Understanding the Yerkes-Dodson law of arousal, affect, and workload when performing a motor sequencing task

Jakob J. Smoor (s2889307)

Bachelorthesis Human Factors and Engineering Psychology (202000383)

Supervisor: Russell W. Chan

Second Supervisor: Simone Borsci

26.06.2025

Table of Contents

Abstract

The Yerkes-Dodson law was initially proposed by Yerkes and Dodson (1908) and predicts a quadratic relationship between arousal and performance, where moderate arousal facilitates performance while insufficient and excessive arousal diminish it. Although over a century of research in motor sequence learning has yielded mixed results, the theory is widely known among psychologists (Winton, 1987). This thesis evaluates the theory with more naturalistic observations by adopting the Dance Step-Sequence Production (DS-DSP) task, a dance motor sequence learning paradigm developed by Chan et al. (2025). We recruited 16 healthy individuals to perform the experiment and combined self-report data with reaction time data to determine the functional relationships that arousal, pleasure, and mental workload share with performance. The results of linear and quadratic regression analyses provided minimal evidence for the theory in the experiment. Only frustration, a subdimension of mental workload, was significantly related to reaction time in a quadratic function ($p = 0.03$), and only temporal demand significantly affected reaction time in a linear function ($p = 0.03$). These findings suggest that psychologists should approach theories and frameworks with caution and emphasize the need for more replication studies. We discuss potential explanations for these findings and provide directions for future work.

1. Introduction

From tying shoelaces to performing complex surgery, every behavioural sequence must be learned before we can produce it efficiently. However, through the process of motor sequence learning (MSL), people can eventually perform these sequences quickly, accurately, and with minimal cognitive effort (Abrahamse et al., 2013; Verwey, 2023). The study of MSL is relevant because it relates to both routine behaviours and the acquisition of complex skills. Researchers aim to optimize learning and performance during motor sequencing tasks by examining the potential effects of emotional and cognitive constructs such as arousal, pleasure, and workload. Nevertheless, the empirical findings remain mixed, and the exact relationships remain difficult to quantify.

A prominent theory on the relationship between arousal and performance is the Yerkes-Dodson law, which proposes a quadratic relationship where both too much and too little arousal impair performance, whereas moderate arousal enhances it (Yerkes & Dodson, 1908). In the present thesis, we test this theory during a dance motor sequencing task and manipulate arousal through varying sequence lengths. In doing so, we assess whether increasing nominal task difficulty, the inherent fixed challenge of a task, induces arousal that affects reaction time (RT) performance in a quadratic pattern and attempt to identify the optimal level of difficulty for learning within our experimental context.

1.1 The Relationship between Arousal and Performance

Arousal refers to the psychological and physiological state of alertness or activation (Ashcraft & Radvansky, 2013). Over a century of research has investigated the potential relationship between arousal and performance, but the concrete relationship remains difficult to quantify due to intervening effects of task characteristics, experience levels, and individual differences. Additionally, arousal is a multidimensional construct that involves physiological measures such as heart rate, skin conductance, and pupil size (Critchley et al., 2013), that may not always align with the individual’s subjective experience (Mauss & Robinson, 2008; Wang et al., 2018). For example, someone might interpret an increase in heart rate or skin conductance as the result of caffeine intake or physical activity instead. Although researchers are limited in their ability to isolate arousal due to its interactions with other physiological, cognitive, and emotional constructs (Staal, 2004), it is generally assumed that arousal can influence performance through a reallocation of attention and focus (Bond et al., 2023;

Mather et al., 2015) or through an effect on the individual's preparedness of the upcoming task (Petersen & Posner, 2012).

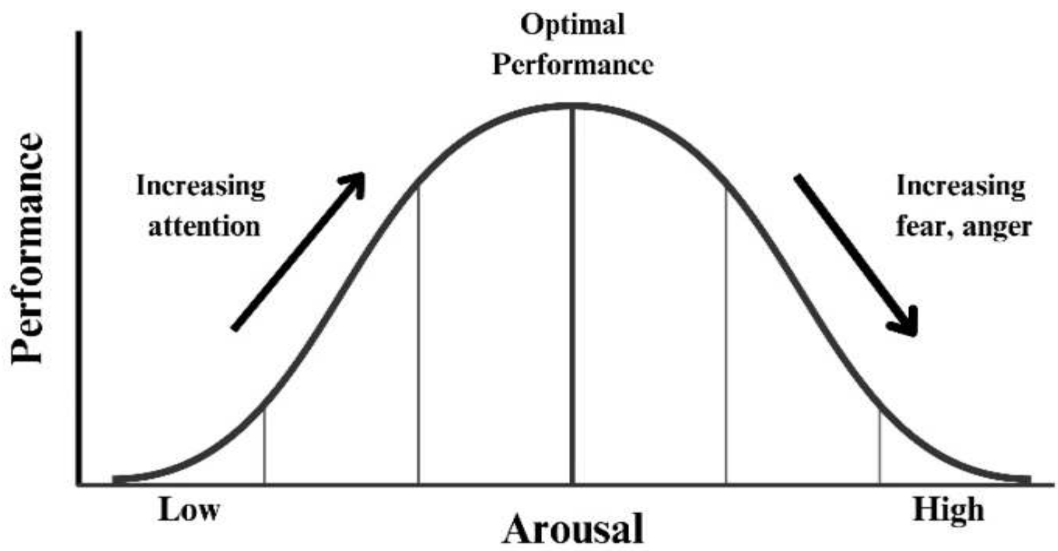
The study of arousal in MSL is important because it can inform various learning and rehabilitation programs. A clearer understanding of how arousal affects performance during MSL could help to create interventions in fields such as surgery, sports, or neurorehabilitation where there is a need to regulate one's emotions adequately for optimal performance. This is particularly important for clinical populations, such as individuals with ADHD, anxiety, or the elderly, who may benefit from optimizing cognitive systems involved in learning and memory (Martella et al., 2020; McGaugh, 2004).

1.1.1 The Yerkes-Dodson Law

The most prominent theory regarding the relationship between arousal and performance is the Yerkes-Dodson law (famously called the Inverted-U hypothesis). The original idea proposed by Yerkes and Dodson (1908) states that arousal and performance follow the previously described quadratic relationship that can be identified by its U-shape (see Figure 1). According to the law, insufficient arousal results in feelings of boredom, disengagement, or sleepiness, whereas excessive arousal leads to stress, anxiety, or overstimulation. Moderate arousal, on the other hand, presents an optimal balance of activation and available capacities that enables optimal performance. Despite methodological limitations, such as its reliance on laboratory mice, the law has become the most popular theory of arousal (Winton, 1987). We review evidence for and against it from fields related to MSL to explore when and how it applies, and how methodology may affect the results. Then, we proceed to other constructs of interest and the experimental design of this thesis.

Figure 1

A Typical Visualization of the Yerkes-Dodson Law



1.1.2 Empirical Evidence on the Yerkes-Dodson Law

Several studies support the Yerkes-Dodson law in various contexts. However, the literature regarding the impact of arousal on performance during MSL is scarce. Therefore, we refer to literature that investigates the effects of arousal on motor performance or performance during cognitive tasks which share key aspects of sequence learning. Additionally, arousal can be measured in numerous ways, making it more challenging for researchers to draw inferences about its general effects. We present two studies that have supported the effect of arousal on motor performance and learning rates and compare their characteristics to studies that provide evidence against the Yerkes-Dodson law.

Firstly, in a study by Rietschel et al. (2010), college-age adults performed a target aiming task. They induced arousal, which they measured via heart rate, skin conductance, and self-report, by applying a social-evaluative condition in which participants believed they were being observed and evaluated by an outside force. Additionally, they assessed performance by assessing the accuracy and smoothness of the movements. According to the researchers, moderate arousal corresponded to more accurate and fluid movements, supporting the Yerkes-Dodson law. They concluded that moderate arousal may cause neural changes that enhance motor planning efficiency.

Secondly, Nassar et al. (2012) conducted a predictive-interference task study to investigate the effects of arousal on learning rates in healthy individuals. They assessed arousal via pupil diameter changes and occasionally induced an arousal spike with an unexpected but irrelevant sound that reliably increased the participants' pupil size. The researchers found that the arousal spike of the auditory cues enhanced the participants' learning rates when their baseline arousal was low. On the other hand, the cues decreased their learning rates when their baseline arousal was already high. These results support the Yerkes-Dodson law by suggesting that moderate arousal is the ideal level of arousal for learning and indicate that arousal may affect learning in addition to performance.

In contrast to the previous two studies, Bond et al. (2023) performed an experiment using the Sustained-Attention-to-Response Task (SART) to assess sustained attention and response inhibition of healthy individuals by identifying lapses in attention. They administered the Karolinska Sleepiness Scale (KSS) to measure arousal via self-report and hypothesised that moderate arousal would lead to optimal attention and performance. However, the researchers did not find evidence to support a quadratic relationship between SART performance and daily arousal fluctuations and rejected the Yerkes-Dodson law for their experiment. These results could suggest that a unitary measurement of arousal may not be sufficient to lead to significant results.

Furthermore, Chang et al. (2009) explored the relationship between exercise-induced arousal and performance on simple and choice RT tasks. For these purposes, participants cycled at eight different levels of exertion, measured via heart rate, to induce wide ranges of arousal. RT was divided into a premotor, motor, and movement phase to identify the exact stage that was affected by arousal. The researchers found that arousal had a negative linear effect on RT during the motor and the movement phase, whereas the premotor phase remained unaffected by arousal. They implied that the effects of arousal on performance may be linear instead of quadratic and highlighted the potential stage-dependent effects of arousal. The researchers discuss that quadratic functions may be more likely for tasks that focus on cognitive processing and that linear functions may apply more to tasks that require greater motor processing instead.

In conclusion, there is mixed evidence to support a curvilinear relationship between arousal and motor performance and learning among humans. These studies vary in task characteristics and apply measures of arousal that range from physiological to self-report. Most studies examine the effects on healthy and young participants. Older and neurodivergent populations remain relatively unexplored. Additionally, there is a lack of studies that directly evaluate the effects of arousal on performance in MSL.

While there exist numerous studies that replicate a quadratic relationship between arousal and performance, the notion that a negative linear relationship may be more appropriate and is well supported by existing meta-analyses (Corbett, 2015; Muse et al., 2003). This inconsistency supports research that does not treat the Yerkes-Dodson law as a universal law (as the name would suggest) but rather as a framework whose validity depends on several factors, including the specific task demands and individual characteristics. Researchers should be aware that linear relationships may be more appropriate to describe these relationships than quadratic ones, depending on their experimental task. Secondly, there are numerous ways to conceptualize and operationalize arousal and researchers should disregard the concept of arousal as a unitary concept (Hancock & Ganey, 2003). Lastly, researchers must capture wide variability of data to be able to explain the entire relationship between arousal and performance adequately (Muse et al., 2003). For example, if only low to moderate levels of arousal were captured in the data, then it is possible that a linear relationship is observed instead of a quadratic one because the turning point of the quadratic relationship might not have been reached.

1.1.3 The Optimal Level of Challenge

The Challenge Point Framework (CPF), initially proposed by Guadagnoli and Lee (2004), underlines the value of identifying the exact ranges of difficulty that elicit moderate arousal by highlighting their potential positive impact on learning potential. It applies the concept of the Yerkes-Dodson law to MSL and highlights that individuals can perceive a “sweet spot” where their challenges and their available capacities align in a way that enable optimal learning. A fundamental concept of CPF is functional task difficulty, the level of challenge that a task places on an individual in relation to their current skill-level experience and environment. According to CPF, functional task difficulty and learning potential exhibit the same quadratic relationship as arousal and performance in the Yerkes-Dodson law. This effect is presumably caused by changes to attention and other cognitive resources, which affect the encoding and execution of learned motor sequences (Guadagnoli & Lee, 2004).

Nominal task difficulty on the other hand refers to the inherent, fixed challenge of a task and researchers can manipulate it by adjusting sequence length or target size in a MSL experiment, for example. In contrast to functional task difficulty, the literature has consistently demonstrated a negative linear relationship between nominal task difficulty and performance. For example, a study by Rhein and Vakil (2017) compared the transfer of learned sequences during a serial reaction time task (SRTT). Healthy young adults learned to press either a 6-item or 12-item sequence on a keyboard and then transferred their knowledge to the other sequence. The researchers found that transferring short sequences to longer ones resulted in slower RT and interpreted this finding as evidence that nominal task difficulty impairs performance. These findings align with a study by Raw et al. (2019), which found that sequence length was increased RT among healthy individuals who moved a computer mouse from a central location to one of eight target locations in a sequencing task. Therefore, increased nominal task difficulty leads to a decrease in performance because more complex sequences require greater cognitive and motor preparation. Similarly, it can be assumed that nominal task difficulty increases arousal and workload while simultaneously decreasing pleasure.

1.2 The Relationship between Affect and Performance

One way to enhance the interpretability of self-reported arousal is to apply the concept of affect, the underlying feelings behind an individual's response to a certain situation. Psychologists typically conceptualize affect along two dimensions: arousal and pleasure (or valence). Whereas arousal refers to an individual's level of activation, pleasure refers to how positively or negatively the individual perceives that situation. When measuring arousal in isolation, researchers generally cannot distinguish between positive arousal like excitement or negative arousal like stress. When measuring affect, however, researchers can make a distinction between positive and negative arousal and can therefore capture the participants' inner world more accurately and can receive greater insight into how or why performance varies as the result of emotional and cognitive states.

Having explained the relationship between arousal and performance, the relationship between pleasure and performance is less clearly described by literature. Chan et al. (2018) assessed the effects of pleasure on a computerised attention task (CAT) and an SRTT to evaluate the effects of focused attention meditation on MSL performance. Among their findings was that pleasure remained relatively unaffected between the phases of the experiment, and that its effects on performance depend on the kind of pleasure and performance observed. To demonstrate, although the effects were weak, task pleasure related to greater disruption in responses to changes of the sequence, and sequence learning pleasure related to greater sequence-specific learning but less general learning. For the latter, a possible explanation is that participants who focused more strongly on the sequence may have done so at the cost of general performance improvements, possibly by attending more to pattern structure than to overall speed.

Music is one of the most common tools to reliably induce pleasure among participants. For example, Valencia et al. (2018) instructed healthy adults to perform the Purdue Pegboard dexterity test of fine motor skill, and participants listened to either a high- or low-tempo sonata. They administered the Affect Grid to measure the participants' arousal and pleasure but did not find a significant difference between the two conditions. In contrast, another study has found that general motor output during a drumming task, as measured via hand acceleration and Electromyography (EMG), increased with corresponding increases in self-reported pleasure (Palumbo et al., 2024). In summary, studies exist that support and contradict the assumption that pleasure affects motor output and performance, and the exact relationship remains elusive and requires further research to quantify.

1.3 The Relationship between Workload and Performance

In addition to the Affect Grid, we apply a measure of workload as an indicator of task difficulty to receive insights into the specific demands that are relevant during the present experiment. According to Gopher and Donchin (1986), workload refers to the discrepancy between the information processing system's capacities that are needed to complete a certain task and its capacities that are available at that time. It reflects the various demands that a task places on individuals' resources, such as attention, memory, or physical capacities in relation to their available cognitive and physical capacities. Literature generally supports the notion that a moderate degree of workload is optimal and that too little workload leads to inattentiveness. In contrast, overload can occur when the task demands exceed the available resources (Vidoulich & Tsang, 2012).

Despite the notion that moderate workload enables optimal performance, several studies could not support a quadratic function between the two variables. In one study, adults performed a novel reaching task with two levels of nominal difficulty, high and low. The researchers assessed workload using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) questionnaire, and motor performance with accuracy and improvement rates over practice. Although all participants improved over time, higher mental workload was consistently associated with poorer motor performance and slower improvement rates and there was no indication of a quadratic relationship (Shuggi et al., 2017). In another study, higher NASA-TLX scores among novice surgical trainees were related to more frequent errors during a surgical task (Yurko et al., 2010). Therefore, similar to affect, there is some evidence regarding workload that is in favour of a linear relationship over a quadratic one with performance.

1.4 The Present Experiment

The present experiment utilizes the Dance-Step Discrete Sequence Production (DS-DSP) task, a dance motor sequence learning paradigm initially proposed by Chan et al. (2025), to investigate the functional relationships between self-report and behavioural variables and performance. The task enables more naturalistic and ecologically relevant observations by instructing participants to engage in full-body movements that simulate real-world motor learning scenarios like sport or rehabilitation more clearly. We manipulate nominal task difficulty through sequence length, measure performance through RT, and employ the Affect Grid and the NASA-TLX to assess affect and workload.

The present research follows three primary aims: (1) to evaluate the effects of nominal task difficulty on RT performance; (2) to evaluate the effects of nominal task difficulty on self-reported arousal, pleasure, and workload; and (3) to determine the functional relationships that arousal, pleasure, and workload share with RT performance. We formulate the following hypotheses: (1) Nominal task difficulty follows a linear positive relationship with RT; (2) Nominal task difficulty follows a positive linear relationship with arousal and workload while following a negative linear relationship with pleasure; and (3) arousal, pleasure, and workload follow a quadratic relationship with RT.

2. Methods

2.1 Participants

We recruited 23 students from the University of Twente to take part in the study. Convenience sampling was applied by posting an invitation on Sona Systems, a web-based platform to recruit research participants. Participants were informed of the following exclusion criteria: Be healthy and between 18 and 35 years old; be physically uninjured and able-bodied; have had no incidents of falling or heart problems in the past year; have no learning disabilities, diagnosed mental health conditions, or neurological disorders (e.g., Alzheimer's, Parkinson's, stroke, multiple sclerosis, brain tumours, brain injuries, seizures, concussions, or coma); have not previously participated in motor learning experiments involving dance-step sequence tasks in the BMS or via SONA; be available for one data collection session lasting up to 3 hours and willing to learn a dance-step sequence; be comfortable with motion capture sensors being attached to your legs, feet, and pelvis; be generally well, with no serious sleep issues requiring medication, and no history of depression, anxiety, drug, alcohol, or tobacco addiction; have no visible physical injuries or impairments that would affect performance; have not consumed alcohol in the last 24 hours before the experiment.

The study was approved by the University of Twente's Behavioural, Management and Social Sciences Ethics Committee (No. 240848) (See Appendix B). The final sample consisted of 16 eligible and complete responses (mean Age = 21); six participants had to be excluded because of technical issues in capturing the data, and one needed to be excluded because more than half their trials were removed by RT and accuracy filters. The final sample was largely made up of female (56.3%), right footed (93.8%) and non-smoking (81.2%) participants. Despite the exclusion criteria, 31.3% of participants mentioned they drank alcohol in the last 24 hours before the experiment.

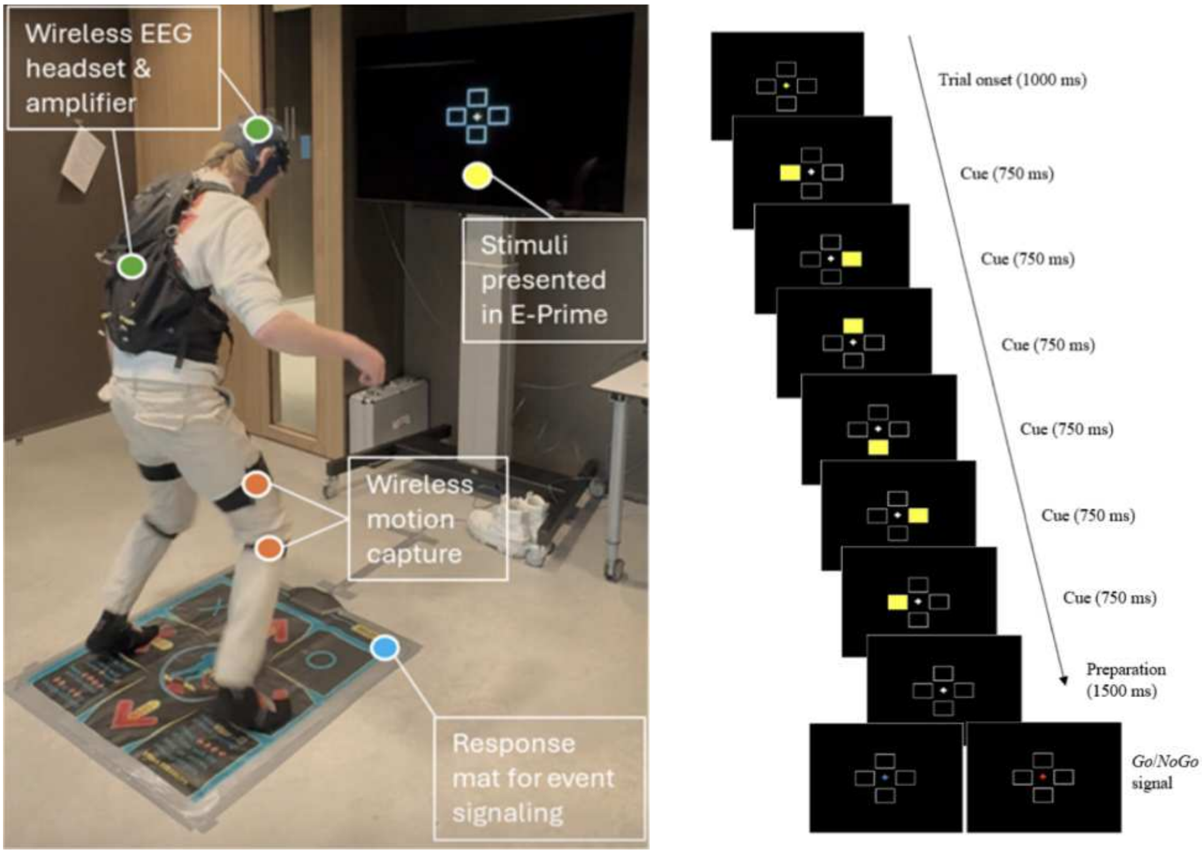
2.2 Materials

2.2.1 The DS-DSP Task

The DS-DSP Task was programmed in the E-Prime 3.0 software based on a script originally developed by Chan et al. (2025). E-Prime measured the RT of each step and tracked whether the performed step was correct or false. Participants performed the task on a commercially available D-Force Nonslip USB Dance Pad (see Figure 2A). The dance pad was attached to the floor with tape and connected to the laptop running E-Prime via USB and interfaced using JoyToKey, which mapped the directional pad inputs (↑, ←, ↓, →) to keyboard keys ('w', 'a', 's', 'd'). This procedure enabled E-Prime to register footsteps as regular keyboard inputs. The stimuli were presented on a large monitor that was positioned 1.2 meters level in front of the participants centred at eye level. The monitor was connected to the laptop running E-Prime via HDMI. See Appendix B for a complete visualisation of the synchronization of task device protocols.

Figure 2

A) The EEG Cap and The Dance Pad *B) The Experimental Paradigm*



* originally from Chan et al. (2025) and adapted for this report

2.2.2 EEG and XSens

This study included Xsens motion tracking and electroencephalography (EEG) equipment in addition to the collection of self-report data. Participants wore a mobile 22-electrodes EEG cap by ANT Neuro while carrying its amplifier in a sports backpack, and 7 XSens sensors were attached to the participants with Velcro Binds: One at both feet, shins, and thighs and one on the pelvis (See Figure 2A). However, the analysis of the EEG and motion capture data is the subject of other manuscripts.

2.2.3 Affect and Workload Questionnaire

We used two self-report questionnaires to measure the participants' subjective experiences during the experiment. The first questionnaire was used to gather subjective assessments of the participants' arousal, pleasure, and workload before and over the course the task. For these purposes, we applied the Affect Grid to measure arousal and pleasure and the NASA-TLX to measure workload.

The Affect Grid

We applied the Affect Grid to assess participants' levels of pleasure and arousal during the experiment (Russell et al., 1989). The Affect Grid is a single-item, two-dimensional measure of core affect, with Pleasure on the x-axis (unpleasant to pleasant feelings) and Arousal on the y-axis (sleepiness to high arousal). Participants selected a square on a 9-by-9 grid that best reflected their perceived current state. Higher values indicated greater pleasure or arousal. The Affect Grid also specifies the states that might occur because of an individual's level of arousal and pleasure. For example, high arousal and low pleasure leads to stress whereas low arousal and high pleasure leads to relaxation (see Appendix C). The Affect Grid provides a quick and straightforward tool for capturing momentary emotional experiences, making it suitable for longer studies with repeated measurements.

The NASA-TLX

The NASA-TLX questionnaire, developed by Hart and Staveland (1988) of the NASA Ames Research Centre, was used to assess the participant's subjective workload before and during the task. It is the most widely used assessment of workload and originated from the aviation industry, where it was derived from experiments on aircraft simulation, laboratory tasks, and control tasks. The TLX is a multidimensional measure of workload and accounts for six subdimensions: Mental demand, physical demand, temporal demand, performance, effort, and frustration (See Appendix C for the official definitions of the six subdimensions). Over the years, the TLX became the most cited measure of workload (Grier, 2015), and Researchers have successfully applied the TLX to various industries, including healthcare (Hoonakker et al., 2011; Lowndes et al., 2020). Furthermore, research has consistently and independently shown that the TLX is the most valid and reliable assessment of workload (Grier, 2015; Hoonakker et al., 2011).

Studies using the TLX typically investigate the effects of the subdimensions in isolation and calculate a weighted aggregate score that represents general workload by letting participants perform a pairwise comparison task (Hart, 2006; Hart & Staveland, 1988). However, many studies instead calculate the ‘Raw TLX’ (RTLX), in which participants do not perform the pairwise comparisons. The RTLX and weighed TLX are strongly correlated measures of general workload and were independently validated by literature (Grier, 2015; Said et al., 2020). In the present experiment, participants rated each dimension on a scale of 1 (very low) to 21 (very high). We analysed the subscales independently and computed a Raw TLX (RTLX) score.

Both the Affect Grid and the NASA-TLX have limitations, but they are particularly suited for the present experiment. On the one hand, they rely on introspection which introduces potential biases like the social desirability bias where participants may give answers that they believe will make them look better instead of expressing their true opinion or experience (Donaldson & Grant-Vallone, 2002). Specifically, participants may report lower arousal and workload and higher pleasure than they truly perceive. Additionally, omitting the weighted comparisons could obscure individual differences in the perceived importance of the TLX subscales. On the other hand, we chose these measures because they offer quick repeated measures that are well suited for a lengthy block-based experiment. The Affect Grid’s single item format enables participants to report their perceived arousal and pleasure in just a few seconds (Russell et al., 1989). Due to the omission of paired comparisons, the RTLX is less demanding for the participants to fill out but still retains a strong correlation to weighted TLX scores (Grier, 2015; Said et al., 2020). Therefore, we chose these measures to minimize the length of the experiment and the demands that are placed on the participants’ cognitive systems between blocks.

2.2.4 Memory Recall and Strategy Questionnaire

To assess memory of the performed sequence, participants were asked to enter the directions of the steps in the correct order. They were first asked to enter the steps of the 6-step sequence, then the 12-step sequence, and finally the entire 18-step sequence. After each sequence, they were asked to indicate how sure they are about the correctness of their given answer on a scale of 1 (unsure) to 10 (sure).

Regarding the techniques that participants used to memorize the sequence, they were given the following options to choose from: *"I remembered the order of the arrows", "I remembered the position of the arrows", "I remembered the position of the blocks on my screen", "I tapped the sequence in my mind",* and *"I re-enacted the sequence with my body"*. Participants were allowed to choose multiple and to name further techniques if they were not mentioned.

Afterwards, participants were asked about their prior experience with motor sequence learning. On a scale of 1 (definitely not) to 5 (definitely yes), participants answered the questions *"Have you ever participated in an experiment having to deal with learning sequences before?"* and *"Do you have any personal experience with learning sequences? (Think of playing an instrument)"*. Furthermore, participants estimated the average amount of hours they spend playing video games each week and assigned themselves to one of the following categories: *"Complete beginner / I do not game", "Beginner", "Intermediate", "Advanced",* or *"Expert"*. Lastly, participants were asked to write any further remarks about the experiment into an open textbox.

2.3 Procedure

Before starting the experiment, participants were shortly briefed about the study purpose, their right to withdraw at any time, and that they will be competing for two cash prizes given to the fastest and most accurate participants. Participants were given and asked to fill out the informed consent (see Appendix B) and were then instructed to wash their hands with shampoo and to take off their shoes so the researchers could measure their feet size. Furthermore, the participants' head size was measured to select the most appropriately sized EEG cap and conductive gel was applied using blunt-tipped syringes to bridge the electrodes and ensure acceptable impedance levels and conductivity.

The researchers entered the participants height and weight into the MVN Analyze software to accurately model their 3D figure and centre of mass. The seven Xsens sensors were attached to the participants with Velcro Binds: One at both feet, shins, and thighs and one on the pelvis. Afterwards, Xsens was calibrated by telling the participants to stand in the N-pose (standing straight with both arms hanging from the sides) and to wait for the researchers to start MVN Analyze. Then, the participants were asked to walk 4 meters in a straight line, to turn around, and to walk back into the starting position. This process was repeated until MVN informed the researchers that at least an acceptable calibration was achieved. Participants filled out the pre- and peri task questionnaire to determine their baseline levels of arousal, pleasure, and workload.

Afterwards, the E-Prime script was executed, and participants were told to stand with their feet firmly on the centre of the dance mat and look at the monitor. Participants watched a sequence of stimuli until a blue or red plus sign told the participants to respond or not to respond with their feet on the dance mat. The stimuli consisted of arrows that pointed into one of four directions: ↑, ←, ↓, or →. See Figure 2B for a visual overview of the experimental paradigm as proposed by Chan et al. (2025).

Participants underwent 5 blocks: 3 training blocks and 2 testing blocks. During the training phase, participants learned a sequence that was extended with each new block. During the first block, participants learned the first 6-steps of the sequence. During the second block, participants learned an additional 6-steps, so 12-steps in total, and during the third block, participants performed the entire 18-step sequence. In the testing phase, the first block (Testing-1) tested the already learned sequence, while the second block (Testing-2) tested completely unfamiliar sequences. The trials within both testing blocks randomly varied between either 6-, 12-, or 18-steps. Every block consisted of exactly 48 trials. At the beginning of every block, participants were reminded to respond as quickly and accurately as possible and to then return to the starting position. After the first four blocks, participants were given a short break of approximately 3 to 5 minutes to recover and to fill out the first questionnaire consisting of the Affect Grid and NASA-TLX. After the last block, participants filled out the second questionnaire about their memory of the performed sequence, their prior experiences, techniques, and remarks on the study. The study took approximately three hours for each participant. (See Figure 3 for an overview of the experimental procedure)

Figure 3

The Experimental Procedure

		Learning Phase:			Testing Phase:	
Preparation	Baseline	6-Steps	12-Steps	18-Steps	Testing-Familiar	Testing-Unfamiliar
Consent	Demographics	Reaction Time	Reaction Time	Reaction Time	Reaction Time	Reaction Time
EEG Cap	Affect Grid	Affect Grid	Affect Grid	Affect Grid	Affect Grid	Memory Retention
Xsens Calibration	NASA-TLX	NASA-TLX	NASA-TLX	NASA-TLX	NASA-TLX	Experience & Technique
	EEG & Xsens	EEG & Xsens	EEG & Xsens	EEG & Xsens	EEG & Xsens	EEG & Xsens

* *We administered the questionnaires after each block*

2.4 Data Preparation & Analysis

Data cleaning and analysis has been performed on RStudio by Posit Software, Version 2024.12.1+563. The following packages were required: Tidyverse, readxl, readr, ggplot2, tidyr, dplyr, lme4, emmeans, rstanarm, effects, and car. The self-report data was downloaded from the cloud-based platform Qualtrics, and the RT data was extracted from E-Prime 3.0. Both data frames were exported as .csv files. The complete R code can be accessed on Rpubs (https://rpubs.com/jakob_jonael/1295431).

Two filters have been applied to the RT data frame to account for outlying trials in terms of RT and accuracy. First, all trials with an average RT higher than 2 standard deviations within the participant and block data were excluded. Secondly, all trials with a mean accuracy below 80% were excluded as well. By applying these filters, 29.7% of observations within the data were removed. The amount of removed trials were calculated for all participants. It was identified that one participant got 54.5% of trials removed by applying the filters. Additionally, another participant’s trials were incorrectly administered by the script. Both these participants were removed from the final data frame.

We computed the variable BMI by dividing the participants’ weight in kilograms by their height in meters squared, as well as the variable RTLX by adding the values of all NASA-TLX subscales and dividing the result by six. To ensure that high scores always correspond to higher workload, we reversed the Performance scale by subtracting its values from 21 before computing the RTLX. We then merged the self-report and the RT data to create a complete data frame for further analysis and calculated the sample demographics. Specifically, we computed the mean participants' age, and the relative proportions of gender, footedness, smoking, and alcohol consumption in the last 24 hours.

2.4.1 Behavioural: Nominal Task Difficulty on RT Performance

Firstly, we performed a linear mixed-effects model to assess whether nominal task difficulty significantly affects RT performance during the present experiment. Nominal task difficulty was manipulated through the sequence length of each block: 6 steps corresponding to easy, 12 steps to medium, and 18 steps to hard. We included block as a fixed effect and added a random intercept for participant ID to account for repeated measures and within-subject variability. The significance of fixed effects was tested using Type II Wald chi-square tests with the Anova() function. Post-hoc analysis involved pairwise comparisons of estimated marginal means using the emmeans() function, with Tukey adjustments applied for the block factor with five levels. Lastly, we visually inspected histograms of residuals and quantile-quantile (QQ) plots to assess the distribution and normality of residuals (see Appendix D).

2.4.2. Self-report: Nominal Task Difficulty on Affect and Workload

Secondly, we performed nine linear mixed-effects models to assess whether nominal task difficulty significantly affected self-reported arousal, pleasure, workload as represented by RTLX, and the six NASA-TLX subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration). In each model, block was included as a fixed effect and participant ID as a random intercept to account for repeated measures and within-subject variability. We tested fixed effects with Type II Wald chi-square tests using the Anova() function. When a significant main effect of block was found, post-hoc analyses with pairwise comparisons of estimated marginal means were performed using the emmeans() function, applying Tukey adjustments for the block factor with five levels. Histograms and QQ plots of residuals were inspected to evaluate normality and distribution (see Appendix D).

2.4.3 Functional Modelling of self-reports to predict RT performance

Lastly, we performed nine linear regression models and nine quadratic regression models to investigate whether the independent variables, self-reported arousal, pleasure, RTLX, and the six NASA-TLX subscales, significantly affected the dependent variable average block RT. We carried out these analyses to test the Yerkes-Dodson law and assess whether the relationship between the independent variables and performance followed a linear or quadratic function to determine the most appropriate functional relationships. We fitted each model using the lm() function. In the quadratic models, we included a squared term to capture potential quadratic effects. We assessed the assumption of normality and distribution of residuals by inspecting the models’ histograms of residuals and quantile-quantile (QQ) plots (see Appendix D).

3. Results

3.1 The Final Sample

The final sample consisted of 16 eligible responses. After applying the RT and accuracy filters, the mean RT reduced from 564.47ms to 556.57ms, and the mean accuracy improved from 84.53% to 98.28%. The block that kept the most trials was 6-Steps, whereas 18-Steps and Testing-Unfamiliar kept the fewest trials (see Table 1). Every remaining participant kept at least 50% of their trials.

Table 1

The Number of Remaining Trials per Block after the Filtering

Block	Remaining Trials
6-Steps	614 (97.95%)
12-Steps	524 (68.23%)
18-Steps	475 (61.85%)
Testing-Familiar	582 (75.78%)
Testing-Unfamiliar	480 (62.50%)

* Each block originally consisted of exactly 768 trials

3.2 Descriptive Statistics

The means of the variables arousal, pleasure, and workload for each block are in Table 2. Besides the baseline measurements, most variables were lowest during 6-Steps and Testing-Familiar, whereas they peaked during blocks 12-Steps and 18-Steps. Notably, the reverse is the case for pleasure and performance which are the only positive constructs among the other negative ones. Generally, when considering that the scales ranged from 1 to 21, most variables received moderately low scores over the course of the experiment, ranging from about 3 to 10.

Table 2

Descriptive Statistics

	Baseline	*6-Steps*	*12-Steps*	*18-Steps*	*Testing-F*	*Testing-U*	*Overall*
RT	/	482.65	501.33	534.46	465.13	549.74	515.66
Arousal	5.56	6.06	6.25	5.25	4.46	/	5.54
Pleasure	6	6	5.75	6.25	6.63	/	6.13
RTLX	5.5	6.85	8.1	8.66	7.38	/	7.3
MD	3.75	6.5	9.44	10.31	9.25	/	7.85
PD	3.38	4.5	6.75	7.5	6.25	/	5.675
TD	3.5	5.5	6.25	5.94	5.25	/	5.29
Per	6.94	10.81	12.5	11.87	12.94	/	11.01
Eff	3.93	7.75	10.13	10.25	9.06	/	8.23
Frus	4.38	6.69	7.56	8.88	6.44	/	6.79

** MD = mental demand; PD = physical demand; TD = temporal demand; Per = performance; Eff = Effort; Frus = Frustration*

** To compare performance with other TLX variables subtract the values from 21*

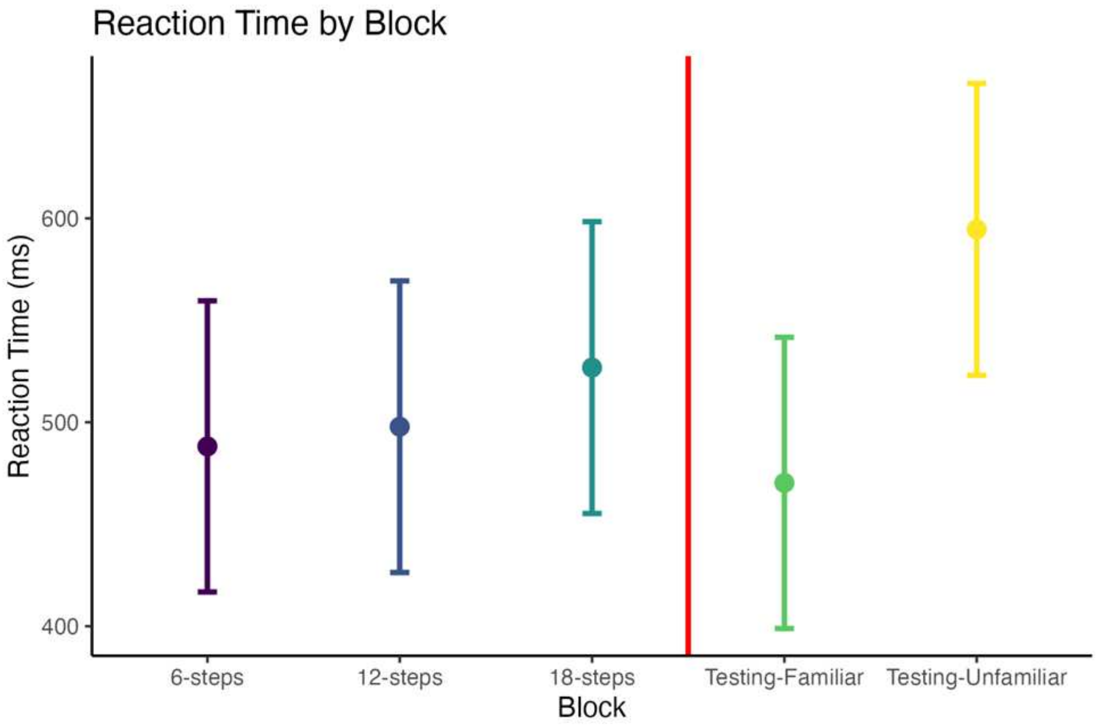
3.3 Statistical Analyses

3.3.1 Behavioural: Nominal Task Difficulty on RT Performance

Block RT changed significantly over the blocks of the experiment. The linear mixed-effects model revealed a significant main effect of nominal task difficulty on RT ($\chi^2 = 174.11$, $p < .01$). Pairwise comparisons (emmeans with Tukey adjustment) revealed that RT increased steadily from baseline to 18-Steps, with a significant contrast from 12-Steps to 18-Steps ($p = 0.04$), a momentary decrease from the hardest block 18-steps to Testing Familiar sequences ($p < 0.01$), and then another increase from Testing Familiar to Testing Unfamiliar ($p < 0.01$) (see Figure 4).

Figure 4

RT increases significantly over the blocks and only decreases from 18-steps to testing of familiar sequences



3.3.2 Self-report: Nominal Task Difficulty on Affect and Workload

See Table 3 for an overview on the results of the ANOVA of the linear mixed-effects models that were performed to test whether nominal task difficulty significantly affects self-reported arousal, pleasure, and workload. Arousal changed significantly over the blocks (χ^2 = 16.5, $p < .01$). It generally increased from Baseline to 12-Steps before decreasing linearly until Testing Familiar sequences. The results of the pairwise comparisons did not outline a significant change from one block to the next (see Figure 5). In contrast, pleasure did not significantly change over the blocks of the experiment ($p = 0.51$).

Table 3

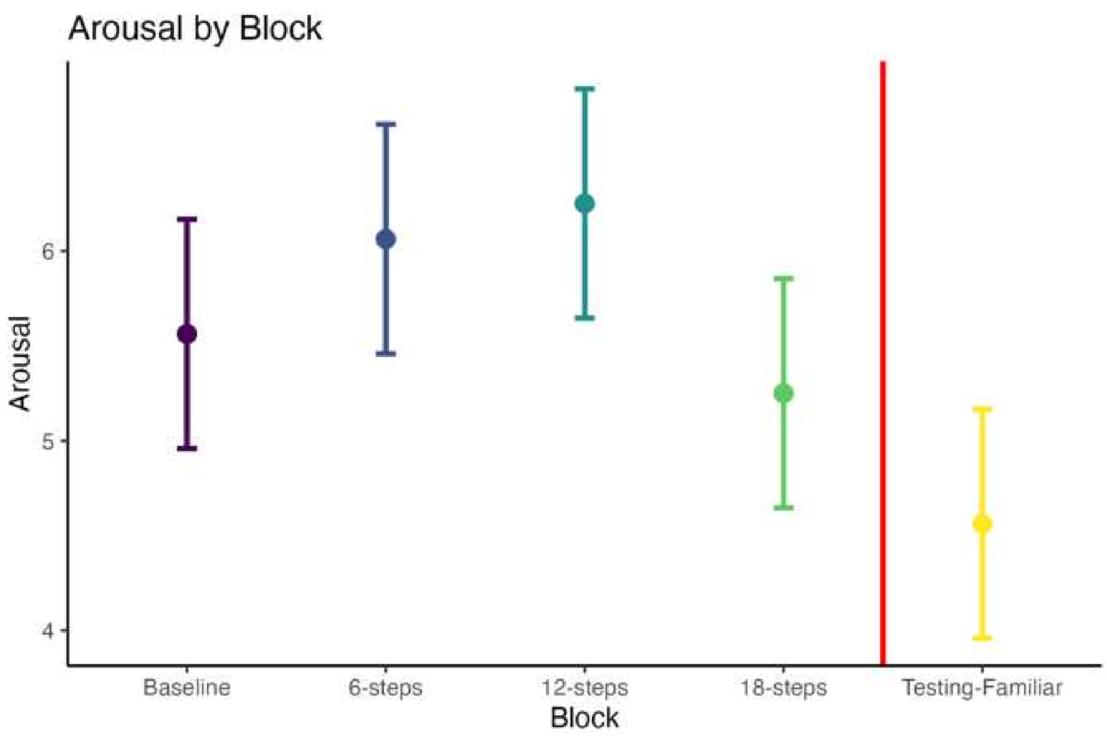
ANOVA of the linear mixed-effects models: Nominal Task Difficulty on Affect and Workload

	χ^2	*p-value*
Arousal	16.5	<0.01*
Pleasure	3.3	0.51
RTLX	23.17	<0.01*
MD	36.23	<0.01*
PD	27.81	<0.01*
TD	12.16	0.02*
Per	19.35	<0.01*
Eff	30.98	<0.01*
Frus	12.22	<0.01*

** To compare performance with other TLX variables subtract the values from 21*
** Df of each analysis was equal to 4*

Figure 5

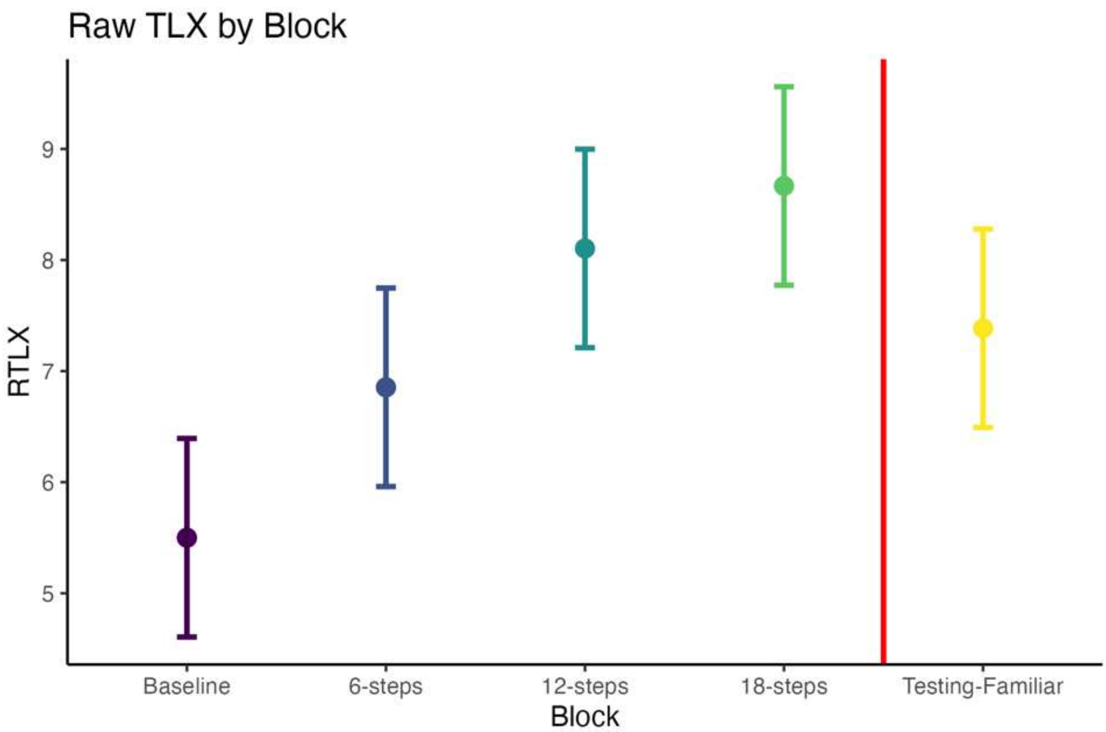
Arousal increases from baseline to 12-steps and then decreases until testing of familiar sequences



Both the RTLX and each of the TLX subscales changed significantly over the blocks (see Table 3). Generally, workload increased over the course of the blocks and decreased just slightly during the testing phase. Pairwise comparisons did not indicate a significant contrast from one block to the next (see Figure 6). Most workload subscales were most demanding during 12-Steps and 18-Steps and least demanding during 6-Steps besides Baseline (See Figures 7-12). In terms of the TLX subscales, the only significant contrast was regarding effort and RT from Baseline to 6-Steps (p = 0.04).

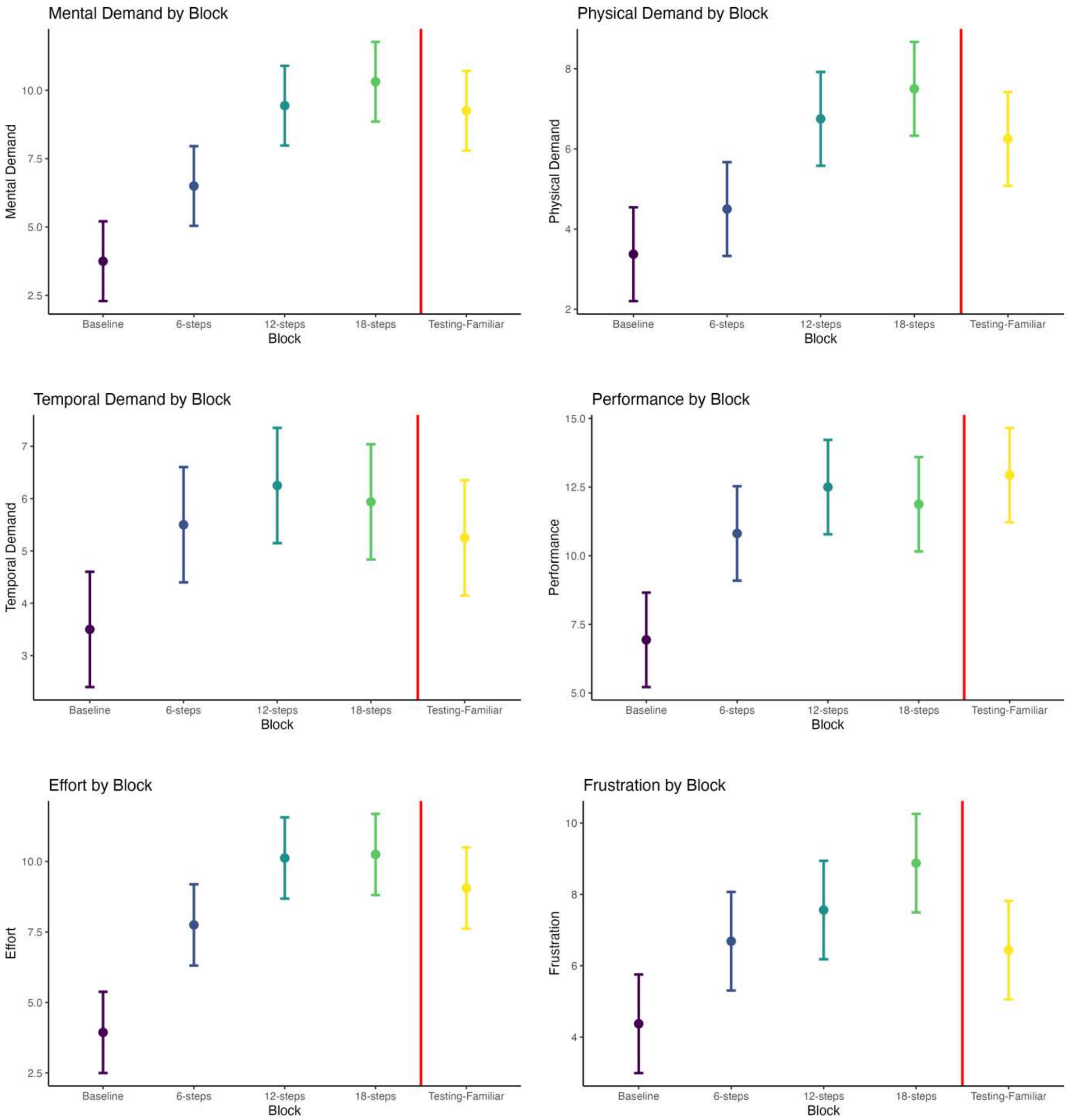
Figure 6:

RTLX increases from baseline to 18-steps and then decreases to testing of familiar sequences



Figures 7-12:

TLX subscales generally increase from baseline to either 12- or 18-steps, then decrease until testing of familiar sequences



3.3.3 Functional Modelling of self-reports to predict RT performance

See Table 4 for the results of the linear and quadratic regression analyses to determine the functional relationships of self-reported arousal, pleasure, and workload with RT. We only found two statistically significant functional relationships between the independent variables and RT. The effect of temporal demand on RT was significant in a negative linear function, meaning that an increase in temporal demand was consistently related to faster RT ($p = 0.03$).

However, temporal demand only explained a small amount of variance in RT (adjusted R^2 = 0.06). (See Figure 13).

Table 4

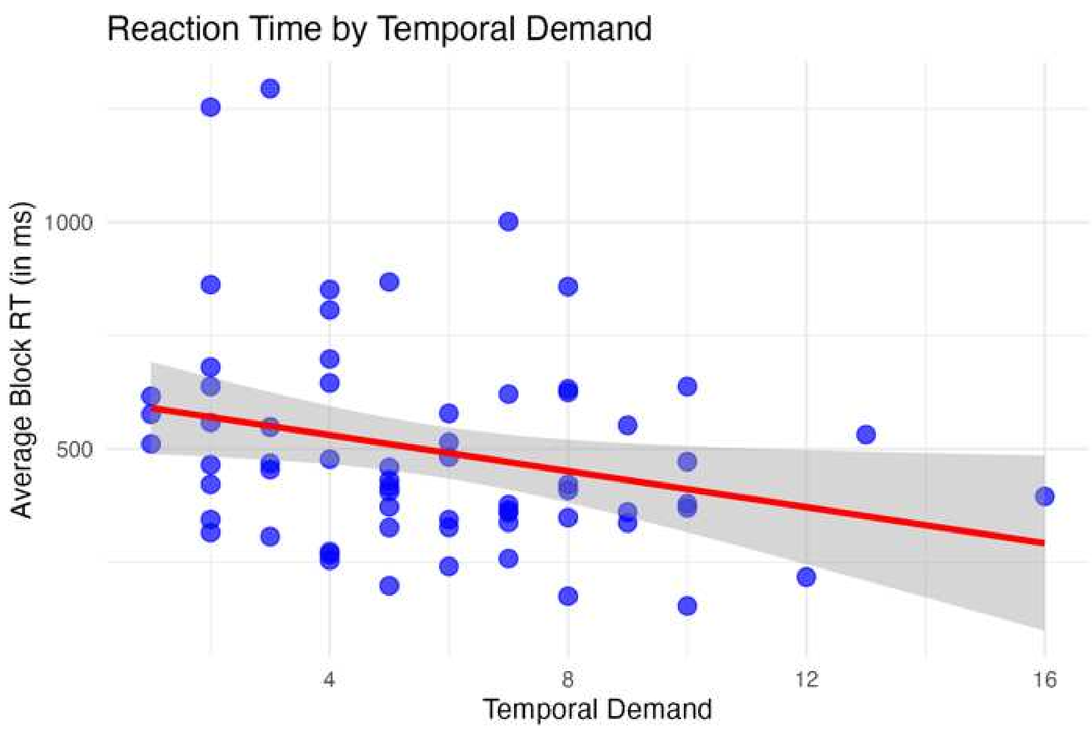
Prediction of RT performance in the different self-reports using linear and quadratic terms.

	Linear Regressions		*Quadratic Regressions*	
	p-value	*Adj. R^2*	*p-value*	*Adj. R^2*
Arousal	0.84	-0.02	0.69	-0.03
Pleasure	0.64	-0.01	0.63	-0.03
RTLX	0.96	-0.02	0.79	-0.03
MD	0.36	-0.0	0.44	-0.01
PD	0.08	0.03	0.2	0.05
TD	0.03*	0.06	0.4	0.05
Per	0.72	-0.01	0.55	-0.02
Eff	0.41	-0.0	0.98	-0.02
Frus	0.44	-0.01	0.03*	0.05

** To compare performance with other TLX variables subtract the values from 21*

Figure 13

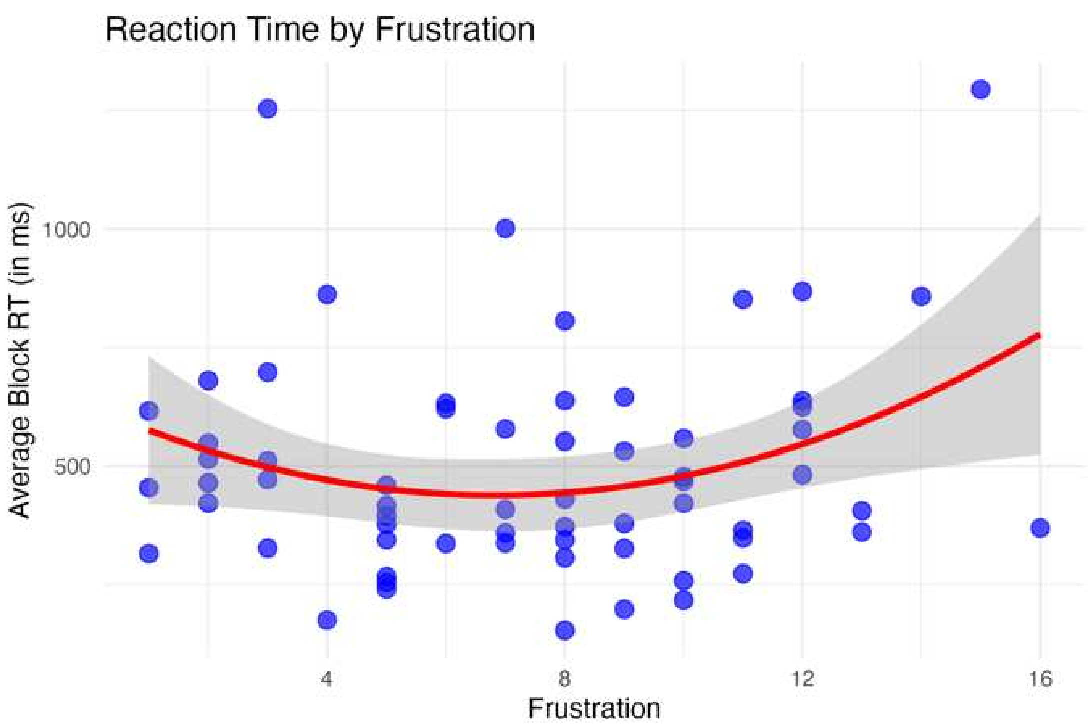
RT decreases linearly as a function of temporal demand



Frustration, on the other hand, demonstrated a statistically significant effect on RT in a quadratic function. Specifically, RTs were slowest at both low and high levels of frustration, while moderate frustration was associated with the fastest responses. The shape of the relationship follows a regular U-shaped pattern. However, as with temporal demand, frustration only explained a small amount of variance in RT (adjusted $R^2 = 0.05$) (see Figure 14).

Figure 14

RT follows a quadratic relationship with frustration, peaking at low and high levels



4. Discussion

The primary goals of the present study were (1) to evaluate the effects of nominal task difficulty on RT performance; (2) to evaluate the effects of nominal task difficulty on self-reported arousal, pleasure, and workload; and (3) to determine the functional relationships that arousal, pleasure, and workload share with RT performance. We hypothesized that (1) nominal task difficulty follows a linear positive relationship with RT; (2) Nominal task difficulty follows a positive linear relationship with arousal and workload while following a negative linear relationship with pleasure; and (3) arousal, pleasure, and workload follow a quadratic relationship with RT.

The results indicated that RT increased with greater sequence length, demonstrating that nominal difficulty negatively affects performance in a linear function. Furthermore, we identified the same linear effect for self-reported arousal and workload, where higher nominal difficulty related to greater demands on the individual's processing capacities. Regarding the functional relationships between the self-report variables and performance, we found support for a negative linear relationship between mental demand and RT which implies that perceived time pressure made the participants significantly faster. Lastly, the results supported a quadratic relationship between frustration and performance, where moderate frustration corresponded to slowest RT and best performance. In the following sections, we will outline the summary of findings, relation to hypotheses, theoretical implications, and directions for future work for all three aims of the study before moving on to the practical implications, limitations, and meaningful contributions of the present study.

4.1 Behavioural: Nominal Task Difficulty on RT Performance

Before we explain the results of the analyses, we want to discuss the participants' accuracy and their perceived demands during each block. The accuracy of each block was determined by calculating the difference in the amount of trials per block before and after applying the filters. The most accurate block was 6-Steps, whereas the least accurate blocks were 18-Steps and Testing-Unfamiliar. The least demanding blocks were 6-Steps and Testing-Familiar, whereas the most challenging blocks were 12- and 18-Steps. In essence, 6-Steps was the most accurate and least demanding block, whereas 18-Steps was the least accurate and most demanding block. These observations already suggest that nominal task difficulty and performance follow a linear function and that 12-Steps might have been the optimal level of challenge since 6-Steps was too easy and 18-Steps was too difficult.

We found statistical support that nominal task difficulty and RT follow a linear negative function, aligning with usual findings about the effects of nominal task difficulty on performance (Raw et al., 2019; Rhein & Vakil, 2017). There actually was a reduction in RT when transitioning from block 18-Steps to Testing Familiar sequences, implying that performing an already known sequences places fewer demands on an individual’s processing capacities. However, this result can also be explained by the sequence lengths of the Testing Familiar sequences block, where sequences randomly varied between either 6, 12, or 18-steps. Therefore, we accept the hypothesis that nominal task difficulty follows a linear positive relationship with RT.

4.2 Self-report: Nominal Task Difficulty on Affect and Workload

Our findings that workload changed significantly over the blocks also align with the literature on the effects nominal task difficulty (Raw et al., 2019; Rhein & Vakil, 2017). The same was not the case for pleasure, which remained relatively stable over the course of the experiment, similarly to the previously discussed study by Chan et al. (2018). Conclusively, we accept the hypothesis that an increase in sequence is related to an increase in workload and reject the hypothesis that an increase in sequence length leads to greater arousal and less pleasure.

The relatively small spread of pleasure compared to arousal and workload could explain the observation that pleasure was not significantly affected by nominal task difficulty. It is possible that nominal task difficulty might have affected pleasure, but that the participant’s progress in the study negated this effect. In other words, sequence length could have decreased pleasure, but the participants’ progress, both in terms of skill- and study progression, might have increased their pleasure as well, creating a null sum. Future research should more closely investigate the exact sequence lengths that are sufficient to cause such changes in perceived pleasure and should avoid the participants’ progress as a confounding variable. To circumvent this, researchers should consider counterbalancing the order of blocks across participants.

Applying these results to the CPF, we outline the optimal level of challenge during the present experiment. By investigating arousal scores in isolation, it appears that 18-Steps should be the optimal level of challenge since the arousal scores were moderate during that block in comparison to the other blocks. Considering that the arousal and workload scores were generally moderately low during the experiment, it is possible that participants experienced the task as relatively easy and that the optimal level of challenge would be higher than we could identify. In that case, the difficulty of the task would not have enabled us to sufficiently test the entire U-shaped relationship; a common methodological flaw according to Mather et al. (2015). However, since the decreases in performance over the blocks of the experiment were significant, and since there was a significant collapse in performance during 18-Steps, we have evidence to suggest that 12-Steps was indeed the optimal level of challenge for learning during the experiment.

4.3 Functional Modelling of self-reports to predict RT performance

We only found two significant relationships regarding performance: Firstly, with temporal in a linear function, indicating that participants' perceived time pressure corresponds to significantly faster responses. Secondly, the only significant quadratic relationship with performance was with frustration, where moderate levels of frustration resulted in faster responses, while high and low frustration levels resulted in slower responses. No other dimension of workload, nor arousal or pleasure, shared a significant quadratic relationship with RT. Therefore, we found minimal evidence in support of the Yerkes-Dodson law and reject our hypotheses that arousal, pleasure, and workload share a quadratic relationship with RT.

The finding about temporal demand represents both an alignment and a contrast with prior literature. On the one hand, the significant linear relationship between temporal demand and RT aligns with previous literature which suggest that perceived time pressure can enhance task focus and reduce hesitation, thereby improving speed., According to Wickens et al. (2012), people often prioritize speed over accuracy when under time constraints. Although this tendency might seem beneficial, real-world scenarios like performing surgery or sports reveal a greater risk for errors and injuries. Generally, however, this observation forms a contrast with prior literature which suggests that workload decreases performance in a linear function (Shuggi et al., 2017; Yurko et al., 2010).

Regarding frustration, the significant quadratic relationship with RT aligns well with the notion that moderate workload leads to optimal performance whereas insufficient workload leads to inattentiveness and excessive workload leads to overload (Vidoulich & Tsang, 2012). Halperin et al. (2020) found similar results when instructing recreational lifters to perform maximal elbow flexions under positive, negative, or no feedback. They found that those participants who received mildly critical feedback to induce moderate frustration significantly increased motor output and argue that the negative feedback expressed a need for the participants to improve. Our findings suggest that inducing frustration could be beneficial for performance and prompt the question whether this should be done in practice. It should be pointed out that frustration is very subjective and context-dependent, and that there is no optimal level of frustration that works best for everyone in every task. Therefore, researchers and practitioners should not induce a certain, fixed amount of frustration to increase performance. Instead, they could measure both performance and frustration and could manipulate task difficulty or frustration slightly until the optimal level is found (Akizuki & Ohashi, 2015). In essence, frustration can arise when there are obstacles to an anticipated goal or behaviour. When frustration is moderate, this can cause a reappraisal of the situation and might activate additional resources that could benefit focus and the monitoring of performance.

The result that arousal and pleasure did not correlate to performance in a linear or quadratic model during the present experiment aligns with studies that also could not replicate such an effect. Specifically, Bond et al. (2023) failed to support a relationship between arousal and SART performance, and Valencia et al. (2018) did not report an effect of musically induced pleasure on motor performance. Chan et al. (2018) found only a weak effect of pleasure on performance and no such effect of arousal. Interestingly, all those studies as well as this present thesis share a similarity in that they all rely solely on self-report. In contrast, all previously discussed studies that did not rely solely on self-report did find a statistical relationship between performance and arousal (Chang et al., 2009; Nassar et al., 2012; Rietschel et al., 2010), performance and pleasure (Palumbo et al., 2024). This observation aligns with the criticism of Hancock and Ganey (2003) that researchers should disregard the unitary concept of arousal. It should be noted that this applies only to studies investigating affect and not workload, as the other self-report studies investigating workload and performance with the NASA-TLX did report statistically significant relationships (as did our study partly). Therefore, the TLX may be accurate enough to measure workload whereas the Affect Grid in isolation might not be accurate enough to measure affect.

4.4 Limitations of this Study

It should be mentioned that we could have supplemented the self-reports with EEG, kinematic-, or physiological data, like skin conductance or heart rate. This could have provided a triangulation of results that may have increased accuracy and specificity. To demonstrate, Rietschel et al. (2010) measured arousal through a combination of heart-rate, skin conductance, and self-report, and out of all the studies we discussed that measured arousal through self-report, their study was the only one which reported a significant effect of arousal on performance.

We did not perform the pairwise comparison task to obtain weighted TLX data and calculated the RTLX instead. Although Grier (2015) and Said et al. (2020) validated the RTLX and reported a strong correlation with the weighted TLX scores, different results may have been obtained by instructing participants to weigh the workload subscales.

The sample of the present study was sufficient, but larger sample sizes are advisable for participant-level analyses. Additionally, the present sample can be characterized as a WEIRD sample since the majority of participants were wealthy, educated, industrialized, rich, and democratic. This composition may have altered the data and may have led to a misrepresentation of the general population (Wiggins & Christopherson, 2019).

The applied dance mat was a regular commercially available one. Although sufficient for this thesis, the participants have occasionally voiced complaints that the mat did not register their steps correctly, and the spaces for the participant's feet were relatively small. This effect might have been more drastic for taller people with larger foot sizes, potentially distorting the results.

4.5 Conclusion

This thesis only provided minimal support for the Yerkes-Dodson law during a whole-body sequencing task. Only frustration, a subdimension of workload, significantly predicted RT in a quadratic function. Additionally, the only significant linear function was between temporal demand and RT. We strongly advise a more multidimensional view on the theory and advise future research to consider individual and contextual characteristics more closely.

This study highlighted that researchers should not mistake a theory's intuitive logic for truthfulness and advocates for a critical approach to existing theories. This approach is especially important for psychologists, as psychological constructs like arousal are complex, multidimensional, and often imperceptible, leaving room for the researchers to conceptualize and operationalize them differently. For example, researchers could measure arousal using self-report, EEG, or physiological measures, and combining these measures could result in a more accurate model of arousal and performance. Moreover, the relationship is influenced by external factors such as the environment, and by individual differences like skill-experience level. These modulating factors can cause the relationship to change over time, for example as participants learn or adapt over the course a task. Therefore, understanding of arousal and performance benefits from a multidimensional approach that considers both internal and external influences.

Given the nuances of the Yerkes-Dodson law, the literature requires more conceptual replication studies. Future studies should aim to move on from the controlled laboratory setting and evaluate the theory in more naturalistic scenarios. These studies should aim to explore the contextual and individual characteristics when the law applies and when it does not. Additionally, researchers should consider qualitative research to explore the reasons behind the observed results, apply think-aloud methods to gain insights into participants' interpretations of questionnaires, or conduct systematic reviews of existing literature. By uncovering the subtle differences in the contexts, methodologies, and results of studies that support the theory and those that do not, and by exploring the reasons behind the observed results, researchers can better illustrate the relationships between variables such as arousal and performance in sequence learning.

References

Abrahamse, E. L., Ruitenberg, M. F. L., Kleine, E. d., & Verwey, W. B. (2013). Control of automated behavior: insights from the discrete sequence production task. *Frontiers in Human Neuroscience*, *7*, 1-16. https://doi.org/10.3389/fnhum.2013.00082

Akizuki, K., & Ohashi, Y. (2015). Measurement of functional task difficulty during motor learning: What level of difficulty corresponds to the optimal challenge point? *Human Movement Science*, *43*, 107-117. https://doi.org/10.1016/j.humov.2015.07.007

Ashcraft, M., & Radvansky, G. A. (2013). *Cognition*. Pearson New International.

Bond, I. G., Machida, K., & Johnson, K. A. (2023). Daily arousal variation has little effect on sustained attention performance. *Current Psychology*, *43*(3), 2690–2703. https://doi.org/10.1007/s12144-023-04473-9

Chan, R. W., Lakomski, V., Pannermayr, J. V. R., Wiechmann, E., Klooster, J.-W. J. R. v. t., & Verwey, W. B. (2025). Multimodal mobile brain and body imaging for quantification of dance motor sequence learning. *MethodsX*, *14*, 1-11. https://doi.org/10.1016/j.mex.2025.103324

Chan, R. W., Lushington, K., & Immink, M. A. (2018). States of focused attention and sequential action: A comparison of single session meditation and computerised attention task influences on top-down control during sequence learning. *Acta Psychologica*, *191*, 87-100. https://doi.org/10.1016/j.actpsy.2018.09.003

Chang, Y.-K., Etnier, J. L., & Barella, L. A. (2009). Exploring the relationship between exercise-induced arousal and cognition using fractionated response time. *Research Quarterly For Exercise And Sport*, *80*(1), 78-86. https://doi.org/10.1080/02701367.2009.10599532

Corbett, M. (2015). From law to folklore: Work stress and the Yerkes-Dodson Law. *Journal of Managerial Psychology*, *30*(6), 741-752. https://doi.org/10.1108/JMP-03-2013-0085

Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. *The Lancet Neurology*, *12*(3), 160-173. https://doi.org/10.1016/B978-0-444-53491-0.00006-7

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245-260. https://doi.org/10.1023/A:1019637632584

Gopher, D., & Donchin, E. (1986). Workload - An exmanination of the concept. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of perception and performance* (Vol. 2, pp. 1-49). John Wiley & Sons.

Grier, R. A. (2015). How high is high? A metanalysis of NASA TLX global workload scores. *Human Factors and Ergonomics Society*, *59*(1), 1727-1731. https://doi.org/10.1177/1541931215591373

Guadagnoli, M., & Lee, T. (2004). Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning. *Journal of Motor Behavior*, *36*(2), 212-224. https://doi.org/10.3200/JMBR.36.2.212-224

Halperin, I., Ramsay, E., Philpott, B., Obolski, U., & Behm, D. G. (2020). The effects of positive and negative verbal feedback on repeated force production. *Physiology & Behavior*, *225*, 1-5. https://doi.org/10.1016/j.physbeh.2020.113086

Hancock, P. A., & Ganey, H. C. N. (2003). From the Inverted-U to the Extended-U: The Evolution of a Law of Psychology. *Journal of Human Performance in Extreme Environments*, *7*(1), 4-14. https://doi.org/10.7771/2327-2937.1023

Hart, S. G. (2006). Nasa-task load index (Nasa-TLX); 20 years later. *Proceedings Of The Human Factors And Ergonomics Society Annual Meeting*, *50*(9), 904-908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hoonakker, P., Carayon, P., Gurses, A., Brown, R., McGuire, K., Khunlertkit, A., & Walker, J. M. (2011). Measuring workload of ICU nurses with a questionnaire survey: the NASA Task Load Index (TLX). *IIE Transactions on Healthcare Systems Engineering*, *1*(2), 131-143. https://doi.org/10.1080/19488300.2011.609524

Lowndes, B. R., Forsyth, K. L., Blocker, R. C., Dean, P. G., Truty, M. J., Heller, S. F., Blackmon, S., & Hallbeck, M. S. (2020). NASA-TLX Assessment of Surgeon Workload Variation Across Specialties. *Annals of Surgery*, *271*(4), 686-692. https://doi.org/10.1097/SLA.0000000000003058

Martella, D., Aldunate, N., Fuentes, L. J., & Sánchez-Pérez, N. (2020). Arousal and Executive Alterations in Attention Deficit Hyperactivity Disorder (ADHD). *Frontiers in Psychology*, *11*, 1-7. https://doi.org/10.3389/fpsyg.2020.01991

Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2015). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *39*, 1-75. https://doi.org/10.1017/s0140525x15000667

Mauss, I. B., & Robinson, M. D. (2008). Measures of emotion: A review. *Cognition & Emotion*, *23*(2), 209-237. https://doi.org/10.1080/02699930802204677

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*(1), 1-28. https://doi.org/10.1146/annurev.neuro.27.070203.144157

Muse, L. A., Harris, S. G., & Feild, H. S. (2003). Has the Inverted-U Theory of Stress and Job Performance Had a Fair Test? *Human Performance*, *16*(4), 349-364. https://doi.org/10.1207/s15327043hup1604_2

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040-1046. https://doi.org/10.1038/nn.3130

Palumbo, A., Groves, K., Munoz-Vidal, E. L., Turry, A., Codio, R., Raghavan, P., Schambra, H., Voebel, G. T., & Ripollés, P. (2024). Improvisation and live accompaniment increase motor response and reward during a music playing task. *Scientific Reports*, *14*(1), 1-15. https://doi.org/10.1038/s41598-024-62794-6

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(1), 73-89. https://doi.org/10.1146/annurev-neuro-062111-150525

Raw, R. K., Wilkie, R. M., Allen, R. J., Warburton, M., Leonetti, M., Williams, J. H. G., & Mon-Williams, M. (2019). Skill acquisition as a function of age, hand and task difficulty: Interactions between cognition and action. *PLOS ONE*, *14*(2), 1-27. https://doi.org/10.1371/journal.pone.0211706

Rhein, Z., & Vakil, E. (2017). Motor sequence learning and the effect of context on transfer from part-to-whole and from whole-to-part. *Psychological Research*, *82*(3), 448-458. https://doi.org/10.1007/s00426-016-0836-4

Rietschel, J. C., Goodman, R. N., King, B. R., Lo, L.-C., Contreras-Vidal, J. L., & Hatfield, B. D. (2010). Cerebral cortical dynamics and the quality of motor behavior during social evaluative challenge. *Psychophysiology*, *48*(4), 479-487. https://doi.org/10.1111/j.1469-8986.2010.01120.x

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, *57*(3), 493-502. https://doi.org/10.1037/0022-3514.57.3.493

Said, S., Gozdzik, M., Roche, T., Braun, J., Rössler, J., Kaserer, A., Spahn, D. R., Nöthiger, C., & Tscholl, D. W. (2020). Using the Raw NASA Task Load Index Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: A Meta-Analysis Using a Mixed Model. (Preprint). *Journal of Medical Internet Research*, *22*(9), 1-13. https://doi.org/10.2196/19472

Shuggi, I. M., Oh, H., Shewokis, P. A., & Gentili, R. J. (2017). Mental workload and motor performance dynamics during practice of reaching movements under various levels of task difficulty. *Neuroscience*, *360*, 166-179. https://doi.org/10.1016/j.neuroscience.2017.07.048

Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework (NASA/TM-2004-212824).* NASA Ames Research Center Retrieved from https://www.academia.edu/110222679/Stress_Cognition_and_Human_Performance_A_Literature_Review_and_Conceptual_Framework?

Valencia, C., Currier, K., Lindsay, S., Lemperis, P., Hughes, T., & Sowers, J. (2018). Levels of arousal in positive moods: Effects on motor performance. *The Open Journal of occupational Therapy*, *6*(3), 1-10. https://doi.org/10.15453/2168-6408.1447

Verwey, W. B. (2023). C-SMB 2.0: Integrating over 25 years of motor sequencing research with the Discrete Sequence Production task. *Psychonomic Bulletin & Review*, *31*(5), 931–978. https://doi.org/10.3758/s13423-023-02377-0

Vidoulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 243-273). John Wiley & Sons. https://doi.org/10.1002/0470048204.ch9

Wang, C.-A., Baird, T., Huang, J., Coutinho, J. D., Brien, D. C., & Munoz, D. P. (2018). Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Frontiers in Neurology*, *9*, 1-13. https://doi.org/10.3389/fneur.2018.01029

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering Psychology and Human Performance* (4th ed.). Psychology Press.

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202-217. https://doi.org/10.1037/teo0000137

Winton, W. M. (1987). Do introductory textbooks present the Yerkes-Dodson Law correctly? *American Psychologist*, *42*(2), 202-203. https://doi.org/10.1037/0003-066X.42.2.202

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit Formation. *Journal of Comparative Neurology and Psychology*, *459-482*. https://doi.org/10.1002/cne.920180503

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, *5*(5), 267-271. https://doi.org/10.1097/SIH.0b013e3181e3f329

Appendix A

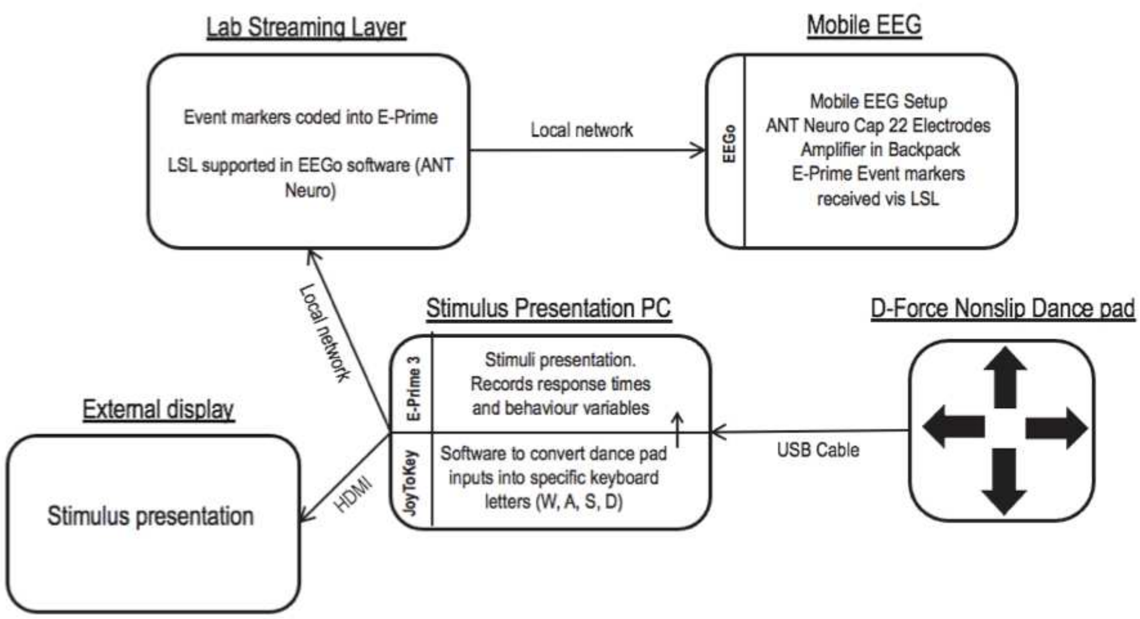
AI Disclaimer:

During the preparation of this work, I used ChatGPT and Grammarly in order to brainstorm and ideate, develop and test argumentations, copy edit including minor revisions for conciseness and clarity of writing, receive feedback on the structure and flow of the text, and assist in programming and debugging of code in R. After using these tools, I reviewed and edited the content as needed and take full responsibility for the content of the work.

Appendix B

Figure 15

Synchronization of Device Protocols



Originally developed by Chan et al. (2022)

Figure 16

The Ethical Review

Status: Positive advice by reviewer

The BMS ethical committee / Domain Humanities & Social Sciences has assessed the ethical aspects of your research project. Based on the information you provided, the committee does not have any ethical concerns regarding this research project.

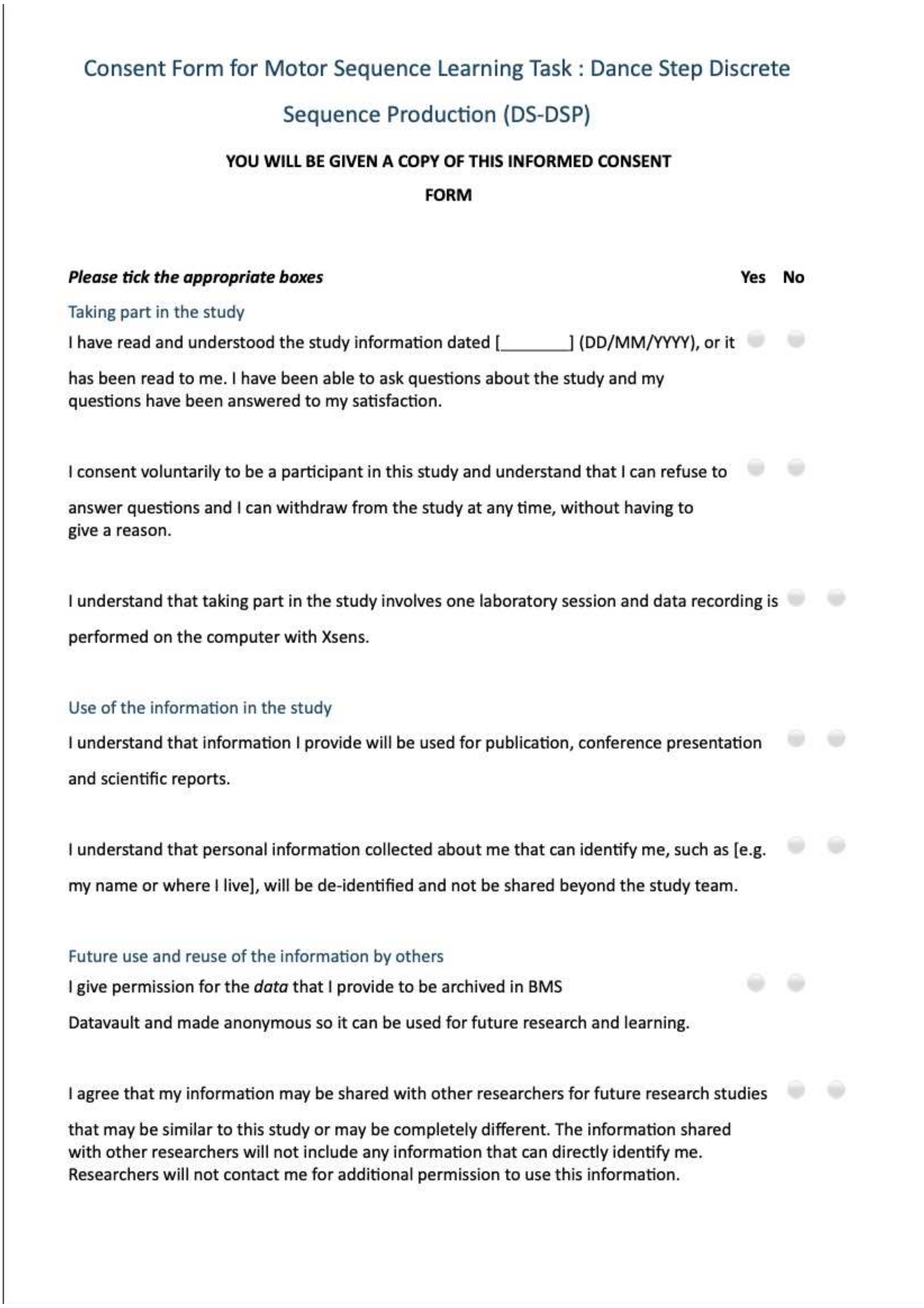
It is your responsibility to ensure that the research is carried out in line with the information provided in the application you submitted for ethical review. If you make changes to the proposal that affect the approach to research on humans, you must resubmit the changed project or grant agreement to the ethical committee with these changes highlighted.

Moreover, novel ethical issues may emerge while carrying out your research. It is important that you re-consider and discuss the ethical aspects and implications of your research regularly, and that you proceed as a responsible scientist.

Finally, your research may be subject to research compliance regulations such as the EU General Data Protection Regulation (GDPR), Codes of Conduct at UT related to (Scientific)Integrity or other codes of conduct that are applicable in your field, and the obligation to report a security incident (data breach or otherwise) at the UT.

Figures 17&18

The Informed Consent



Consent Form for Motor Sequence Learning Task : Dance Step Discrete Sequence Production (DS-DSP)

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated [_______] (DD/MM/YYYY), or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.		
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.		
I understand that taking part in the study involves one laboratory session and data recording is performed on the computer with Xsens.		
Use of the information in the study		
I understand that information I provide will be used for publication, conference presentation and scientific reports.		
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will be de-identified and not be shared beyond the study team.		
Future use and reuse of the information by others		
I give permission for the *data* that I provide to be archived in BMS Datavault and made anonymous so it can be used for future research and learning.		
I agree that my information may be shared with other researchers for future research studies that may be similar to this study or may be completely different. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information.		

I give the researchers permission to keep my contact information and to contact me for future research projects.

Signatures

____________________ ____________________ ________

Name of participant [printed] Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

____________________ ____________________ ________

Researcher name [printed] Signature Date

Study contact details for further information: p.schwarzmann@student.utwente.nl or c.erdogan@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl

Appendix C

Table 5

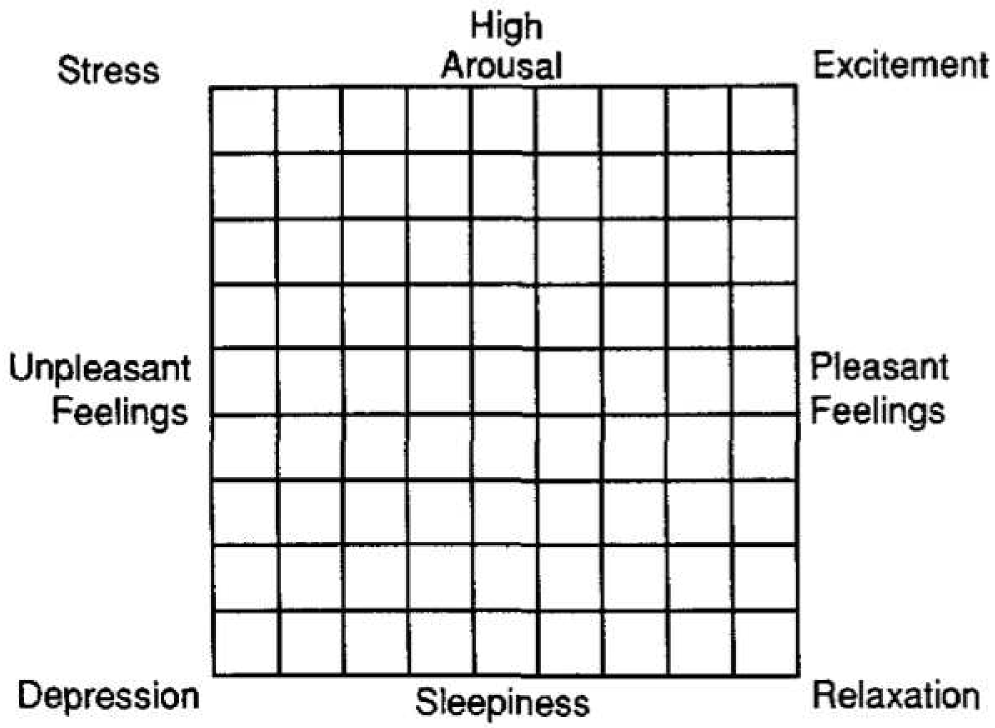
The Definitions of the NASA-TLX Subdimensions

Subdimension	Wording
Mental Demand	*How mentally demanding was the task?*
Physical Demand	*How physically demanding was the task?*
Temporal Demand	*How hurried or rushed was the pace of the task?*
Performance	*How successful were you in accomplishing what you were asked for?*
Effort	*How hard did you have to work to accomplish your level of performance?*
Frustration	*How insecure, discouraged, irritated, stresses, and annoyed were you?*

*Originally proposed by Hart & Staveland (1988)

Figure 19

The Affect Grid of Pleasure and Arousal



*Originally proposed by Russell et al., (1989)

Appendix D

Figure 20

Histogram & Q-Q Plot of Residuals: RT by Nominal Task Difficulty

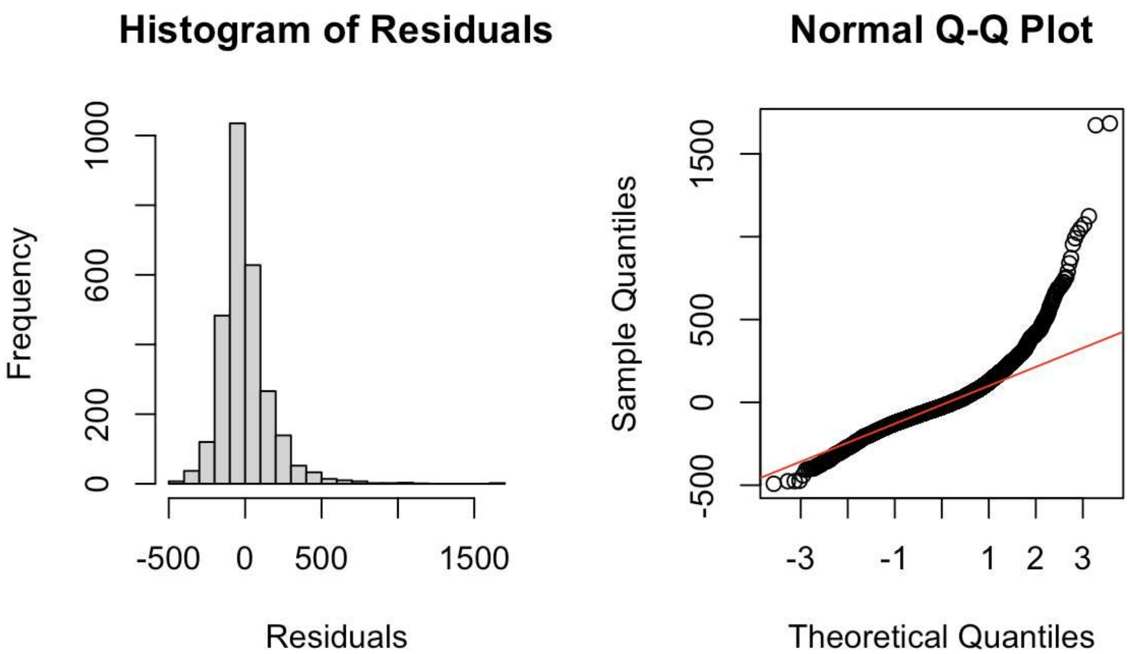


Figure 21

Histogram & Q-Q Plot of Residuals: Arousal by Nominal Task Difficulty

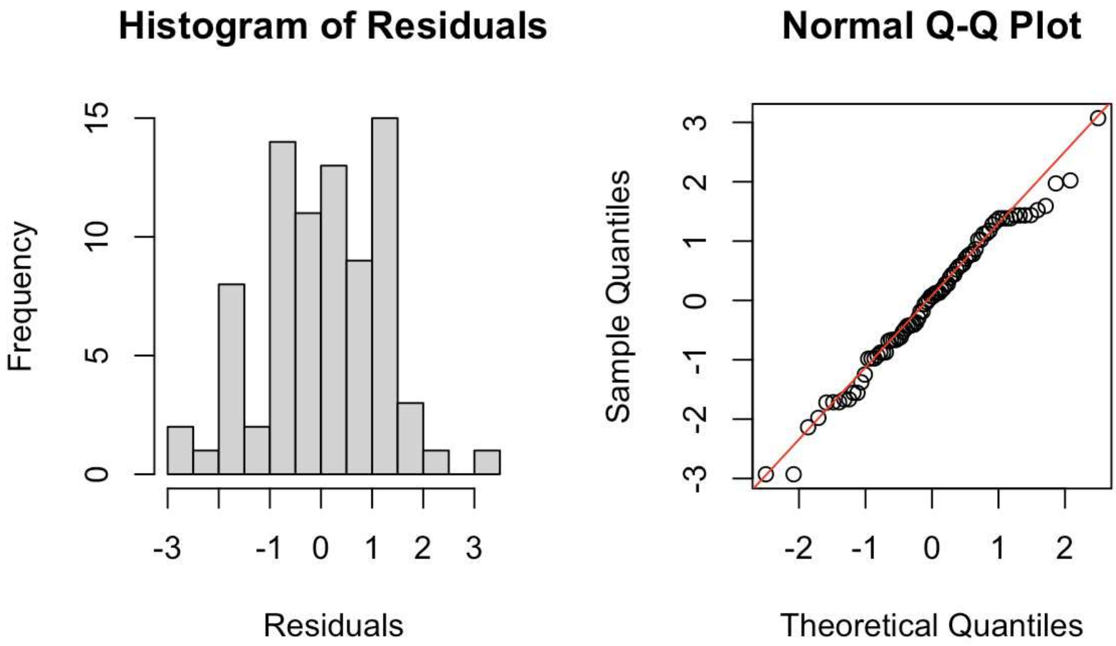


Figure 22

Histogram & Q-Q Plot of Residuals: Pleasure by Nominal Task Difficulty

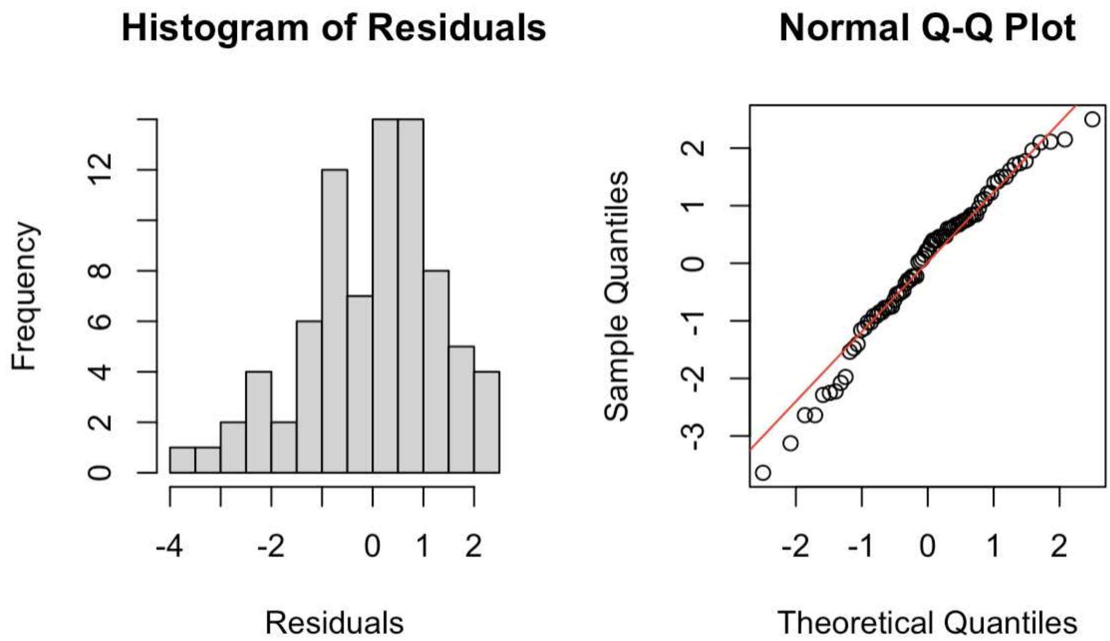


Figure 23

Histogram & Q-Q Plot of Residuals: RTLX by Nominal Task Difficulty

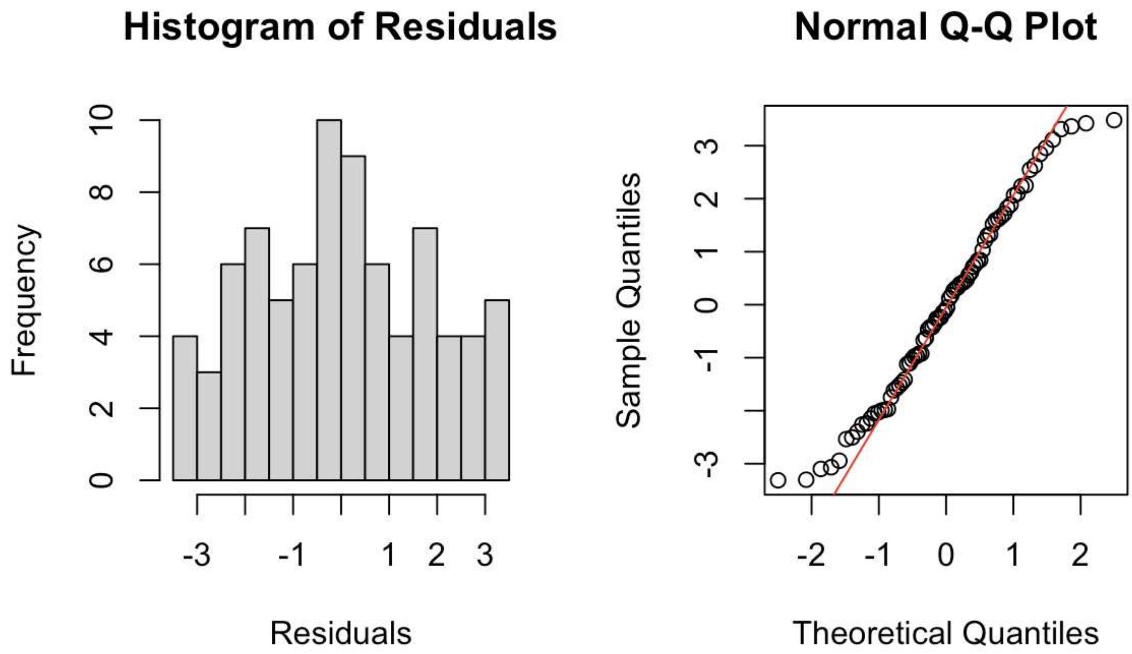


Figure 24

Histogram & Q-Q Plot of Residuals: Mental Demand by Nominal Task Difficulty

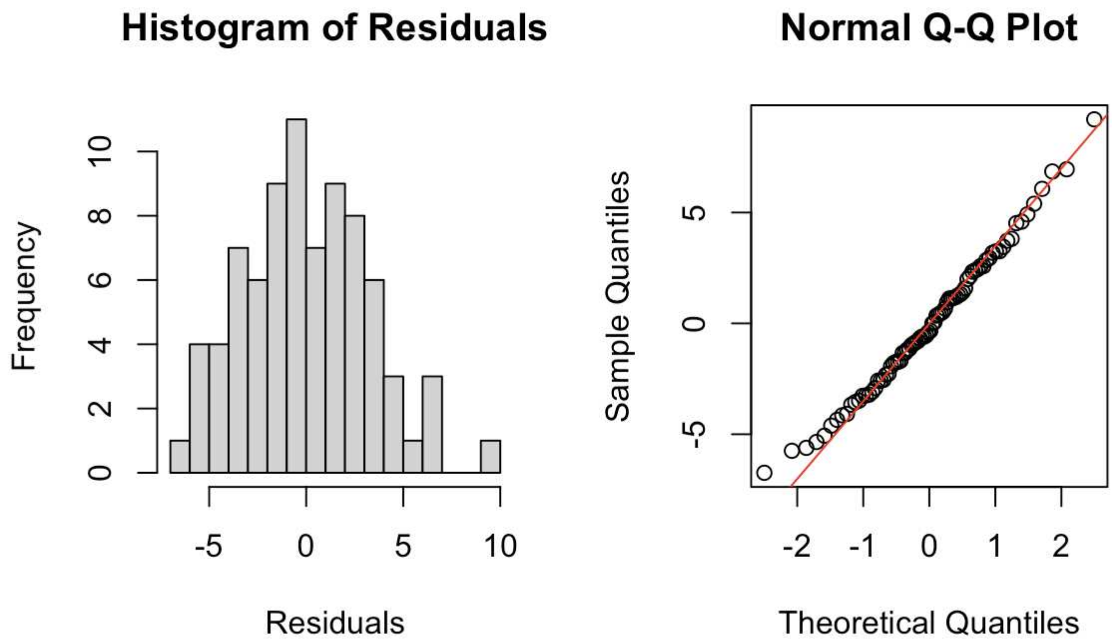


Figure 25

Histogram & Q-Q Plot of Residuals: Physical Demand by Nominal Task Difficulty

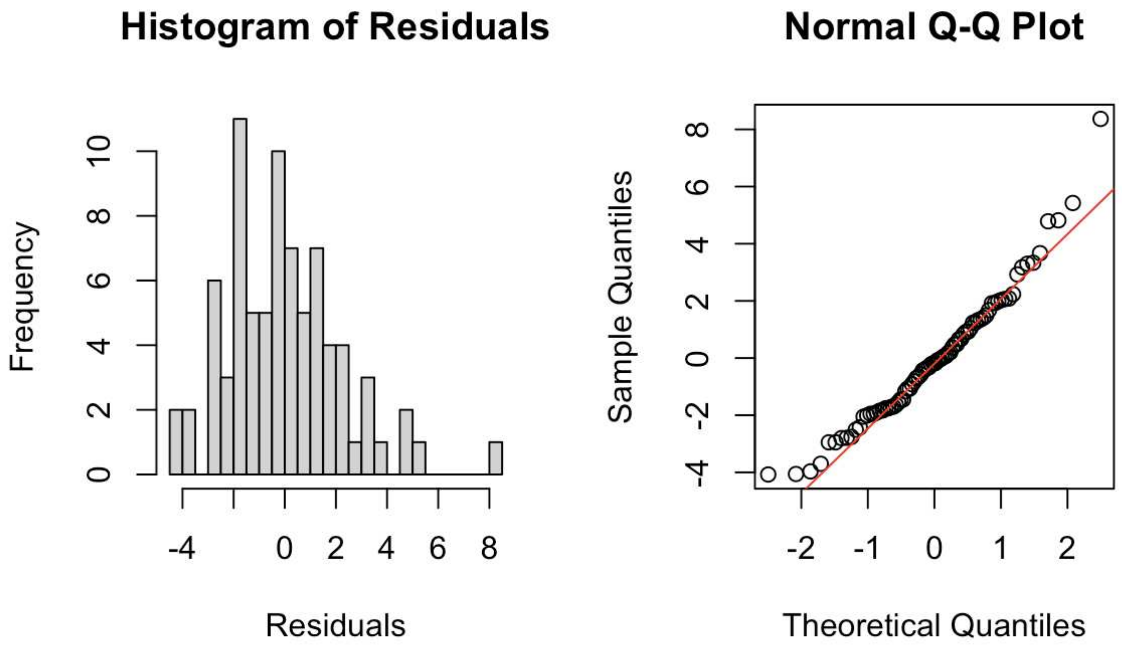


Figure 26

Histogram & Q-Q Plot of Residuals: Temporal Demand by Nominal Task Difficulty

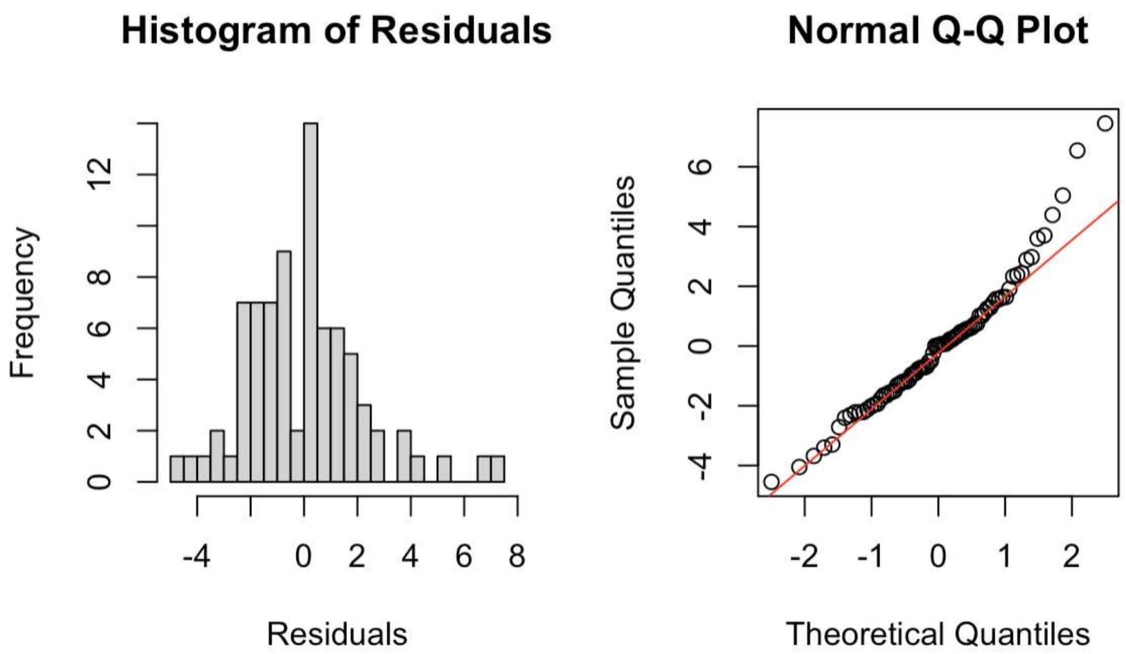


Figure 27

Histogram & Q-Q Plot of Residuals: Performance by Nominal Task Difficulty

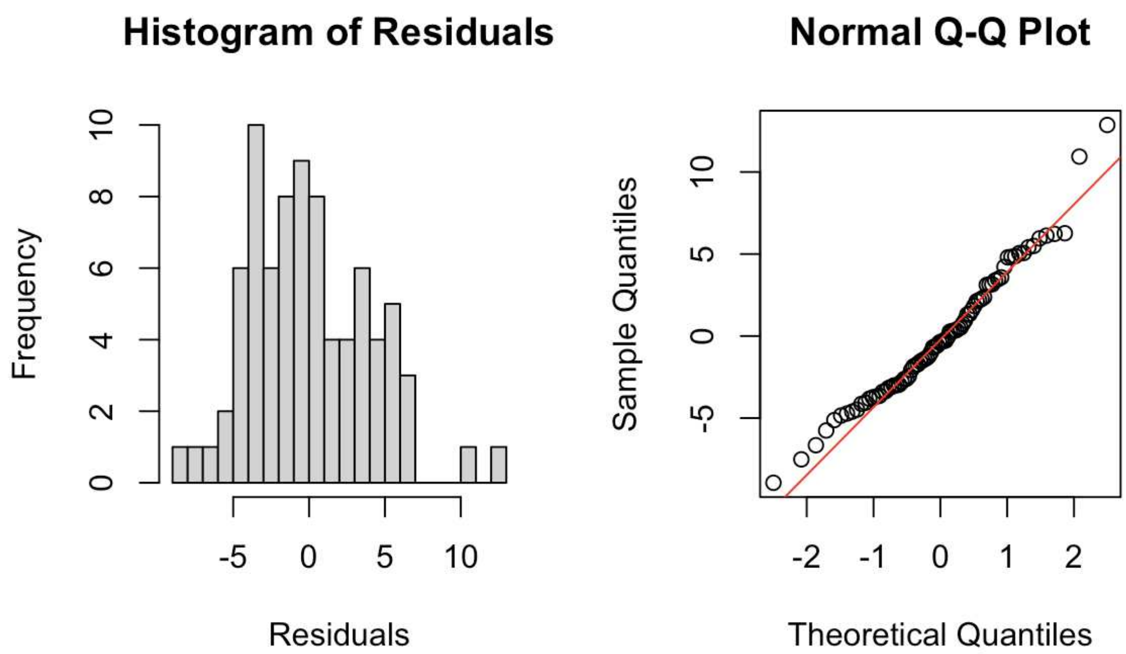


Figure 28

Histogram & Q-Q Plot of Residuals: Effort by Nominal Task Difficulty

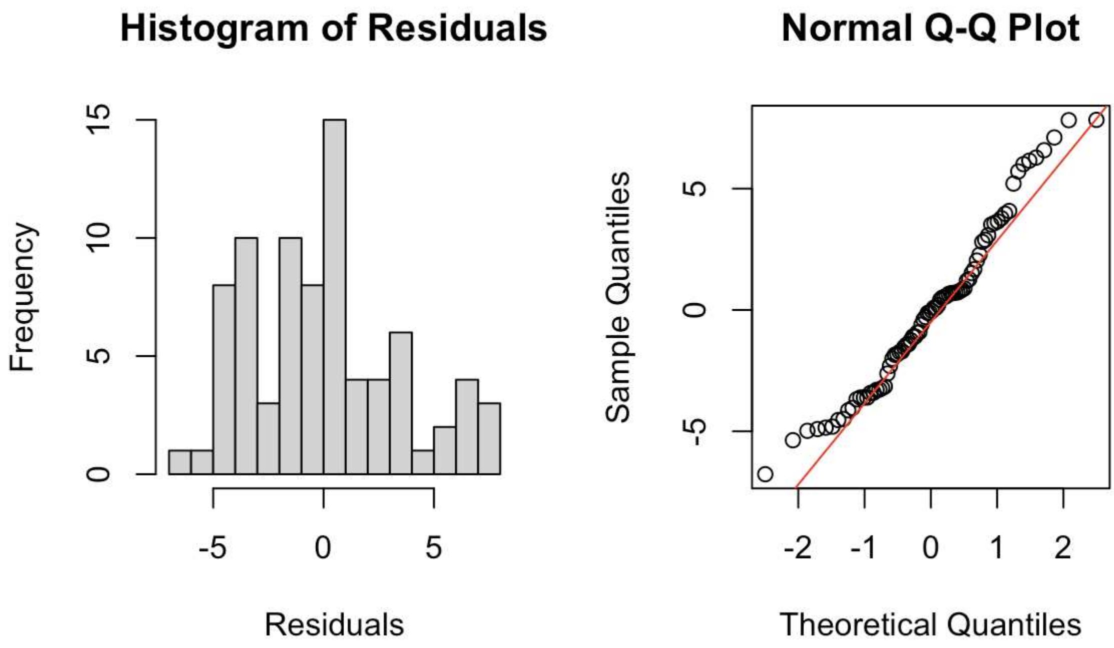
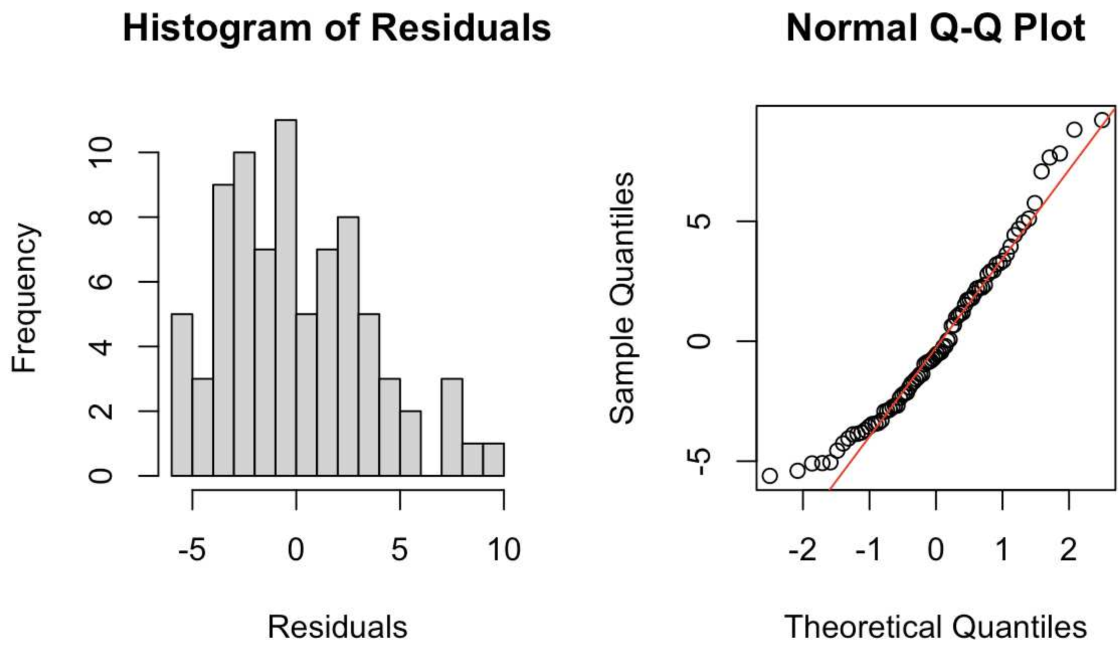


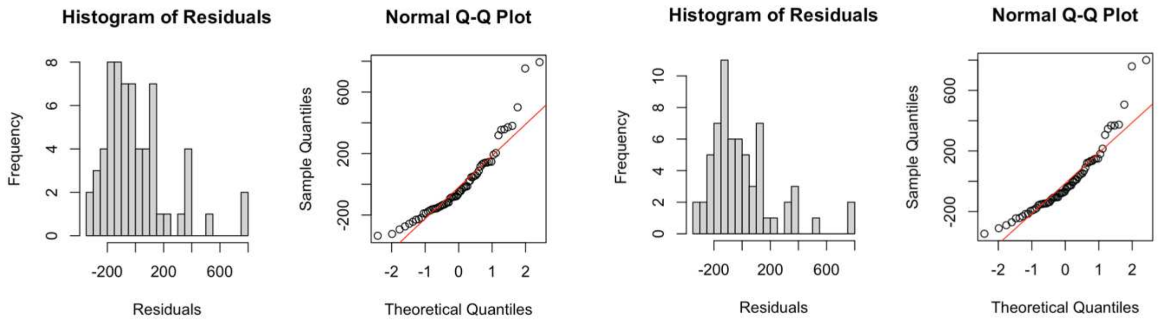
Figure 29

Histogram & Q-Q Plot of Residuals: Frustration by Nominal Task Difficulty



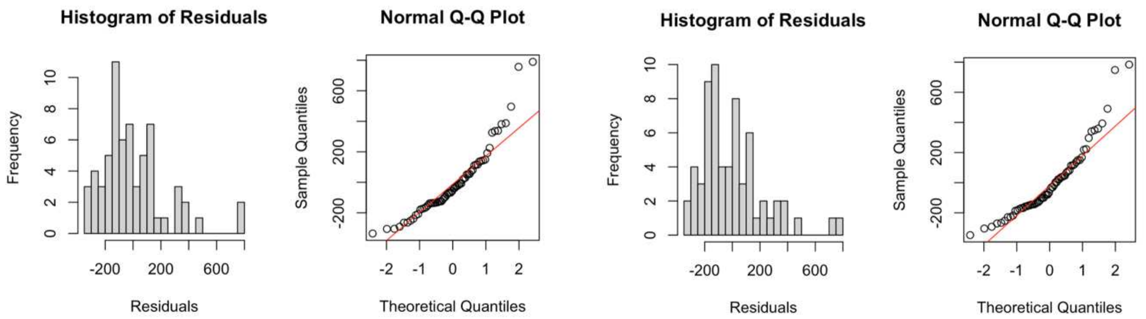
Figures 30&31

Histogram & Q-Q Plot of Residuals: Arousal by RT



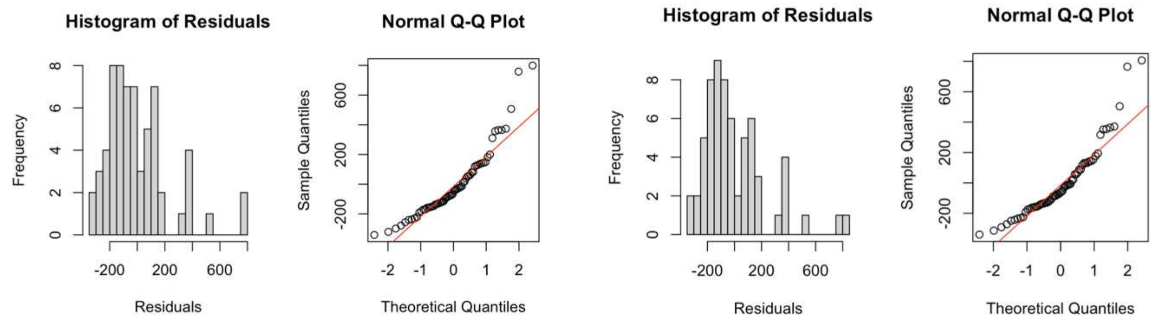
Figures 32&33

Histogram & Q-Q Plot of Residuals: Pleasure by RT



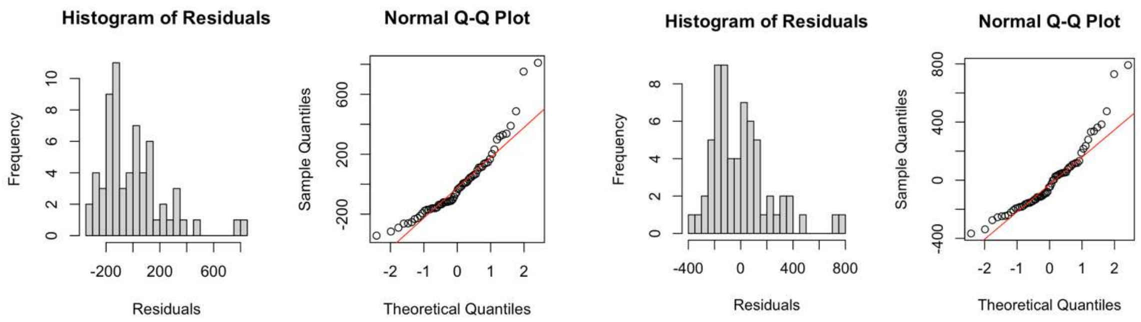
Figures 34&35

Histogram & Q-Q Plot of Residuals: RTLX by RT



Figures 36&37

Histogram & Q-Q Plot of Residuals: Mental Demand by RT



Figures 38&39

Histogram & Q-Q Plot of Residuals: Physical Demand by RT

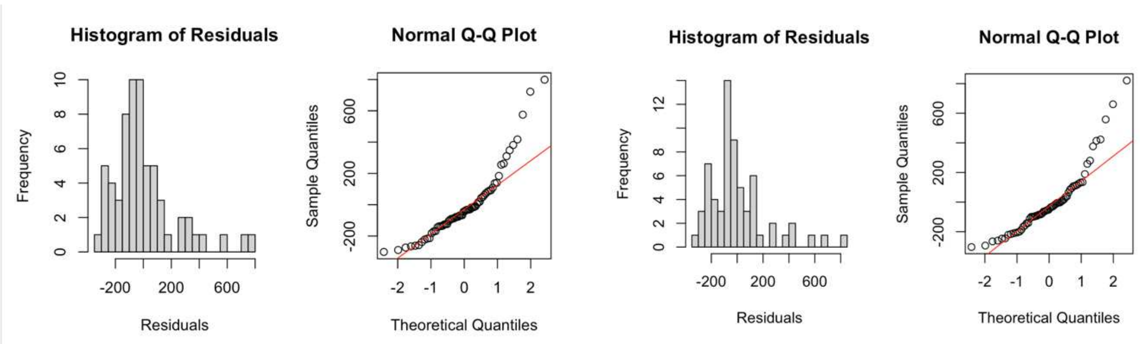
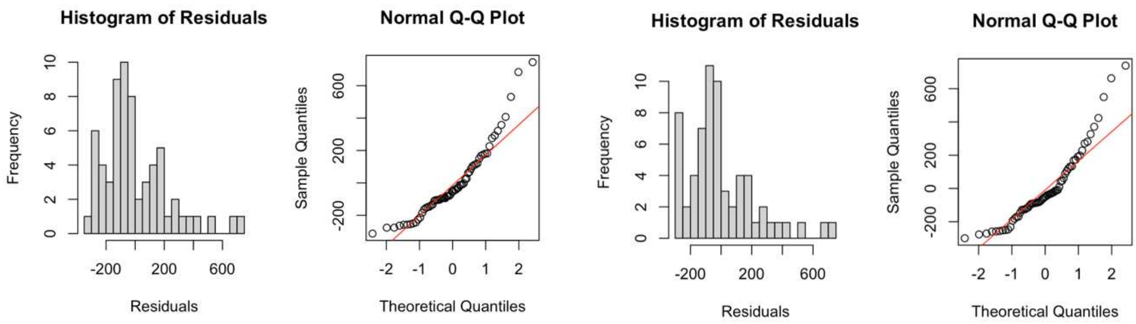


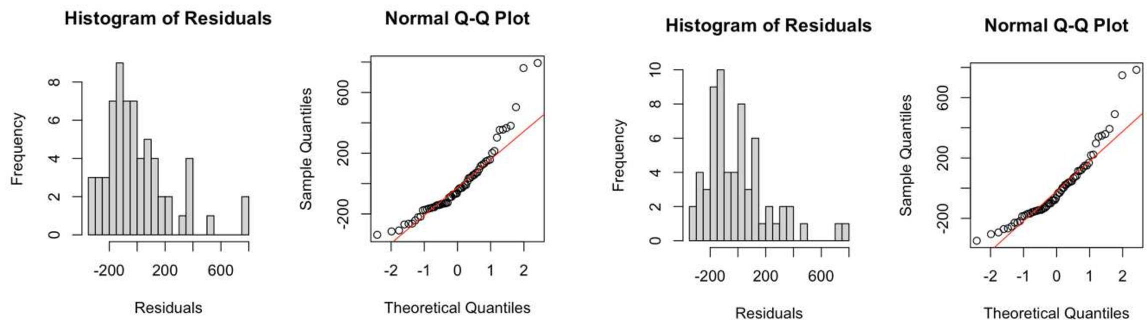
Figure 40&41

Histogram & Q-Q Plot of Residuals: Temporal Demand by RT



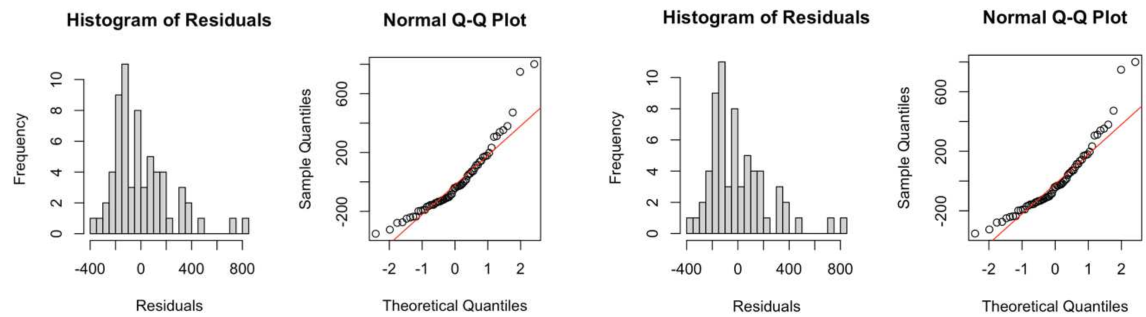
Figures 42&43

Histogram & Q-Q Plot of Residuals: Performance by RT



Figures 44&45

Histogram & Q-Q Plot of Residuals: Effort by RT



Figures 46&47

Histogram & Q-Q Plot of Residuals: Frustration by RT