p=0.052, df=89.7) but further exploration within each subgroup of putt assignments (34, 50, & 67) highlighted that only the 34 putt groups approached significance (p=0.07, df=28.8) and the latter two failed to reject the null hypothesis of equivalent accuracy with p-values exceeding 0.45. Given these results, no adjustments were made before additional inferential statistics were run on the data set.

Psychometric scores were tested for global and within intervention group effects between baseline and post intervention. Table 3 breaks out those measures. No significant shifts were measured.

<table>
<thead>
<tr>
<th>Table 2: Participant Descriptive Statistics by Various Groupings (Cont)</th>
<th>Intervention Assignment</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychometrics AAQ-II (mean (SD))</td>
<td>16.56 (6.69)</td>
<td>17.08 (7.21)</td>
</tr>
<tr>
<td>Psychometrics MAAS (mean (SD))</td>
<td>64.27 (11.05)</td>
<td>64.62 (11.14)</td>
</tr>
<tr>
<td>Psychometrics CFQ (mean (SD))</td>
<td>20.24 (9.45)</td>
<td>20.81 (9.20)</td>
</tr>
<tr>
<td>Psychometrics OESQ (mean (SD))</td>
<td>13.56 (6.83)</td>
<td>14.15 (8.51)</td>
</tr>
<tr>
<td>Assignment</td>
<td>Putt Attempts (Count(%))</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>15 (33.3)</td>
<td>17 (35.4)</td>
</tr>
<tr>
<td>50</td>
<td>14 (31.1)</td>
<td>16 (35.3)</td>
</tr>
<tr>
<td>67</td>
<td>16 (35.6)</td>
<td>15 (31.2)</td>
</tr>
<tr>
<td>Intervention Assignment (Count(%))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>21 (48.8)</td>
<td>25 (46.9)</td>
</tr>
<tr>
<td>b</td>
<td>22 (51.2)</td>
<td>26 (53.1)</td>
</tr>
<tr>
<td>Table 3: Participant Psychometric Measures Within Group Across Time</td>
<td>Round</td>
<td>Within FUNCTIONAL AWARENESS Group Across Rounds</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>n</td>
<td>Baseline</td>
<td>Post</td>
</tr>
<tr>
<td>Psychometrics AAQ-II (mean (SD))</td>
<td>16.83 (6.69)</td>
<td>16.33 (7.37)</td>
</tr>
<tr>
<td>Psychometrics MAAS (mean (SD))</td>
<td>65.42 (11.07)</td>
<td>64.64 (12.32)</td>
</tr>
<tr>
<td>Psychometrics CFQ (mean (SD))</td>
<td>20.54 (8.35)</td>
<td>19.32 (7.96)</td>
</tr>
<tr>
<td>Psychometrics OESQ (mean (SD))</td>
<td>13.86 (7.71)</td>
<td>13.65 (7.76)</td>
</tr>
</tbody>
</table>
During additional exploratory analyses of effective variability, baseline levels were compared across intervention groups in the same way as accuracy was tested above. Results did not indicate any significant pre-experimental differences.

**Primary Intervention Effects Analysis**

**Accuracy Within and Between Subjects Over Time**

The strongest evidence of an intervention effect from either Control or Functional Awareness instructions would be a significant shift in the frequency of made putts by any one individual repeated across individuals that correlated with each participant experiencing the intervention. The multiple baseline design helps clarify intervention effects by employing the logic of unlikely coincidences. As intervention timing is varied between participants, consistent change in the dependent variable is increasingly unlikely to be a product of a confounding coincidence across all the participants. This does not control for all confounds, such as selection bias, but in this case can provide insight into if measured effects are due to intervention versus fatigue or immediate learning history.

**Visual Analysis**

Accuracy data within subjects first went through a visual analysis for possible large effects. Large effects within subjects would exhibit significant slope changes occurring at the point of intervention on a cumulative record. Figure 3-5 below are the cumulative records for all participants grouped by putt and intervention assignment. Each putt assignment group also includes a combined cumulative record for each intervention
group (Right Y-Axis). Line type indicates intervention group. Dashed participants were assigned the Functional Awareness intervention.

On a cumulative record representation, an accuracy improvement would exhibit as a steeper curve after the intervention and accuracy declines would exhibit as a flattening curve after the intervention. Across putt assignment groups and interventions, no consistent visible shift was noticeable at either the within individual or within group level. Said more simply, neither coaching intervention had an observable impact.

Figure 3: Cumulative Record of Hits 34 Putt Attempt Group.
Figure 4: Cumulative Record of Hits 50 Putt Attempt Group.

Figure 5: Cumulative Record of Hits 67 Putt Attempt Group.
**Statistical Analysis**

Within-subject between groups statistical analysis was conducted using a mixed effects linear regression model with random intercepts and slopes for each participant comparing their accuracy across pre and post intervention rounds. R statistical software (R Core Team, 2020; RStudio Team, 2019) was used running the nlme package (Pinheiro et al., 2021) for mixed effects modelling. A copy of all anonymized coded data and R code is available upon request.

Data was truncated for this method to sample from the putt attempts within 34 attempts of the intervention point. That meant that data from participants assigned to 34 attempts were wholly included, but participants in the 50 and 67 putt attempt groups had their first and last results removed from their accuracy calculations in order to standardize the number of observations per participant per round. The logic used to make this decision was that the multiple baseline design was used to discriminate between intervention, fatigue, and learning effects occurring within the experimental participation process. Once data was collected in this way, other than for visual analysis, truncating the data such that pre and post had the same number of observations would not invalidate the logic that across the intervention point, any statistically significant effect would be discriminated as more likely due to intervention than fatigue or learning.

Analysis was run on three variants of the linear mixed effects model. A simple model where only random effects (participants random slope and intercept) were modeled. A 3x2x2 intervention assignment by putt assignment by round fixed effect with the previously described random effect. And a final model with the addition of one of the psychometric measures taken before putting was included. That final model was run both
as multiplicative and additive fixed effects variants as well as with all four different psychometric measures (AAQ-II, CFQ, MAAS, & OESQ) singularly to check if there were any significant model differences between measures.

Comparison between the simple and 3x2x2 models suggested that pre-experimental variables by participants explained over 90% of the variance in accuracy between rounds and p-values never crossed the predetermined p<0.05 significance threshold. In addition, once any of the psychometric measures were included in either additive (intercept) or multiplicative (slope) ways, variance explained by fixed effects decreased even further and any p-values that approached significance in the 3x2x2 model only moved away from the significance threshold.

Given these results, we cannot reject the null hypothesis that differences in accuracy measured within participants across rounds and intervention and putt assignment groups are not different. Said more simply, in confirmation of the visual analysis, neither coaching intervention had a statistically significant impact.

**Exploratory Analysis**

*Variation within subject over time*

The most effective behavior is not always the same behavior. Through an evolutionary lens of selection by consequences, some degree of variation of behavior is more likely to result in consequences contributing to long term success of the species and individual. In the same way, the shaping of a new operant behavior, say a skilled performance, is efficiently accomplished through quick successive contact with consequences of somewhat similar behaviors. Through the lens of learning a skilled
performance, there is minimal value in doing the same unsuccessful thing over and over. Any behavior that results in a novel version of failure is relatively more valuable. Even when success is achieved, there is incremental value in continuing to engage in behavioral variation. Clarifying the boundary between success and failure of stimuli influencing success results in behaving more responsive to dynamic conditions and therefore more likely to be robustly effective relative to a behavior shaped by more limited variation.

This interpretation of behavioral variability suggests that a precursor indicator of intervention effects may be engaging in more effective variation. That is, while putting the same miss over and over is not particularly informative after the first time, putting a variety of misses can bring behavior under the control of variables that in the long-term lead to more hits. This pattern of contextually effective variation was not expected to occur when participants were instructed to focus on control efforts and was hypothesized to possibly even decline under the Control condition.

Effective variation as a behavior of interest has not previously been operationalized in the behavior analysis literature so a number of subjective assumptions and choices had to be made in order to subject it to an exploratory analysis in this data set. Those are as follows:

1. Raw variation does not represent effective variation because it lacks a contextual benchmark to evaluate relative efficacy.
2. In the context of putting, the long-term goal is all hits.
3. Any variation of putt outcomes in temporal sequence, or lack of variation thereof, that results in a hit is relatively the most effective for that context.
4. Misses following a miss that were different from the last were relatively less effective than a hit but more effective than repeating the last miss exactly.

Using those four assumptions, an arbitrary scoring system was implemented where any Hit earned a full point, any Different Misses earned a half point, and any Same Misses earned zero points. This scoring system produces a possibilities boundary where 100% hits OR 50% hits alternating between hits and misses maximizes the Effective Variability score (100%) and perfect repetition of the same miss results in a zero score. Any deviation from those three extreme cases would result in an Effective Variability score >0% but <100%. This was the scoring system implemented with the data set.

**Statistical Analysis**

This scoring system inherits accuracy as an inbuilt confounding variable.

Statistical and visual analysis should not be run directly on this measure.

Through collaboration with Dr. Markus Kemmelmeier, we were able to eliminate the accuracy confound by regressing a participant’s Effective Variability score to their Accuracy measures over the same putts and then storing the residual values that remained as an Effective Variability Residual. The Effective Variability Residual was calculated for each participant for both Baseline and Post-Intervention rounds.

Within-subject between groups statistical analysis was conducted using a mixed effects linear regression model with random intercepts and slopes for each participant comparing their Effective Variability Residual across pre and post intervention rounds. This analysis was identical to the analysis using Accuracy as the dependent variable.
described above and only changed the dependent variable to this exploratory measure of variability. This includes the multiple model generation and comparison steps.

Consistent with the accuracy analysis results, changes in Effective Variability Residual within individuals were largely accounted for by pre-experimental differences and no intervention or intervention related interaction element of any of the models crossed the p<0.05 level of significance. Overall variation within participants over time attributed to fixed effects in these models was less than 10% of the entire model. In harmony with all analytic methods previously employed there was no evidence of an impact on behavioral variation for either intervention.

**Discussion**

**Intervention Effects**

This study aimed to use treatment as usual as a control condition and draw conclusions as to the relative efficacy of Control vs Functional Awareness interventions. From the results, one might be drawn to the conclusion that since neither intervention produced a different result, then Functional Awareness may be just as good as the status quo. This is a logical but misguided interpretation of the data. The linear mixed effects models estimated that fixed effects, which included intervention effects, accounted for no more than 5.5% of Accuracy change and 3.7% of Effective Variability changes. These small amounts of possible effects are then spread across multiple elements including, most of all, participant learning history (time). This would suggest that neither intervention had any measurable effect and as such we cannot compare the two when we did not detect either one in this study.
The choice to use a single page written instructional intervention was a strategy to be able to conduct a well powered experiment at scale with high internal consistency across participants while validating a novel use of the golf putting procedure. Results of both visual and statistical analyses of ultimate (accuracy) and theorized proximate (effective variability) measures of behavior suggest that that strategic choice may have reduced the efficacy of either intervention to near zero. While the procedure itself enabled a high volume of participants and captured the outcome of many instances of behavior over time, the intervention will require significant changes in future iterations.

The fact that the status quo Control of Form instructions did not impact participant performance to a detectable level was surprising considering that near identical instructions emphasizing control of both form and emotional experience are widespread within the golf literature. It is similarly interesting to note that qualitative responses to the intervention focus points were positive evaluations regardless of intervention assignment. No participants expressed negative opinions of the focus points or otherwise think the instructions would negatively impact their putting outcomes.

In a similar parallel, the golf history questionnaire asked participants to self-report their skill level. When that was used to visualize their first-round performances broken out by skill level, there was no clear indication that self-report was predictive of first round performance or within subject changes across rounds. Similarly, time since last golf experience and last golf type were not correlated with baseline performances or within subject changes of any tested dependent variables across rounds.

These interesting asides may not be significant revelations. The task was conducted away from most of the environmental cues that would normally co-occur with
golf and mini golf activities. It is entirely likely that much of the ancillary stimulus influence that would support subjects’ normal golf putting repertoires were not in effect. Along that logic, a reasonable argument could be made that this laboratory preparation is only a vague analog to actual golf putting and if future variants of this procedure do produce observable intervention effects, they should be validated across increasingly naturalistic settings as well.

**Study Limitations & Strengths**

The study inherited limitations due to mid-stream procedural changes due to COVID-19. The original study was two sessions per participant run a week apart from each other. It was in a controlled lab space with a trained observer. All participants would use the same equipment and putting green and participants were expected to be all college students. Final approval for that original procedure occurred days before the US instituted lock down. Students dispersed. University activity stopped. It was weeks before COVID guidance was provided by the university and months before the final version of this study was approved. This compressed sessions to one, spread participation across remote and in person modalities, diversified the participant population, and environments and equipment became much more varied. It is possible this reduced procedural fidelity introduced confounds. That possibility can only be tested by repeating the experiment under more controlled conditions.

The experiment also had strengths. No participant, regardless of skill level, completely maximized or minimized the task. Only one participant managed to make every putt into the hole for one of their two rounds and no participants missed all of their
putts in any round. The seemingly simple task of a short putt to an elevated hole has a broad range of sensitivity. It may detect performance changes in participants ranging from highly skilled to relative novices. This sensitivity may be improved with reliable and precise measurement of the ball’s travel. This study attempted to address that imprecision. The computer vision solution pursued required additional investigation before it can be considered a reliable alternative to the more imprecise human coding. That solution is discussed below in more detail.

**Inferential Limitations**

This study employed a number of methods in an attempt to infer intervention effects as well as discriminate those from likely confounds such as learning and fatigue. Visual analysis in combination with the multiple baseline design, multinomial categorical outcomes coding, and inferential statistics were used but each also introduced limitations.

Visual analysis relies on visual discrimination of an effect that correlated to the intervention time point. This requires characteristics of the data including a large effect, accurate choice of behaviors measured, use of a measurement tool sensitive enough to quantify possible effects, and measurement enabling discrimination between effect variation and artifacts due to the tool. In this case, one or more of these requirements may be violated. Most likely, and discussed previously, is the interventions did not perturb participants’ behavior enough. There was no large effect of the intervention.

Next most likely is the multinomial method of categorizing putt outcomes. This was based on the trained, though still subjective, assessment of speed and angle of the ball by human observers. Very likely, this introduced artifacts into the data such that fine
variation within participants’ behavior was masked by the categorical method of measurement. Critical variability may have gone unmeasured simply because the measurement tool had low precision. A post-hoc technological solution was pursued in the form of a computer vision algorithm. This algorithm applied ball identification and tracking to locate the ball in each video frame. It then applied homography to measure the location of the ball based on known dimensions of the putting surface landmarks. In this way, it was hoped that all putts could be coded to a much finer degree of precision, possibly as accurate to a few millimeters and degrees. While our collaborators in computer science produced a working algorithm, of the 93 participant video sets, less than 25 were able to be coded for 100% of attempted putts. More than half of the participants’ data sets failed to be measured even to 50% of attempted putts. While this may be a viable method of precise measurements with a large number of participants in future iterations of this research, it did not provide that solution for this particular data set. Anecdotally, the failure to measure most of the putt instances may be due to the variation introduced by the COVID adjustments discussed above. All of these factors may be identified with a systematic review of failure modes within this data set and addressed in future work.

Of the three requirements for successful visual analysis, the selection of behavior, in this case the behavioral product of putt outcome, is the least likely to have been at issue here. The categorical nature of the make/miss discrimination easily translates into a cumulative record graph, the traditional behavior analytic tool for visual analysis. Traditional cumulative records are plotted as count on the y-axis and time on the x-axis. In this case, time was only approximately coded in this data set via putt attempt number.
Because of this, identifying temporal patterns, such as a participant increasing or decreasing their rate of putts over time, is confounded by the increment of time between putts not being represented on the x-axis. An in-depth time coding of the data set may allow such analysis, but due to the otherwise null results discussed above, this was not prioritized. Relatedly, the fact that the data set is continuously video recorded in an anonymized format allows for revisiting related questions without the need to run a new cohort.

Statistical analysis inherits additional potential weaknesses. Model selection, measure selection, variance within and between participants and groups all impact the strength and validity of inferential conclusions. Measure selection applies here in much the same way as discussed above. The major difference is that inferential methods can isolate and magnify critical variation that is visually undetected. Model selection becomes critical because while data can be transformed for different models, transformation can magnify confounding variance, strip critical variance, or both at the same time. For example, early exploratory analysis of Effective Variability used a raw effective variability score for each participant in the first half of their baseline putts to correlate to their accuracy measure in the second half of their baseline putts. This was done to explore if one measure (Eff Var) had utility in predicting the outcome of the other. In this correlation, an r-squared value of 0.5 was computed suggesting that it may have predictive utility. To further investigate this pattern, baseline round effective variability was correlated to post intervention accuracy within participants and the correlation flipped ($r^2=−0.4$). This contradicted the within round predictive correlation. Upon review with Dr. Kemmelmeier, he quickly identified the accuracy confound in the
measure. In this case, transformation from raw data introduced unintended confounds that
were eventually detected and controlled for. That detection and adaptation doesn’t always
occur.

In the primary statistical analysis, something similar may have resulted in lost
statistical variance. While the participant data was taken as many individual putts over
time, in order to run the 3x2x2 linear mixed effects model, some data were discarded and
the remaining data was transformed to one pre and one post value for accuracy, or
effective variability residual per participant. This cut the number of observations from 34
per round per participant, to 1. This may have stripped variance that otherwise could have
indicated a detectable effect. Much like the other design decisions, this data
transformation was a strategic choice due to the data being in a categorical form at the per
putt level. Future analysis of this, or similar data, may utilize tools such as the
AutoRegressive Integrated Moving Average (ARIMA) or Bayesian methods to analyze
categorical variance within subjects over time. Additional methodologies that arose
during analysis include One Hot Encoding for machine learning forecasting as well as
Frequency Analysis (also known as Laplace Transforms). Both handle multinominal
categorical data well but are not commonly used with causal inference. While current
documentation for these methods is limited in this use case, they may be viable avenues
of reanalysis.

Related to the variance of data, the variability within and between subjects was a
topic of concern for multiple individuals briefed on the project. Specifically, with respect
to the range of participants’ baseline skill at putting. Consultation with Dr. Mark
Guadagnoli, associate dean of the UNLV School of Medicine and longtime researcher
into learning and performance with golf, brought up his concern that the variability introduced into the data by novice putters is likely to overwhelm that of the more skilled putters. If the participants were of roughly uniform high or low skill level, the study may be powerful enough to detect an intervention effect but commingling of levels may underpower the analysis for both skill groups and result in a null statistical finding. This cannot be ruled out and further analysis of the data set may be warranted to clarify this suggestion. Within the visual analysis, evidence of this may appear with groupings within intervention groups where low or high skill participants consistently have a visibly steeper post intervention graph. In the above graphs, this would produce solid, or dashed, lines crossing or diverging each other post-intervention as well as diverging from their pre-intervention trends and yet across putt groups, most of the post intervention data show largely parallel trends continuing along the pre-intervention trend.

**Questions to Follow Up On**

The outcome of this experiment leaves mostly unanswered questions. Chief among those is if a more precise measure of the ball location would surface an effect and if so, what would it be? Additionally, would recruiting from a more limited population of high skill golfers result in a more nuanced outcome? Does the effective variability construct provide any utility as a proximate measure of intervention effect? And do missed putts have functionally different outcomes relative to the distance of the miss? All of these questions ostensibly rely on that more precise measure. Further investigation may prioritize the refinement of that video homography algorithm such that many participants can attempt many putts quickly without incurring the high human cost of
training, data cleaning, data coding, and interrater reliability monitoring that rapidly consumed resources after data collection was completed in this iteration of the experiment.
References


https://doi.org/10.1901/jeab.1989.51-385


https://doi.org/10.18761/PAC.2021.v12.RFT.02

https://doi.org/10.1002/jeab.64


https://doi.org/10.1080/02640414.2016.1254809


https://doi.org/10.1080/10413208908406413


U.S. Department of Commerce.


https://doi.org/10.1080/00336297.2006.10491876

https://doi.org/10.1016/j.beth.2014.07.002

