2024

WHAT PSYCHOTHERAPY DOSE IS GOOD-ENOUGH? COMPARING THEORETICAL MODELS BY WEEKS AND SESSIONS

Ashley Rottkamp

WHAT PSYCHOTHERAPY DOSE IS GOOD-ENOUGH? COMPARING THEORETICAL MODELS BY WEEKS AND SESSIONS

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PSYCHOLOGY

to the faculty of the

DEPARTMENT OF PSYCHOLOGY

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY
New York
by
Ashley Rottkamp

Date Submitted 5/2/2024 | Date Approved 5/14/2024

____
Ashley Rottkamp

____
Dr. Raymond DiGiuseppe

ABSTRACT

WHAT PSYCHOTHERAPY DOSE IS GOOD-ENOUGH? COMPARING THEORETICAL MODELS BY WEEKS AND SESSIONS

Ashley Rottkamp

To determine optimal psychotherapy doses, researchers commonly rely on one of two competing theories, the Dose-Effect (DE) or the Good-Enough Level (GEL) models, to guide their research. Which model one selects is clinically meaningful, as the models have conflicting evidentiary support. Thus, it is essential to ascertain which theoretical model best fits empirical data. The present study compares the applicability and fit of the DE and GEL theories using total sessions and weeks in treatment as predictors within a community-based, unlimited-duration psychotherapy clinic. Significant determinants of treatment success were investigated. Adult participants' symptomatology was periodically assessed using the Outcome Questionnaire 45, Second Edition (OQ 45.2). A subsample of clients who completed over two outcome measures without a significant lapse in treatment (n = 311) was analyzed via multilevel modeling. Linear, log-linear, quadratic, and cubic models were constructed and evaluated within each theoretical framework. The superior models from each theory (DE, traditional GEL, and modified GEL) were compared, and the predictive value of using sessions versus weeks was examined. Sociodemographic and clinical variables were integrated into the best-fitting model to explore potential interaction effects. Further, determinants of treatment success and deterioration were analyzed for the full sample of all clients treated (n = 434) including those with early drop out. The findings of this study suggest that the traditional,

log-linear GEL model provided the best relative fit to the data for both week and session-based predictors. When compared, week-based predictors had a better model fit and were suggested to explain more of the variability than session-based predictors; however, potential inaccuracies in measuring session variables may have affected these results. Variables related to depressive symptoms (namely, hopelessness, suicidality, previous suicide attempts, and attention problems) showed a significant, positive correlation with initial symptom scores and enhanced the fit of the traditional GEL model. Furthermore, factors including low emotional stability and high symptom scores at baseline and improvements in emotional stability, hope, gratitude, and quality of life during treatment were all associated with better treatment outcomes. Conversely, low patient ratings of the therapeutic alliance during initial sessions were linked to treatment deterioration and dropout.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Importance of Studying Model Fit

Many studies on psychotherapy effectiveness aim to identify the optimal level of session variables that predict treatment outcomes (Anderson & Lambert, 2001; Erekson et al., 2022; Reese et al., 2011; Robinson et al., 2020; Wolgast et al., 2005). These “session” variables include session number, length, and frequency within or across diagnoses (Anderson & Lambert, 2001; Erekson et al., 2022; Reese et al., 2011; Wolgast et al., 2005). Research in this area typically hinges on the foundational theories of either the Dose Effect (DE) or the Good Enough Level (GEL) models, which researchers use to guide hypotheses, methods, and interpretations of findings (Reese et al., 2011; Wolgast et al., 2005). While both theories have ample empirical support, they have significant differences (Baldwin et al., 2006; Bone et al., 2021; Robinson et al., 2019).

The DE model suggests that the likelihood of client change is correlated with and primarily attributed to the total number of therapy sessions attended by the client, and the model posits there are insignificant differences in change rates across clients (Howard et al., 1986). The DE model is often used to define an "optimal dose" of psychotherapy, or the total number of sessions required for 50% of patients to exhibit a significant change, or “effect,” of treatment (Howard et al., 1986; Wolgast et al., 2005). Previous research suggests that the optimal number of therapy sessions can vary widely, from 4 to 54 sessions, depending on the severity of the initial impairment (Gotaas et al., 2021; Levy et al., 2020; Lincoln et al., 2016; Robinson et al., 2019).

Conversely, the GEL model posits that clients experience varying rates of change and typically discontinue treatment once they determine a "good enough" improvement has been achieved for them, so a greater number of sessions does not equate to a higher likelihood of change (Barkham et al., 2006). Alternatively, this model proposes that clients who change quickly likely terminate treatment with fewer required sessions than those with slower rates of change (Barkham et al., 2006). Studies using this model often explore the optimal frequency or duration of sessions, finding that more frequent or longer sessions generally lead to a greater reduction in symptoms compared to their less frequent or shorter counterparts, even when the total treatment duration is equivalent (Barkham et al., 2006; Bruijniks et al., 2020; Cuijpers et al., 2013; Okun & Glidewell, 2020; Reese et al., 2011; Robinson et al., 2019).

The results of studies exploring optimal session variables are significant, as they can influence clinical practice by altering recommended treatment lengths and session frequencies. Furthermore, they may affect public health policies and insurance coverage limits, as well as guide the provision of counseling services in educational settings (Owen et al., 2014). Given that the choice between the DE and GEL models affects the method and interpretation of studies investigating optimal therapy practices and doses, it is essential to determine which model best fits and generalizes to empirical client data.

LITERATURE REVIEW

Core Assumptions of the Theoretical Models for Psychotherapy Effectiveness

Dose Effect (DE) Model

Howard et al. developed the DE model in 1986, establishing a theory to estimate the required number of psychotherapy sessions, or doses, for significant patient improvement or a significant effect. In their foundational study, researchers analyzed data from 299 patients at a Chicago psychiatric outpatient clinic. They conducted a meta-analysis involving 2,431 patients across various settings, including community mental health clinics, private practices, university counseling centers, university psychiatric centers, and a Veterans Affairs clinic. Their research yielded five primary findings.

First, Howard et al. (1986) found that there is a significant, positive correlation between the total number of sessions (dose) and the percentage of patients who experienced significant improvement (effect). Second, contrary to the assumption that each additional session uniformly benefits all patients, researchers findings align with the economic principle of the Law of Diminishing Returns, which posits that each additional unit of therapy (i.e., sessions) results in progressively smaller gains with the largest gains per session occurring early in treatment (Shepard & Fare, 1974). Third, Howard et al. (1986) proposed that a negatively accelerating improvement curve is also present for individual patient progress, so each additional psychotherapy session yields diminishing returns on client improvement. As such, when the relationship between sessions and patient improvement is graphed, the DE model posits that both the aggregate and individual patient improvement curves follow a negatively accelerating trend. Fourth, researchers found that across various session ranges, there were no significant differences in improvement percentages among patients regardless of their number of total sessions (i.e., whether patients attended 1-100 sessions, the percentage of patients who improved during 1-3 sessions ranged from 29%-38%, and the percentage of patients who improved during 4-7 sessions ranged from 48%-58%). This suggests a relatively consistent rate of change regardless of session number (Howard et al., 1986). Fifth, Howard et al. (1986) found that the optimal total number of sessions varies by diagnosis and symptom severity; specifically, researchers found that the patients with anxious or depressive disorders had an optimal dose of 8-13 sessions while patients diagnosed with "borderline-psychotic" disorders had an optimal dose of 13-52 sessions. However, the DE model does not propose that symptom severity or diagnosis significantly alters the other model aspects, including the rate of change.

Good Enough Level Model

Introduced by Barkham et al. in 1996, the GEL model contrasts with the DE model in five fundamental ways. First, unlike the DE model, the GEL model posits that rates of change vary significantly among patients, and patients terminate treatment once they achieve a "good-enough" level of improvement for themselves. Specifically, the GEL model proposes that individuals who respond quickly to treatment and have faster rates of change, referred to as early responders, experience earlier discharge (Barkhum et al., 2006). Secondly, the GEL model proposes that the negatively accelerating curve is only evident in aggregate data and is skewed by early responders who exit treatment sooner, thus leaving behind slower-responding patients (Barkhum et al., 2006). Third, the GEL model proposes that individual patient progress is linear as opposed to negatively accelerating (Barkhum et al., 1996). Fourth, the GEL model proposes that setting session limits on treatment can positively influence patients' rate of progress (Barkhum et al., 1996). Specifically, Barkhum et al. (2006) found that 212 patients treated with eight total sessions showed significantly more improvement at session eight (59%) than those treated with sixteen total sessions (40%). This further supports the GEL theory that increasing total session does not increase likelihood of patient improvement, as change rates primarily lead to improvements. Fifth, the GEL model acknowledges that individuals with different diagnoses and symptom severity exhibit different rates of change (Barkhum et al., 2006)

Table 1 summarizes the key differences between both models.

Studies Directly Comparing Both Models

Many studies directly compare the DE and GEL models by fitting them to actual client data. Nine such comparative studies are included in this literature review. Eight of these studies had a mean session number ranging from 3.7 to 12.16 sessions, while the ninth included patients with severe diagnoses treated with a mean of 52 sessions. The number of participants in each study ranged from 263 to 64,319. Multilevel modeling (or similar methods) was utilized to analyze the data, which allows researchers to nest individual observations (i.e., individual symptom scores) within individual clients over time without violating the independence assumptions for longitudinal data. All studies reported that psychotherapists used a variety of theoretical orientations for treatment (Baldwin et al., 2009; Falkenström et al., 2016; Lee et al., 2022; Nordmo et al., 2021; Reese et al., 2011; Owens et al., 2016; Schuler et al., 2021; Stultz et al., 2013).

Table 2 describes the characteristics of these studies.

Studies Predominantly Supporting the GEL Model

Despite almost all studies reporting some evidentiary support for both models, most studies (62.5%) concluded that the GEL model best fits their data, and they detailed several findings supporting that conclusion. First, of the studies supporting the GEL model, many of these studies found significant variability in change rates across participants, which often varied as a function of their total number of sessions attended throughout treatment (Baldwin et al., 2009; Falkenström et al., 2016; Lee et al., 2022; Owens et al., 2016; Reece et al., 2011). Specifically, participants with faster rates of improvement typically attended fewer sessions, while individuals with slower rates typically attended more sessions (Baldwin et al., 2009; Falkenström et al., 2016; Lee et al., 2022; Owens et al., 2016; Reece et al., 2011). Second and further consistent with the GEL model, most studies found a minimal and insignificant correlation between total sessions attended and the likelihood of clinical improvement, as many participants reached similar symptom levels upon termination regardless of their total session number (Baldwin et al., 2009; Lee et al., 2022; Owens et al., 2016). Third, several studies found that patients did not experience gains of diminishing returns; instead, their participants' change rates during treatment appeared linear (Baldwin et al., 2009; Nordmo et al., 2021; Reese et al., 2011), and in one study, the aggregate curve even appeared linear (Falkenström et al., 2016). Fourth, and most importantly, using MLM, these researchers found that the relative fit of the GEL model was significantly better than the DE model.

Although these studies concluded that the GEL model had the best fit, some of these studies found support for aspects of the DE model. For example, some researchers found that more sessions were associated with greater treatment gains and symptom reduction (Falkenström et al., 2016 & Lee et al., 2022). Additionally, consistent with the DE model, one study found that a negatively accelerating curve was present even when stratifying for the total number of sessions attended; thus, there were diminishing returns even after removing the effect of early responder dropout (Nordmo et al., 2021).

Modified GEL Model. Further, two studies explored how session frequency impacts change rates and total session numbers for participants by introducing a modified GEL model, as researchers hypothesized that session frequency moderates the relationship between total sessions and change rates (Baldwin et al., 2009 & Reece et al., 2011). Researchers' adaptation started with the traditional GEL, multi-level model and incorporated additional terms for session frequency and an interaction between individual client change rates and total session number. Findings from these studies indicated that the modified GEL model had a significantly better fit than the DE and traditional GEL models. As such, researchers concluded that session frequency significantly influences participants' change rates, so clients who attend sessions more frequently (e.g., weekly over twice monthly) experience more rapid, significant improvement than their counterparts (Baldwin et al., 2009 & Reese et al., 2011).

Studies Predominantly Supporting the DE Model

One study found the DE model to have the best fit (12.5%) while two others concluded mixed findings with vast evidence for the DE model (25%). Consistent with the DE model, these studies found diminishing marginal returns for each additional session of psychotherapy across both aggregate and individual client data represented by a negatively accelerating curve (Niileksela et al., 2021; Schuler et al., 2020; Stultz et al., 2013). Additionally, researchers found a positive, significant correlation between the total number of sessions in treatment and the percentage of clients achieving significant improvement (Niileksela et al., 2021; Schuler et al., 2020; Stultz et al., 2013). Conversely, in support of the GEL model, two studies found significant differences in change rates across participants (Niileksela et al., 2021 & Stultz et al., 2013), while one found insignificant differences consistent with the DE model (Schuler et al., 2020)

Systematic Reviews and Meta-Analyses of DE & GEL Models

In comprehensive reviews, Bone et al. (2021) and Robinson et al. (2019) analyzed the DE or GEL literature across 15-26 studies. Both reviews found evidence supporting aspects of both theoretical models. Consistent with the DE model, aggregate and individual participant data often reflected diminishing returns often reflected by a curvilinear relationships between sessions and patient improvement. Further consistent with the DE model, many researchers identified an optimal dose for treatment responders. Conversely, consistent with the GEL model and across the majority of studies, researchers found that participants had significantly different individual change rates during treatment. Many studies found that the relationship between sessions and individual patient improvement reflected a linear relationship in longer treatments and a nonlinear (i.e., curvilinear) relationship in shorter treatments. Additionally, the majority of studies found a significant, negative correlation between change rates and total sessions, as those with faster rates had fewer sessions of treatment. As such, both authors found evidentiary support for elements of both the GEL and DE models (Bone et al., 2021; Robinson et al., 2019).

Limitation of Previous Studies

The Impact of Time in Treatment and Session Frequency

Previous studies comparing the DE and GEL theories have several limitations which the present study aims to address. The first limitation is related to the sole reliance on sessions as a predictor which disregards and overlooks the impact of missed, infrequent, and irregular sessions. These missed, infrequent, and irregular sessions often occur at high rates in treatment, so they likely impact theoretical model fit. As such, it is important to consider these factors, as they can significantly impact treatment outcomes in the following ways (Reese et al., 2011).

First, the literature suggests that missed psychotherapy appointments are common. Many studies assessing client absences suggest that approximately 20% of appointments are missed, canceled, or not attended. These rates are consistent across short-term [Mean (M_{Sessions} = 5-6)] and longer-term (M_{Sessions} = 20) treatments (Kivlighan et al., 2019; Stiles et al., 2003; Xiao et al., 2017). As such, ignoring missed sessions may affect the accuracy of model comparisons.

Second, according to many studies, the frequency of psychotherapy is clinically meaningful, as more frequent sessions are associated with significantly better outcomes than less frequent sessions (Bruijniks et al., 2020; Chase et al., 2012). In a meta-regression analysis consisting of 70 studies and over 5,400 patients, Cuijpers et al. (2013) found a strong association between the frequency of sessions per week and the effect size. Specifically, researchers found that increasing participants' sessions from 1 to 2 each week increased their effect size significantly ($g = 0.45$, $p < .001$) even while keeping the total number of sessions throughout treatment constant. Additionally, these findings were not true across other time or session variables (i.e., time in treatment, total sessions, etc.), further suggesting the frequency of sessions is especially important. As such, it is important to consider the impact of session frequency on model fit.

Third, sometimes, psychotherapy sessions are provided on an inconsistent or irregular basis. In a previously cited DE and GEL comparison study consisting of 13,647 veterans by Lee et al. (2022), patients from their primary care clinics attended an average of 3.7 sessions but remained in treatment for an average of 71.8 days (10 weeks) while patients in their specialty mental health care clinics attended an average of 6.6 sessions but remained in treatment for an average of 103.4 days (14 weeks). As such, on average, participants in their study attended sessions every 2 to 3 weeks, and the irregularity of sessions likely impacted treatment success. As such, this irregularity likely impacts model fit and conclusions.

To address these gaps, the present study will include a modified GEL model and a predictor of time (i.e., weeks) to assess treatment success. As previously mentioned, while most DE and GEL comparison studies did not investigate the impact of time on model fit, two studies considered and assessed the impact of treatment frequency through this modified GEL model (Baldwin et al., 2009; Reese et al., 2011). This model extended the traditional GEL model by including session frequency and an interaction term between individual client rate of change and total number of sessions. These researchers found that the modified GEL model had a significantly better fit to the data, suggesting that more frequent sessions (e.g., weekly) led to more rapid and significant improvements than less frequent sessions (e.g., bi-monthly) (Baldwin et al., 2009; Reese et al., 2011). Therefore, this study adds to the literature by also investigating the fit of a modified GEL model on empirical data. Additionally, while previous studies focus solely on session as a dose, which can disregard the impact of time, the present study will include both session and time (i.e., weeks in treatment) as predictors to encapsulate therapeutic engagement and processes that occur within and outside of sessions.

Participant Homogeneity

Second, due to their homogeneous samples or restricted settings, many DE and GEL are limited in their generalizability to outpatient, community-based clinics (see Table 2). Specifically, four included studies were based in United States College Counseling Centers (US CCCs), which primarily service a uniform demographic of undergraduate and graduate students with minimal variability in their ages and education levels (Baldwin et al., 2009; Niileksela et al., 2021; Owen et al., 2006; Reese et al., 2011; Stultz et al., 2013). One study included data solely collected in US Veterans Affairs centers and included only veterans treated for depression. As such, this sample consists of predominantly male clients with similar diagnoses and job histories, so it is not representative of the patient population treated through a community-based or traditional outpatient clinic (Lee et al., 2022). Additionally, two European studies included participants from broader demographics but are still limited in their generalizability to community clinics in the United States (Falkenström et al., 2016; Nordmo et al., 2021). In Europe, psychological care is divided into three levels. Primary health care is often provided by general practitioners, secondary care is often provided on an outpatient basis by psychiatric service providers, and tertiary care is provided on an inpatient basis by psychiatric specialists (Jané-Llopis & Anderson, 2006). One European study only included patients in the primary and tertiary care levels, while the other restricted their sample to only participants with severe psychopathology receiving secondary care. United States (US) outpatient clinics likely serve patients that present similarly to those served within primary and secondary care settings in Europe, so both studies are limited in their generalizability, as they include ranges too limited or extreme for US clinics (Falkenström et al., 2016; Nordmo et al., 2021). Lastly, only one study within the review included a traditional US outpatient population (Schuler et al., 2020). Since the present study will include participants from a diverse community-based clinic, it will address previous studies' limitations to generalizability and will add to the robustness of the DE and GEL literature for community clinics.

Subclinical Participants

A third limitation of many DE and GEL comparison studies is their strict inclusion and exclusion criteria. Many of these studies excluded participants whose initial symptom scores were subclinical, as a score in the clinical range is needed to experience a clinically significant decline in symptoms (see Table 2; Baldwin et al., 2009; Falkenström et al., 2016; Lee et al., 2022; Owens et al., 2016; Schuler et al., 2021; Stultz et al., 2013). However, eliminating patients with low initial symptom scores from the model decreases generalizability to actual clients, whose initial scores vary.

Predictors of Treatment Success or Deterioration and Interaction Effects

Fourth, while many of these studies report the percentage of participants who reach reliably and clinically significant improvement (RCSI), they do not often detail the factors significantly associated with RCSI or negative outcomes. While some of the previously cited studies conducted exploratory analyses to analyze possible significant interactions or associations, most researchers did not find significant effects, which is likely attributed to low variability between their participants. Further, while these studies sometimes assess associations for participants with significant improvement, they often fail to consider factors associated with a worsening or deteroiation of symptoms. Previous literature suggests that approximately 8.2% of psychotherapy clients experience a reliably and clinically significant deterioration (RCSD) during treatment, and information contrasting those with RCSI and RCSD may show important associations of treatment outcomes (Hansen et al., 2002).

Previous literature suggests that client readiness and commitment variables can predict attrition rates and client outcomes, but the data is mixed (Falkenstrom et al., 2013). Specifically, some prior research has suggested that client attendance is positively correlated with an increased likelihood of staying in treatment; regular attendance is associated with stricter adherence to treatment plans and increased likelihood of achieving desired results; and client appraisals of their working alliance with their therapy provider significantly predicts long term outcomes (Beierl et al., 2021; Falkenstrom et al., 2013; Marker et al., 2013). Meanwhile, other studies suggest that patient and treatment characteristics (such as attendance rates, demographic variables, functional impairment, comorbidities, etc.) have insignificant or inconsistent effects on treatment success and failure across studies (Lindfors, 2022; Reuter et al., 2015). Additionally, some research suggests that moderate or functional levels of initial symptom distress and previous nonresponse in treatment are the most consistent predictors of nonresponse or psychotherapy failure (Reuter et al., 2015). Given the mixed findings, it is likely that further investigation is needed (Lindfors, 2022).

Statement of the Problem

The present study aimed to expand upon the literature by including several components.

- This study compared the fit of the DE, traditional GEL, and modified GEL models within a community-based sample with minimal inclusion and exclusion criteria (i.e., unlike previous literature, the sample included initially sub-clinical clients, treatment non-responders, etc.) to promote generalizability of the findings to traditional outpatient populations.
- This study compared the fit of the DE and GEL models using participants treated with open-ended psychotherapy provided by doctoral psychotherapists primarily using a cognitive-behavioral orientation to treatment.
- This study compared DE, traditional GEL, and modified GEL model fit using both sessions and weeks in treatment as predictor variables.
- This study analyzed and compared the predictive accuracy of session variables versus weeks in treatment variables within the best fitting model.
- This study assessed interaction effects within the model of best fit and identified predictor variables associated with significant treatment success and deterioration.

Research Questions

1. What theory of psychotherapy effectiveness (the DE, traditional GEL, or modified GEL theories) best fits the data from a community clinic using sessions and then weeks as predictor variables?

2. Which predictor, sessions or weeks, has the best relative fit to the superior theoretical model and has the greatest predictive accuracy?
3. What variables or factors interact with the above relationships in the model, and what are other factors that are significantly associated with treatment success and failure?

METHOD

Training Clinic

Archival data from a university-based, outpatient training clinic in a diverse and urban metropolitan city was used. Psychotherapy was provided year-round to university and community members on an income-based sliding scale. The providers were doctoral students studying school or clinical psychology who primarily used cognitive-behavioral approaches and worked under the supervision of licensed psychologists. Psychotherapy was open-ended, but providers were employed for 1-2 years, so some clients had multiple providers across the years.

Participants

From February 2005 to February 2024, 456 research-consenting adult clients (> 18 years) received psychotherapy and were discharged at a community-based, university training clinic. Of these clients, a sample and sub-sample of participants were created for the present study. Of all research-consenting clients treated at the clinic, the first sample had only one exclusion criterion, as this sample excluded participants who failed to complete any outcome measures (n = 22). All other participants were included, even if they only attended one session. As such, the first, primary sample consisted of 434 clients.

The second, smaller sub-sample had more stringent exclusion criteria. Specifically, this sub-sample excluded clients who failed to attend at least three sessions <u>or</u> failed to complete an outcome measure at least <u>three</u> times (n = 99), and clients were excluded if they had over 1 year between consecutive sessions (n = 29). This criterion aimed to minimize significant missing data and allow for appropriate modeling of change. After excluding clients, the sub-sample consisted of 311 participants, and this sample was used to conduct the multilevel modeling analysis.

Prior to treatment, participants consented to the use of their de-identified data in research, and the University's Institutional Review Board and the Psychological Center's Director of Research approved the study. Specific archival data used included the client's descriptive, preliminary information and the result of the client's Outcome Questionnaire-45.2 (OQ-45.2, Lambert et al., 1996) and Bi-Weekly Longitudinal (BIL; see Measures).

Descriptive information was collected for all clients via a preliminary phone screening, though reporting was non-mandatory. Of those who reported this information from the larger sample (n = 434), the study's sample consisted of predominantly female (60%) over male (30%) clients [10% did not report their gender] between the ages of 18 and 73 [Mean (M_{age}) = 33, Standard Deviation (SD) = 12.53]. Most clients were fluent in English (99.7%) and were raised in a monolingual household (78%), but some were not fluent (0.3%), had dual-language households (14%), or had households with a primary language other than English (8%). Services were provided on a sliding scale, and most clients received the lowest fee, which was given to clients whose family income was less than 30,000 per year, who were unemployed, or who were university students (58%). Alternatively, clients paid the highest fee if their family income surpassed $120,000 (12%). Remaining client fees were determined based on income intervals of $10,000 [$30,000-40,000 (9%); $40,000-$50,000 (7%); $50,000-$60,000 (5%); $60,000-70,000 (2%); $70,000-80,000 (2%); $80,000-$90,000 (>1%); $90,000-$100,000 (1%); and $100,000-110,000 (4%)].

Additional intake information was available for approximately 70% of clients. At the time of their preliminary phone screenings, most participants had a history of psychotherapy treatment (53%), while many did not (31%), and some participants were actively receiving treatment elsewhere (16%). Additionally, most clients were not prescribed psychiatric medication (75%), but some were (25%), approximately half of the participants endorsed a history of at least one traumatic event (50%), and some participants had a previous suicide attempt (8%). Client motivations for treatment varied, as some considered their problems emergent (14%), but most did not (86%), and very few sought treatments for substance use (1%) or legal reasons (1%). Further, presenting problems varied across clients and were determined based on client endorsements of symptoms. More than half of all participants endorsed possible anxiety (82%), depression (74%), hopelessness (67%), sleep problems (64%), attention problems (62%), obsessive thoughts (51%), and social problems (50%). Additionally, many participants reported difficulty with eating problems (45%), memory problems (46%), academic difficulties (33%), aggressive behavior (30%), suicidal thoughts (24%), hyperactivity (23%), and hearing voices (5%).

Measures

Participant referral, demographic, and diagnostic information were obtained during preliminary client phone screenings conducted by doctoral students. Participants' self-reported levels of distress were assessed using the Outcome questionnaire-45.2 (OQ-45.2, Lambert et al., 1996) at intake and then at bi-weekly intervals. The OQ 45.2 is a routine outcome measure utilized regularly to assess clients' levels of distress and participant changes throughout psychotherapy. Questions were answered via a Likert scale, and each client's total score reflects distress related to clinical symptoms, interpersonal difficulties, and life satisfaction. Client responses across categories were summed to a total score ranging from 0 to 180. Clients with high (> 105), moderately high (83-104), moderate (64-82), and low/subclinical (< 64) scores were all included in the analyses. Additionally, clients' personality factors and attitudes were assessed using the Bi-Weekly Longitudinal (BIL) self-report at intake and then at biweekly intervals. The BIL is a 37-item Likert scale measure. It consists of various items from publicly available scales, including the Ten Item Personality Inventory (TIPI; Gosling et al., 2003) and the Therapeutic Working Alliance scale (Working Alliance; Duncan et al., 2003).

Procedures

Prior to 2020, the OQ 45.2 was administered at every psychotherapy session. However, in 2020, the OQ 45.2 was administered at intake and then on a bi-weekly (i.e., every other session) basis while the BIL was administered on the opposite weeks. These assessment forms were administered via paper and pencil or via a computer-based iPad system, and clients completed them right before each session. Notably, there are some potential inaccuracies in session numbers and/or change rates due to inconsistency in clinicians giving measures to clients. First, due to the alternating nature of the forms, some clients may not have completed an OQ on their last day in therapy, so their last OQ score is treated as their final OQ. Additionally, prior to 2022, sessions were recorded only when an OQ was complete, so if a clinician neglected to give a measure (at any appointment besides intake, which was always given), the session number may be incorrect. After 2022, sessions were recorded regardless of questionnaire completion.

Data Analysis

Theoretical Model Fit with Session then Week Predictors

To evaluate the primary research questions, two-level longitudinal multilevel models (MLMs), or mixed effects/hierarchical linear models, were used to assess model fit. MLM analyses were conducted in R using the "lme" function in the "nlme" or "lme4" packages. By default, these functions use restricted maximum likelihood (REML) estimation methods. In the MLMs, repeated observations (i.e., OQ scores; level 1) were nested within each individual clients (i.e., participant; level 2). Consistent with Reese et al. (2016), three models were tested: the dose-effect (DE) model, the traditional good-enough level (GEL) model, and the modified GEL model. Consistent with all previous studies, these models were first fitted to the data using the traditional session variable as the predictor. Then, these models were fitted to the data using the novel weeks variable as the predictor.

First, two-level MLMs were created and fit using linear, quadratic, log-linear, and cubic trends across all three models (i.e., DE, GEL, and modified GEL). Then, the model of best fit for each theory (i.e., the log-linear model for the DE theory or the cubic model for the GEL theory) were compared to each other to determine which theoretical model best fits the data (Baldwin et al., 2009; Owen et al., 2016; Robinson et al., 2020; Schuler et al., 2020). For the DE model, an unconditional growth model was used, which estimated an average rate of change or fixed effect of each additive session/week, as the model assumes clients progress at equivalent rates with no significant difference.

For example, the quadratic aggregate (DE) model can be defined as follows:

$$Y_{ij} = B_{00} + B_{100}(x_{ij}) + B_{200}(x^2_{ij}) + u_{0i} + u_{1i}(x_{ij}) + e_{ij}$$

where Y_{ij} is the outcome variable (i.e., OQ Score) at the i^{th} observation for person j; B_{00} is the fixed, initial client mean score at baseline (session 0); B_{100} is the fixed, average linear rate of change; B_{200} is the fixed, average quadratic rate of change; x_{ij} is the linear predictor variable (i.e., session or weeks since initial appointment) for the i^{th} observation for j^{th} person; x^2_{ij} is the quadratic predictor variable for the i^{th} observation for j^{th} person; u_{0i} is the random intercept for the i^{th} observation for j^{th} person; u_{1i} is the random slope for the linear term (x_{ij}) for the i^{th} observation for j^{th} person; and e_{ij} is the residual error for the i^{th} observation for j^{th} person.

Alternatively, in the GEL model, individual client changes (i.e., slopes) are no longer fixed, as the model proposes that there are significant and meaningful differences in change rates across clients. Further, the model posits that these rates impact progress and length of treatment (i.e., total sessions or weeks). As such, individual client change rates and interaction effects were included.

The cubic, good-enough level model can be defined as follows:

$$Y_{ij} = B_{00} + B_{100}(\max x_{ij}) + B_{200}(x_{ij}) + B_{300}(x^2_{ij}) + B_{400}(x^3_{ij}) + B_{500}(\max x_{ij} * x_{ij}) + B_{600}(\max x_{ij} * x^2_{ij}) + B_{700}(\max x_{ij} * x^3_{ij}) + u_{0i} + u_{1i}(x_{ij}) + e_{ij}$$

where the previously defined terms remain and the additional terms are as follows: B_{100}, B_{200}, B_{300}, B_{400}, B_{500}, B_{600}, and B_{700} are the fixed effects coefficients; $\max x_{ij}$ is the maximum value of the predictor variable (i.e., total sessions or total weeks in treatment) for the i^{th} observation for j^{th} person; x^2_{ij} is the quadratic predictor variable for the i^{th} observation for j^{th} person; x^3_{ij} is the cubic predictor variable for the i^{th} observation for j^{th} person; $\max x_{ij} * x_{ij}$ is the interaction term of maximum value of the predictor variable for the i^{th} observation for j^{th} person and the linear predictor variable (i.e., session or weeks since initial appointment) for the i^{th} observation for j^{th} person; maxx_{ij} * x^2_{ij} is the interaction term for the max predictor value and quadratic predictor variable for the i^{th} observation for j^{th} person; maxx_{ij} * x^3_{ij} is the interaction term of the maximum predictor value and cubic term for the i^{th} observation for j^{th} person; u_{0i} is the random intercept for the i^{th} observation for j^{th} person; u_{1i} is the random slope for the linear term (x_{ij}) for the i^{th} observation for j^{th} person; and e_{ij} is the residual error for the i^{th} observation for j^{th} person.

Further, in the modified GEL model, the traditional GEL model was extended by including session frequency, or the average weeks between sessions (calculated as total weeks divided by total sessions) and an interaction term between total sessions and change rate to investigate whether improvement is moderated by frequency or regularity.

The linear, modified good-enough level model can be defined as follows:

$$Y_{ij} = B_{00} + B_{100}(\mathrm{max}x_{ij}) + B_{200}(freqx_{ij}) + B_{300}(x_{ij}) + B_{400}(\mathrm{max}x_{ij} * freqx_{ij}) + B_{500}(x_{ij} * \mathrm{max}x_{ij}) + B_{600}(x_{ij} * freqx_{ij}) + B_{700}(x_{ij} * freqx_{ij} * \mathrm{max}x_{ij}) + u_{0i} + u_{1i}(x_{ij}) + e_{ij}$$

where $freqx_{ij}$ is the frequency of total sessions over time (i.e., weeks) for the i^{th} observation for j^{th} person; and interaction terms involving maxx_{ij}, $freqx_{ij}$, and x_{ij} are included for the i^{th} observation for j^{th} person.

To compare models, -2 log-likelihood estimation (*-2LL*), Akaike Information Criterion (*AIC*), Bayesian Information Criterion (*BIC*), and chi-squared model fit statistics were used (McCoach & Black, 2008) where lower estimates for both *AIC* and *BIC* (with a difference of at least 10), higher estimates for *-2LL*, and significant additive effects for the more complex model or a lack thereof for the simpler model in the Chi-Squared analysis all suggest better model fit (Owens et al., 2016; Schuler et al., 2020).

Assessing the Predictive Accuracy of Session and Weeks Variables

Following the comparison of model fit across the DE, traditional GEL, and modified GEL models for session and week variables, the superior models for sessions and weeks were compared. Then, the predictive accuracy of both weeks and sessions variables as analyzed by partitioning the data into training (85%) and testing (15%) sets. The training set was used to re-create the models, and the testing data was used to evaluate model performance. Then, the fit of both session and week models using the partitioned testing data were compared. Like previous analyses, *-2LL*, *AIC*, and *BIC* were used for comparison. Further, conditional R^2 values were compared, and pseudo-R^2 effect size statistics were calculated for each longitudinal model by taking the squared value of a correlation between the predicted and observed OQ-45 scores for each model (see Peugh, 2010 & Reece et al., 2016). This pseudo-R^2 was comparable to traditional R^2 statistics, as it served as a measure of goodness of fit and represented the proportion of variance in the dependent variable explained by the independent variable.

Additionally, Root Mean Squared Error (RMSE) values were used, as RMSE measures the average magnitude of errors between predicted and observed values and provides a measure of the spread of residuals. Therefore, the RMSE communicated how well the model generalizes to new, unseen data. Lower RMSE values suggested better model performance and smaller prediction errors. Additionally, bias-variance tests were used to further assess overfitting and generalizability by comparing mean squared errors (MSE) for training and testing sets, as overfitting may be indicated if the model performs very well on the training set and poorly on the testing set.

Evaluating Interacting Effects and Additional Predictor Variables

Finally, significant interactions were investigated via an exploratory analysis using the best-fitting model for both sessions and weeks. Possible interactions consisted of data collected during the preliminary phone screening (i.e., demographic information, treatment history, presenting problems, etc.). Due to excessive missing data (72%), potential interaction effects of BIL variables (i.e., working alliance, openness, hope, motivation, etc.) could not be included in the models. As such, to investigate whether factors related to clients' personality characteristics or their motivation for treatment was associated with treatment outcomes, clients who achieved reliably significant improvement [RSI; (i.e., the decline in the OQ score by 14 [or approximately 1 SD)] were compared to all other clients and then to clients with reliably significant deterioration [RSD; (i.e., an increase in OQ score by 14)]. Chi-square analyses and t-tests were used to investigate and identify factors that are associated with treatment success and deterioration.

RESULTS

Descriptive Statistics & Visuals

Preliminary Data Analysis of the Full Sample (n=434)

In the larger sample of 434 clients, the average total number of sessions attended was 22 [Median (*Mdn*) = 11, *SD* = 27, *Range* = 1-163], and the average total time in treatment was 57 weeks (*Mdn* = 30.71, *SD* = 75.67, *Range* = 1-482). Sessions occurred, on average, once every three weeks, but there was significant variation in frequency across clients (*M* = 3.01, *Mdn* = 2, *SD* = 6.48, *Range* = 0.2-105 weeks). The average initial OQ score for all participants was 72 (OQ range = 0-180), which suggests moderate distress *(Mdn* = 73.45, *SD* = 22.51, *Range* = 20-139), and the average final OQ score was 62, which suggests subclinical distress (*M* = 62.46, *Mdn* = 63, *SD* = 23.72, *Range* = 14-163). The overall difference in the initial and final OQ scores across clients was significant, suggesting that, overall, treatment had a small to medium effect *(t* = 10.742, *df* = 433, *d* = -0.407, *p* = < .05). Further, throughout treatment, participants' OQ scores decreased by an average of 10 points *(Mdn* = 6, *SD* = 18.76, *Range* = -85 to +38), which suggests that more than half of the participants did not achieve the 14-point difference on the OQ reflective of reliably significant change. In this sample, the client's change in OQ score was <u>not</u> highly correlated with many predictor variables, including total sessions (*r* = .14), total weeks (*r* = .10), and session frequency (*r* = .03). However, <u>both total sessions and total weeks in treatment had a large and significant correlation (*r* = .884),</u> which supports the notion that the frequency of sessions was similar across participants.

Preliminary Data Analysis of the Sub-Sample Used in the Models (n=311)

After excluding participants with fewer than 3 OQ scores and with more than 1 year between consecutive sessions, clients included in the second, smaller sub-sample (*n* = 311) had an average initial OQ score of 70, which suggests moderate distress *(Mdn* = 72, *SD* = 23, *Range* = 8-139) and an average final OQ score of 59 which suggests subclinical distress *(Mdn* = 60, *SD* = 25, *Range* = 25-163). Changes in OQ scores were often relatively consistent across all subscales of symptom distress (SD), interpersonal relations (IR), and social role (SR), as all subscales were moderately and positively correlated with each other (IR & SR: $r = 0.48$; *SD* & IR: $r = 0.58$; *SD* & SR: $r = 0.62$). In this sub-sample, participants experienced an average decline of 14 points on their OQ scores upon comparison of initial and final OQ scores, <u>so half of clients (50.4%) experienced RSI at termination. However, notably, almost two-thirds (65%) of clients achieved RSI at some point in treatment</u>. The pattern of accumulation of clients achieving RSI by sessions and weeks follows a negatively accelerating pattern, with significant growth initially and diminishing marginal returns thereafter (see Figures 21-26).

Descriptive Statistics for Predictor Variables

Sessions. Although both the DE and GEL models view sessions as a powerful predictor of improvement, the DE model proposes that the total number of sessions is most important. In contrast, the GEL model posits that the rate of change predicts improvement since faster responders typically have fewer sessions. In the sub-sample used for MLMs (*n* = 311), clients attended an average of 26 sessions *(Mdn* = 16, *SD* = 27, *Range* = 3-163) at an average rate of one session every 2 weeks *(Mdn* = 2, *SD* = 1.5, *Range* = 0.69-12.3 weeks). On average, participants' total OQ scores declined by 1.24

points per session [(*Mdn* = -0.62 *SD* = 2.15, *Range* = -11.8-4.16). Further, there was a significant relationship between individuals' rates of change and their total number of sessions [$F(1, 309) = 34$, $p < .001$], as clients with faster declines in symptoms had significantly fewer sessions than those with more moderate or slow change rates ($F(310, 5670) = 1.21e^{27}$, $p < .001$).

Weeks. Further, in the sub-sample used for MLMs, participants spent an average of 60 weeks in treatment (*Mdn* = 40, *SD* = 59, *Range* = 3-163). Sessions occurred, on average, once every two and a half weeks but ranged from more than once per week to once every 12 weeks (*M* = 2.57, *Mdn*=2.08, *SD*=1.52). On average, participants' total OQ scores declined by 0.62 points per week (*Mdn* = -0.27, *SD* = 1.27, *Range* = -8.84 to 2.25). Similar to the session variables, there was a significant relationship between individuals' change rates and their total weeks in treatment, as those with faster declines in symptoms engaged in therapy for significantly fewer weeks ($F(310, 5670) = 68.64$, $p < .0001$).

Visual Inspection

Initial visual inspection of plotted data for each client included in the sub-sample ($n = 311$) suggested a non-linear pattern of change across clients (see Figures 1 and 2). When aggregate client OQ scores were plotted by session, the pattern appeared log-linear or cubic, as participant functioning improved quickly during the first 20 sessions, stabilized until session 100, then sharply improved with significant variability (see Figures 3 and 5). When aggregate client OQ scores were plotted by weeks, the data suggested a cubic pattern with the same significant increase until 30 weeks, then a slight dip in functioning around 65 weeks with a return back to better functioning around week

100, and then, a steady decline in functioning until week 150 that improved sharply until the final week (see Figures 4 and 5).

To eliminate the impact of dropout due to early responders on the trend line, participants attending below and above the median total number of sessions and total weeks were plotted separately. For those with total sessions above the median, the trend line was consistent with the summative data by OQ described above and with the DE model (see Figure 8). For those with less than the median total number of sessions, the pattern appears quartic, with sharp gains or symptom improvement up to session 10, a slight decline in functioning to session 15, and a slight improvement at session 25 (similar to the OQ score in session 10), then a decline in functioning from sessions 30 to 40, likely due to client drop out (see Figure 8). For participants with total weeks in treatment above or below the mean total weeks, cubic relationships were present, which is consistent with the summative data and the DE model (see Figure 7). However, participants who were in therapy for less than the median total weeks had a sharper initial improvement in functioning (until session 10), a smaller decline in functioning (sessions 10-15), then a sharp improvement in functioning until session 40. Participants who attended therapy for more than the median number of weeks had a more gradual initial improvement after the initial gains, with significant gains in OQ scores around week 50, gradual improvement to week 100, a decline in improvement until approximately week 160, then a sharp improvement in symptoms until over 300 weeks (see Figure 7).

Participants were separated into three semi-equal groups (n = 103-104 per group) based on their change rates per session or week to assess differences in change rates across predictors. For participants who experienced moderate change rates per session

(M_{Mod} = -0.61, *Range* = -1.22 to -0.22), a primarily linear decline was present (see Figure 9). Relative to the visual for individuals with moderate change rates, individuals with faster change rates (M_{Fast} = -2.82, *Range* = -11.8 to -1.25) appeared to have more quadratic concave down relationships, as they began at similar points but experienced a sharper decline in symptoms (to approximately an OQ of 50 at session 10) followed by a slow, consistent, and linear worsening of symptoms to approximately 63 OQ points which is likely attributed to early drop out (see Figure 9). Individuals with slow rates of change (M_{Slow} = 0.4, *Range* = -0.21 to 4.16) appeared to have a cubic pattern as a trend line with a small decline in symptoms until session 40, increased symptoms until session 75 (which was slightly higher than initial symptoms), then a sharp decline in symptoms until the final session (see Figure 9). There was a significant relationship between individuals' rates of change and their total number of sessions [$F(1, 309) = 34$, $p < .001$], as faster rates of change were significantly associated with fewer total sessions ($M_{\mathbf{Fast}}$ = 10, *SD*= 7, *Range*= 3-36; $M_{\mathbf{Mod}}$= 31, *SD* = 21, *Range*= 4-102; $M_{\mathbf{Slow}}$ = 38, *SD* = 35, *Range* = 3-163) with significant differences between groups [$F(2, 308) = 38$, $p < .001$].

Further, the differences in change rates across the three groups by weeks (n = 103-104 per group) were also significant, as clients with faster declines in symptoms spent significantly fewer weeks in treatment than those with more moderate or slow change rates ($F(310, 5670) = 68.64$, $p < .001$). For participants that experienced moderate change rates per week (M_{Mod} = -0.27, *Range* = -0.54 to -0.1), a significant improvement in symptoms occurred until week 100 with a small worsening of symptoms around session 50 likely due to outliers; then, at week 100, the symptoms continue to worsen slightly and steadily until around week 200, then the symptoms significantly improve beyond this point to final weeks (see Figure 9). Individuals with faster change rates (M_{Fast} = -1.28, *Range* = -8.84 to -0.55) appeared to have a significant and rapid decline in symptoms until approximately week 40, then there was a slight and steady worsening of symptoms until approximately week 80 followed by another improvement resulting in a similar OQ score to the score at week 40 (see Figure 9). Individuals with slow rates of change (M_{Slow} = 0.4, *Range* = -0.19 to 2.25) appeared to have more steady symptoms with minimal and gradual changes, but participants appeaedr to experience significant improvement around approximately week 200.

Data Analysis

Assessing Relative Fit of MLM by Sessions

Dose-Effect Model. To assess theoretical model fit, the first analysis consisted of creating and fitting linear, quadratic, cubic, and log-linear models across all aggregate (dose effect), traditional good-enough level, and modified good-enough level models, as consistent with previous studies (Reese et al., 2014). Across the aggregate model, a comparison of the linear (*AIC* = 45999, *BIC* = 46039, *-2LL* =-22993), quadratic (*AIC* = 46012, *BIC* = 46059, *-2LL* = -22991) cubic (*AIC* = 45987, *BIC* = 46041, *-2LL* =-22966), and log-linear models (*AIC* = 807, *BIC* = 847, *-2LL* =-389.5) revealed that the log-linear model had the significantly best fit to the data (see Table 4). A comparison of the BIC values and calculation of the Bayes Factor (BF_{ij} = ~0) suggests strong support for and an overwhelming likelihood the log-linear model has the best fit according to the Jeffreys' Scale for Bayes Factor Interpretation (1961). A chi-square test comparing the log-linear and quadratic models further supported the log-linear models' superior fit, as additive components of the quadratic model failed to be significant [ΔX^2 (df =1) = 0].

Additionally, the fixed and random effects considered in the log-linear, dose-effect model explained 97.6% of the variance, which was greater than the linear (93.9%), quadratic (93.6%), and cubic (92.9%) models.

The dose-effect model of best fit (log-linear) suggests that the average symptom count at or before baseline is a moderate OQ score (B_{00} = 67.96, $p < .001$), and the average decline in symptoms for each additional session is nearly 1 OQ point (Session = 0.99) with inconsistent additive benefits of each session (see Table 4). Further comparisons of the linear to quadratic and quadratic to cubic models revealed the second-best relative model fit was the cubic model [$\Delta X^2(df=1) = 50$, $p < .001$]. When comparing all DE models, all effects and pieces had a statistically significant impact, which consisted of linear (Session = -0.51, $p < .001$), log-linear ($\text{Session}_{\log}$ = -0.015, $p < .001$), quadratic (Session^2 = -0.53, $p < .001$), and cubic (Session^3 = -0.67, $p < .001$) effects, as higher-order terms provided additional explanatory power beyond the linear term alone. This suggests a general decline in OQ score as the rate of sessions increases, but the decline is likely not linear or equally consistent in rate over time.

Further, for all linear, quadratic, and cubic models, the model correlations between the initial OQ score and session slope are small and negative. However, their covariances are larger, suggesting a strong, negative relationship [$cov(\text{B}_{00j}, u_{1i})$: linear = -0.98, quadratic = -0.96, cubic = -0.75]. As such, more symptomatic clients are likely to have slower changes than less symptomatic clients. However, the log-linear model suggests a small, positive correlation between participants' intercepts and a low covariance, suggesting slight joint variation between intercepts and slope groups [cov

(B_{00j}, u_{1i}) = -0.98]. Additionally, there was a small, negative correlation between the random effects from the initial OQ score and session slopes (r = -.08).

Traditional Good-Enough Level Model. Further, across the traditional GEL model, a comparison of the linear (*AIC* = 45999, *BIC* = 46039, *-2LL* =-22993), quadratic (*AIC* = 46012, *BIC* = 46059, *-2LL* = -22991), cubic (*AIC* = 45987, *BIC* = 46041, *-2LL* =-22966), and log-linear models (*AIC* = 807, *BIC* = 847, *-2LL* =-389.5) revealed that the log-linear GEL MLM model also had the best relative fit to the data with an extremely strong effect, suggesting almost definitive support for the superior fit of the log-linear model (Kass & Raftery, 1995; see Table 6). A chi-square test comparing the log-linear and quadratic models further supported the log-linear models' superior fit [ΔX^2 $(df = 1)$ = 0, p = <0.05/ not reported]. Further, the log-linear model has a high level of explanatory power as 97.8% of the variance in OQ scores can be explained by the model predictors, which is significantly more than the linear (94.2%), quadratic (93%) and cubic (92.7%) models.

The traditional, log-linear Good Enough Level model suggests that the average symptom count at or before baseline is a very small yet moderate OQ score (B_{00} = 65.2, $p < .001$), the average decline in symptoms for each additional session is nearly 1 OQ point (Session= 0.99, $p < .001$), and the decline in OQ score based on the interaction between total sessions and the session is insignificant yet positive, suggesting a possible deterioration for some patients or inconsistent rates of change [(max(x_{ij}) & max(x_{ij}) * Session = 1.00, $p > .05$]. Further, the fixed and random effects included in the log-linear dose-effect model explains 97.8% of the variance in OQ score based on session (see Table 6). Additional comparisons of the linear to quadratic and quadratic to cubic models revealed that the second best model fit was the quadratic model, as it was significantly better than the linear model [ΔX^2 $(df=2) = 0$] but not the log-linear model [ΔX^2 $(df=2) = 0$, $p > .05$], and the cubic model lacked significant additive effects [ΔX^2 $(df=2) = 2.77$, $p > .05$]. Results from the traditional Good Enough Level models showed many effects to be statistically significant, including the linear (Session = -.94, $p < .001$), quadratic ($Session^2 = 0.008$, $p < .001$), cubic ($Session^3 = -0.01$, $p < .001$), and intercept (B_{00} = 65.16, $p < .001$) fixed effects. However, there was no additive effect of total sessions (p > .05) or the interaction of total sessions and session (i.e., slope) for any model besides the quadratic (0.069, $p < .001$).

Further, within the log-linear model, correlations between session number and initial OQ score ($r = -.01$) and session number with total sessions during treatment ($r = -.05$) were small/negligible and negative. However, there was a large, negative correlation between baseline OQ score and total sessions ($r = -.69$), which suggests that <u>those with higher symptom scores at baseline tended to attend fewer sessions and those who attended high numbers of sessions tended to be less symptomatic at baseline.</u>

Modified GEL Model. Additionally, across the modified GEL model modified by sessions, a comparison of the linear (*AIC* = 45953, *BIC* = 46033, *-2LL* =-22993), quadratic (*AIC* = 45957, *BIC* = 46057, *-2LL* = -22963), cubic (*AIC* = 45931, *BIC* = 46092, *-2LL* =-22942), and log-linear models (*AIC* = 746, *BIC* = 827, *-2LL* =-361) suggested that the log-linear model has the best relative fit, and this was further supported by Jeffreys' Scale for Bayes factor interpretation (which suggests strong and decisive support for the log-linear model fit) and with a chi-square test [ΔX^2 $(df = 1) = 0$, p = not reported] (Kass & Raftery, 1995; see Table 4). The predictors in the modified, log-linear

GEL model were estimated to explain approximately 97.8% of the variance in OQ scores, which is greater than the linear (94.2%), quadratic (93%), and cubic (92.7%) models.

The modified, log-linear GEL model had many significant effects. Specifically, the log-linear GEL model suggests that the average symptom count at or before baseline is a moderate OQ score (B_{00} = 68.72, $p < .001$), the average decline in OQ score by total sessions is -1.00 (Session= -1.00, $p < .05$), and the interaction effects of frequency with session (Session*freqx$_{ij}$, = -0.99, $p > .05$) frequency with total sessions (*freqx$_{ij}$,*max$_{ij}$= 1.00, $p < .05$) and all three terms (Session*freqx$_{ij}$,*max$_{ij}$= 1.00, $p < .01$) are significant. There is an insignificant impact of total sessions, frequency, and the interaction of total sessions and session number within the model. Further comparisons of the log-linear to quadratic models were insignificant, and analyses between the linear to quadratic and quadratic to cubic models revealed the second-best model fit was the cubic model [ΔX^2 $(df = 9) = 44$, $p < .001$]. Compared to the DE and traditional GEL models, the effects of predictors in the modified GEL model were less significant, which may be attributed to multicollinearity, model complexity, or larger residuals indicating possible non-normality. Specifically, in the log-linear, modified GEL model, some scores were still significant, including mean OQ-45 at baseline (B$_{00}$ = 68.7, $p < .001$), quadratic growth rate (Session^2 = 0.016, $p < .05$), cubic growth rate (Session^3 = 0.00002, $p < .001$), However, other predictors' significance were similar to the previously described log-linear model. As mentioned in previous models, the significant effects across growth rates suggest a general decline in OQ score as rate of sessions increases, but the decline is likely not linear or equally consistent in rate over time.

Lastly, a correlation matrix revealed large, significant, and positive relationships between initial OQ with both total sessions (*r* = -.55) and session frequency (*r* = -.83). This suggests that <u>clients with higher initial OQ scores discontinued treatment earlier than other clients and attended treatment more frequently</u> [since session frequency is calculated as weeks/sessions, the inverse relationship suggests clients with high OQ scores have a lower frequency value, meaning they attend sessions with less time between sessions (i.e., weekly over biweekly)]. These relationships also suggest the inverse is true, so participants with high initial symptom scores attend more sessions less frequently over time. Additionally, there was a positive relationship between total sessions and session frequency (*r* = .60), <u>so participants who attend sessions more frequently (i.e., weekly) have fewer total sessions in treatment</u>. In contrast, participants who attend less frequently (i.e., bi-weekly, monthly, etc.) have a significantly higher number of total sessions throughout treatment.

Overall Model Fit for Session Variables. After comparing all the log-linear DE, traditional GEL, and modified GEL models using multiple statistical methods (including chi-square), the log-linear, traditional GEL model was determined to have the best relative fit to the data with an extremely strong effect {[(ΔX^2 $(df = 6) = 96$, $p < 0.001$], ($AIC_{\text{DE}} = 807$, $BIC_{\text{DE}} = 947$); ($AIC_{\text{GEL(Trad)}} = 788$, $BIC_{\text{GEL(Trad)}} = 820$); ($AIC_{\text{GEL(Strat)}} = 747$ $BIC_{\text{GEL(Strat)}} = 827$). Notably, the predictors in the traditional GEL model explain approximately 97.4% of the variance between OQ scores, which is similar to the linear (R^2 = 97.6%) and modified GEL models (R^2 = 97.4%).

Assessing Relative Fit of MLM by Weeks

To assess theoretical model fit using weeks as a predictor, the first analysis consisted of creating and fitting linear, quadratic, cubic, and log-linear models across all three theoretical models consistent with previous studies. Like the session models, the log-linear models had a significantly better fit than the linear, quadratic, and cubic fits across all three models (DE, traditional GEL, and modified GEL models; see Tables 12, 14, and 16). According to Jeffreys' Scale for Bayes Factor Interpretation (1961), there is extreme and decisive evidence for the superior fit of the log-linear models (Kass & Raftery, 1995). Also, across all models, further comparisons of the log-linear to quadratic, linear to quadratic, and quadratic to cubic models revealed non-significant effects in all except the cubic model which was the second best model fit across all theories {[DE: ΔX^2 $(df = 1) = 17$, $p < 0.001$], [Traditional GEL: ΔX^2 $(df = 2) = 42$, $p < .001$], [Modified GEL: ΔX^2 $(df = 10) = 15.7$, $p < .001$]}.

Dose-Effect Model. The dose-effect model of best fit (log-linear) suggests that the average symptom count at or before baseline is a <u>sub-clinical</u> OQ score (B_{00} = 62.15, $p < .001$), and the average decline in symptoms for each additional session is nearly 1 OQ point (Session= 0.99, $p < .001$) with inconsistent additive benefits of each session. Additionally, the DE's log-linear model has a high level of explanatory power, as 95.5% of the variance in OQ scores can be explained by the model predictors, which is significantly more than the linear (92.2%), quadratic (92%), and cubic (91.3%) models (see Table 12). Like the DE model for the session predictors, there is a small, insignificant effect between the initial OQ score and participants' slope (r = -.08).

Traditional Good-Enough Level Model. The log-linear, good-enough level model of best fit suggests that the average symptom count at or before baseline is also sub-clinical OQ score (B_{00} = 61.98, $p < .001$); the average decline in symptoms for each additional session is nearly 1 OQ point (Session= 0.99, $p < .001$) with inconsistent additive benefits of each session. The total number of sessions has an insignificant, positive effect (1.00, $p > .05$). Further, the fixed and random effects of predictor variables included in the log-linear and traditional good-enough level model explain 95.5% of the variance in OQ scores, which is greater than the linear (92.2%), quadratic (95.5%), and cubic (91.3%) models (see Table 14). Similar to the traditional GEL model for session predictors, there are insignificant relationships between weeks with both initial OQ score (r = .02) and total weeks in treatment (r = -.04); however, there is a significant inverse relationship between total weeks in treatment and initial OQ score (r = -.71) which suggests that more symptomatic clients spend less total time in treatment than less symptomatic ones.

Modified Good-Enough Level Model. Further, the modified GEL model of best fit (log-linear) suggests that the average symptom count at or before baseline is a clinical OQ score (B_{00} = 69.33, $p < .001$), and the average decline in symptoms for each additional session is nearly 1 OQ point (Session= 0.99, $p < .001$). Notably, the Weeks variable was significant at all levels ($Weeks^2$ = 0.00079 & $Weeks^3$ = -0.0000017, $p < .001$), which suggests a general decline in OQ score as weeks in treatment progress. However, the decline may not be linear or equally consistent in rate over time. Also, the log-linear, modified GEL model had high explanatory power, as 95.5% of the variation in

OQ scores can be explained by the model predictors, which is significantly more than the linear (92%), quadratic (92%), and cubic (91%) models (see Table 12).

Within the log-linear, modified GEL model, there were large correlations between variables, including a large, inverse correlation between initial OQ scores with both frequency (r = -.87) and total weeks (r = -.74), and a large, positive correlation between session frequency and total weeks (r = .67). As such, those with higher initial OQ scores are likely to spend fewer total weeks in therapy, and they likely have a lower average session frequency, meaning they attend sessions more frequently than others with lower initial scores (e.g., since the variable is coded in terms of weeks, a frequency of 1 means every approximately weekly session while a value of 2 reflects twice-monthly sessions). Further, individuals who attend psychotherapy more frequently are more likely to spend less time in therapy than those attending on a less frequent basis.

Overall Model Fit for Week Variables. After comparing all three models, the traditional GEL model was determined to be the model of best fit for week-based variables based on many statistical comparisons, despite an insignificant chi-square calculation ($AIC_{\mathrm{DE}} = 649$, $BIC_{\mathrm{DE}} = 689$, $R^2 = 0.955$); ($AIC_{\mathrm{GEL(Trad)}} = 631$, $BIC_{\mathrm{GEL(Trad)}} = 685$, $R^2 = 0.955$); ($AIC_{\mathrm{GEL(Strat)}} = 632$, $BIC_{\mathrm{GEL(Strat)}} = 712$, $R^2 = 0.96$)}. This model explains approximately 95.5% of the variance between scores. Notably, all models had similar fits and scores assessing fit.

Comparing the Predictive Accuracy of Weeks and Sessions

To first compare model fit for sessions and weeks as predictors, an analysis of variance (ANOVA) and likelihood ratio test were conducted for the log-linear, traditional GEL models for both variables. Although the ANOVA suggested insignificant differences between models [ΔX^2 $(df=0) = 135.4$, $p < .001$], both the results of likelihood ratio test [ΔX^2 $(df=0) = 128$, $p < .001$] and many other values assessing model fit ($AIC_{\text{Session}} = 767$; $BIC_{\text{Session}} = 820$; $-2LL_{\text{Session}} = -375$; $AIC_{\text{Week}} = 631$; $BIC_{\text{Week}} = 685$; $-2LL_{\text{Session}} = -308$) suggest that the weeks model has a better fit to the data.

However, to compare the predicted power of weeks and session data using the best fitting log-linear model for both sessions and weeks, partitions consisting of training (85%; $n = 264$) and testing (15%; $n = 47$) sets were created using the second dataset ($n = 311$). The resulting observed values (from the training set) and predicted values (from the testing set) based on sessions and weeks data were compared, and the model fit of predicted values on the model created using the training values were compared across variables.

A comparison of the models indicated very similar and mixed data across models [($AIC_{\text{Session}} = 730$; $BIC_{\text{Session}} = 746$; $-2LL_{\text{Session}} = -353$; $AIC_{\text{Week}} = 723$; $BIC_{\text{Week}} = 761$; $-2LL_{\text{Session}} = -357$). Further, the *conditional-*R^2 values and *pseudo* R^2 effect size values were similar, suggesting the models have similar goodness of fit and proportion of variance explained (*conditional* $R^2_{Session} = 0.797$, *pseudo-*$R^2_{Session} = 0.695$); (*conditional* $R^2_{Weeks} = 0.793$, *pseudo-*$R^2_{Weeks} = 0.686$)]. Notably, these values were significantly lower than their original equivalents created with the complete datasets (consisting of *conditional* $R^2_{Session} = 0.973$; *conditional* $R^2_{Weeks} = 0.955$), which may suggest that the original models were overfitted. To investigate this, bias-variance tests were used to determine a balance between the bias and variance in both session and week models, as the mean squared errors (MSE) between both datasets are comparable without a significant effect of standardization (*Training* $MSE_{\text{sessions}} = 4139$; *Testing* $MSE_{\text{sessions}} =$

4122.21; $Training\ MSE_{weeks} = 4139$; $Testing\ MSE_{weeks} = 4095$). These MSE scores further suggest that both the weeks and session models fit to the training data with a similar level of accuracy. However, the session model has a slightly higher yet insignificant difference in MSE. All of these MSE values do not suggest significant overfitting. Given the data (i.e., *AIC*, *BIC*, etc.) and the results of a likelihood ratio test, [$\Delta X^2\ (df = 0) = 6.7$, $p < .001$], the weeks' model appears to be a slightly better fit.

Model Interactions and Predictors of Treatment Success and Deterioration

Interacting Effects

Possible interacting effects were investigated using the traditional GEL model (i.e., the model of best fit) with weeks variables as predictors. Many variables endorsed by clients before intake were examined to assess their possible impact on OQ score, their interaction with the predictors, and their potential additive effect on improving model fit. These variables were predominantly coded dichotomously (e.g., Yes or No) and included information about treatment history and current symptoms. Many variables had an insignificant fixed effect on model fit ($p > .5$), such as current or history of psychotherapy or trauma, income, current use of psychiatric medication, endorsement or denial of anxiety, obsessive thoughts, aggressive behaviors, academic difficulty, hyperactivity, memory problems, hearing voices, and sleep, eating, and social problems (see Table 23). Additionally, therapist and supervisor grouping effects were insignificant and suggested little variation between providers once other fixed and random effects in the model were considered ($SD_{\text{therapist \& supervisor}} = \sim 1$, $var_{\text{therapist \& supervisor}} = \sim 1$).

However, many presenting problems, as endorsed by clients, had a significant fixed effect when included in weeks and session models. The endorsement of depressive symptoms ($t = 3.6$, $p < .001$), hopelessness ($t = 3.5$, $p < .001$), suicidal thoughts ($t = 3.6$, $p < .001$), history of a suicide attempt ($t = 2.401$, $p < .05$), and attention problems ($t = 2.25$, $p = .026$) were all significant in at least one model (i.e., weeks or session models). However, depressive symptoms, hopelessness, and suicidal thoughts had a significant impact on both models. Notably, when all three interactions were included in the model together, the significance of the suicidal thoughts as a predictor declined ($p = .017$), likely due to multicollinearity. However, the model with all three factors (e.g., depression, hopelessness, and suicidal thoughts) fit better than only a model without suicidal thoughts and better than a model without interaction terms for those variables [ΔX^2 (df =1) = 5.56, $p < .01$].

The estimates for each predictor of the traditional weeks model inclusive of the three significant interactions (i.e., depression, hopelessness, suicidal thoughts) when holding all other predictors constant are as follows: the estimated mean of a client's OQ score at baseline (when all other predictors are 0) is 48.2 ($t = 62.7$, $p < .001$); the estimated change for a one-unit increase in total weeks of treatment is an increase in OQ score of 1 per additional week in treatment which is an insignificant effect ($t = 0.97$, $p = .34$); the estimated mean change for a one-unit increase in the current week in treatment is -0.99 ($t = -5.988$, $p < .001$); and the estimated mean change in total weeks given the current week in treatment number is 1.0 ($t = 3.55$, $p < .001$). Additionally, when considering the new, binary predictors (e.g., coded for absence or presence) for individuals who endorse depressive symptoms, their mean OQ score is likely to be 1.165 points higher ($t = 3.56$, $p < .001$); for those endorsing suicidal thoughts, their OQ score is likely to be 1.18 points higher (t = 2.371, p = .012); and for those endorsing hopelessness, their OQ score is likely to be 1.15 times higher (t = 3.56, p < .001).

Although inherently related, the interaction terms for these dichotomous symptom predictors are insignificant. However, many are correlated with each other and with client baseline measures, as both depression and initial OQ score (r = -.65) and hopelessness and initial OQ score (r = -.64) are largely and inversely correlated, meaning those who endorse feelings of hopelessness and depression are more likely to have higher initial OQ scores, and those with low OQ scores are less likely to feel hopeless or depressed. Additionally, there is a positive, significant effect between hopelessness and depression (r = .60), suggesting those experiencing one are more likely to endorse the other, and there is a moderate, inverse effect between total weeks in treatment and initial OQ score (r = -.39), suggesting that those with higher scores at baseline tend to experience earlier discharge from treatment (see Table 25).

Clients with Reliably Significant Improvement

Of all the clients included in the previous studies (n = 311), 65% of clients (n = 203) met the criteria for Reliably Significant Improvement (RSI) [(i.e., a decrease of over 14 OQ points between the first and last OQ as per Lambert (2015)] at some point throughout treatment (see Figures 21 and 24). Consistent with many previous studies, the plot of the accumulation of all clients achieving RSI across sessions and weeks both resulted in log-linear, negatively accelerating curves (see Figures 21 and 24).

Since the GEL model typically attributes this negatively accelerating curve to early responders and termination, more visuals were created for participants with total sessions or weeks above and below the median total time/sessions spent in treatment.

This allowed for visual examination with minimal to no impact of early responders and their termination for the graphs for those above the median. Importantly, all graphs displayed negatively accelerating curves for weeks and sessions above and below the median, which disputes the attribution of the negatively accelerating curve solely to early responders (see Figures 22, 23, 25, and 26). However, there were significant differences in the total time (i.e., weeks) or session number that RSI was achieved when comparing those above and below the median number of sessions/weeks, as those that terminated treatment sooner met RSI faster than those who stayed in treatment longer for both predictors [($t_{\text{Session}} = 4.54$, $M_{\text{Below Session}} = 4.5$, $M_{\text{Above Session}} = 10.6$, $df_{\text{Session}} = 114.4$, $p_{\text{Session}} = .00013$; see Figures 22 and 23); ($t = 4.14_{\text{Week}}$, $M_{\text{Below Week}} = 9.4$, $M_{\text{Above Week}} = 21.1$, $df_{\text{Week}} = 137.1$, $p_{\text{Week}} = .0005$; see Figures 25 and 26)]. This suggests that clients may terminate treatment when they achieve a good enough benefit for themselves.

Of all the participants included in the sub-sample ($n = 311$), only 50% of clients (n = 157) met the criteria for a reliably significant improvement (RSI) or alleviation of symptoms <u>at the time of discharge</u> (i.e., a decrease of 14 points when subtracting their first and final OQ scores). The discrepancy in scores relative to the percentage of participants who met RSI at any point (65%) suggests that participants' OQ scores waxed and waned throughout treatment, which may be attributed to situational factors or timing, among other factors. As such, participants who achieved RSI at discharge ($n = 157$) were compared to all other clients ($n = 171$) across many variables to determine possible determinants of or significant factors influencing change.

There were insignificant differences between those that met or failed to achieve RSI in total number of sessions, total weeks in treatment, session frequency, and highest symptom score throughout treatment. However, compared to those who failed to achieve RSI, those achieving RSI had their highest OQ scores at significantly earlier points in treatment, including at an earlier session number (t = -4.4, M_{RSI} = 5, $M_{NO.RSI\ (NO)}$ = 11, p < .001) and an earlier week in treatment (t = -4.7, M_{RSI} = 10, $M_{NO.RSI\ (NO)}$ = 25, p < .001). Additionally, participants achieving RSI had a significantly greater range in OQ scores throughout treatment (t = 7.5, M_{RSI} = 40, M_{NO} = 26, p < .001) and significantly greater change in total OQ score and across all OQ subscales from beginning to end of treatment [(Total OQ Score: t_{OQ}= -23, M_{RSI} = -29, M_{NO} = -2, p_{OQ} < .001); (IR: (M_{RSI} = -5.8, M_{NO} = -0.89, p_{IR} < .001); (SR: M_{RSI} = -4.3, M_{NO} = -0.7, p_{SR} < .001); (SD: M_{RSI} = -15.8, M_{NO} = -1.9, p_{SD} < .001)]. Also, those who achieved RSI had significant differences in their initial and final subscale and total OQ scores (p < .001), as those achieving RSI had significantly higher initial OQ scores ($t_{Initial\ OQ}$ = 4.8, M_{RSI} = 77, M_{NO} = 65, $p_{Initial\ OQ}$ < .001) and significantly lower final OQ scores ($t_{Final\ OQ}$ = -6.6, M_{RSI} = 50, M_{NO} = 68, $p_{Final\ OQ}$ < .001). Additionally, those achieving RSI had significant differences in their change rates with faster declines in symptoms per session (t = -11.5, M_{RSI} = -2.4, M_{NO} = -0.04, p < .001) and per week (t = -9, M_{RSI} = -1.18, M_{NO} = -0.003, p < .001) than their counterparts.

Further, the impact of individual patient characteristics and attitudes toward treatment were examined using the BIL data. This measure includes scales reflective of participants' attitudes towards therapy (i.e., therapeutic motivation, hope, therapeutic working alliance, etc.) and patient personality (i.e., extraversion, openness, emotional stability, etc.). BIL data was available for approximately 27% (n = 84) of clients. A comparison of clients who achieved and failed to achieve RSI revealed no significant differences between groups in their initial scores, final scores, or change rates across the following scales: openness, agreeableness, conscientiousness, motivation, and perceptions of their working alliance with their therapist ($p > .05$).

However, independent samples t-tests revealed some significant differences between groups. First, participants who achieved RSI were significantly less emotionally stable at baseline ($M_{RSI} = 6.8$, $M_{NO} = 8.1$, $p < .001$). Notably, one of two items on the emotional stability appears to also assess the construct of anxiety (i.e., the item is "I see myself as anxious, easily upset"), so the higher initial score may be reflective of a more symptomatic client than a more emotionally unstable one. Additionally, those who met RSI experienced significantly greater improvement in their emotional stability throughout treatment ($M_{RSI} = -2.2$, $M_{NO.SI} = -0.5$, $p < .001$). Also, they experienced a significantly greater improvement/change in their self-reported quality of life ($M_{RSI} = 3.5$, $M_{NO} = 0.34$, $p < .0001$), hope ($M_{RSI} = 5$, $M_{NO} = 1.3$, $p < .05$) and gratitude throughout treatment ($M_{RSI} = 2.2$, $M_{NO} = 1.2$, $p < .001$). Further, those achieving RSI had significantly greater levels of hope at discharge ($M_{RSI} = 31.7$, $M_{NO} = 29$, $p = .02$). Lastly, participants who failed to achieve RSI had significantly lower levels of hope at discharge ($M_{RSI} = 39$, $\mathrm{M}_{NO} = 35.9$, $p = .02$).

Notably, when participants who achieved RSI at some point throughout treatment (65%) were compared to all other clients (35%), those achieving RSI only differed from all other participants in the following areas: individuals with RSI had significantly lower initial emotional stability at baseline ($M_{RSI} = 7.1$, $M_{NO} = 8.4$, $p < .001$), a significantly greater change in emotional stability throughout treatment ($M_{RSI} = 1.69$, $M_{NO} = 0.58$, $p =$

.02), and a significantly greater change in self-reported quality of life over treatment (M_{RSI} = 3.5, M_{NO} = 0.34, $p < .0001$).

Clients with Reliably and Clinically Significant Deterioration

Of the 311 clients included in the sub-sample, approximately 4% ($n = 13$) of clients experienced a reliably significant deterioration (RSD) during treatment characterized by a significantly higher OQ score at discharge than higher at baseline (i.e., an increase in OQ score by more than 14 points; Lambert, 2015). Relative to participants achieving RSI, those experiencing RSD had a significantly lower initial OQ score (t = 2.80, M_{RSD} = 62.6, M_{RSI} = 76.7, df = 15, p = .01) and a significantly higher final OQ score (t = -7.88, M_{RSD} = 86.8, M_{RSI} = 50.3, df =16, $p < .001$). On average, clients experiencing RSD attended 33 total sessions ($M_{MaxSession}$ = 33, Mdn = 23, SD = 28, *Range* = 4 to 100) across 79 weeks in treatment ($M_{MaxWeeks}$ = 79.49, Mdn = 67.57, SD = 67.57, *Range* = 9 to 279) with sessions occurring on average once every 2 and a half weeks. On average, those who experienced RSD first met the RSD criteria at their 7th session ($M_{Session}$ = 6.85, Mdn = 5, SD = 4.16, *Range* = 3 to 17) and during their 16th week (M_{Week} = 16.46, Mdn = 15, SD = 11.17, *Range* = 2-34). When plotted, those achieving RSD by sessions and weeks both followed a negatively accelerating curve (see Tables 10 and 11).

To assess other factors related to treatment failure, exploratory analyses were conducted for participants with RSD. When compared to those who achieved RSI, those who experienced RSD presented as more symptomatic at one point in treatment, with a significantly higher maximum symptom/OQ score (M_{RSI} = 83, M_{RSD} = 96, p = .04). However, when participants experiencing RSD were compared to participants achieving RSI and all other clients, there were no other notable, significant differences across other predictor variables (i.e., total sessions, frequency, etc.). Further, limited BIL data precluded comparisons between those experiencing and those not experiencing RSD on that measure.

Therefore, additional exploratory analyses were conducted for those experiencing RSD in the larger participant sample (n = 434), which included participants with less than 3 OQ scores. As such, early discharge for these clients may be attributable to early dropout, a mismatch with their therapist, the decision they have reached a good enough level of treatment for them, or other factors.

Clients with Reliably Significant Deterioration and Early Dropouts

Of the 434 participants in the study's larger sample, 101 clients had a significantly higher OQ score in their final session than in their first session. When compared with those experiencing RSI, clients experiencing RSD had many insignificant differences across predictor variables and scales on the BIL. Specifically, when comparing those with RSD to those with RSI, there were no significant differences in measures of emotional stability, gratitude, or hope between groups. However, participants who experienced RSD had a significantly lower initial working alliance with their therapist than all other clients (t = 2.3525, df = 14.157, M_{RSI} = 24.7, M_{RSD} = 21.9; p = .033). As such, consistent with the literature, an early, unsatisfactory therapeutic alliance is associated with poorer treatment outcomes (i.e., early dropout, RSD, etc.).

DISCUSSION

Summary

The purpose of the current study was to compare the DE, traditional GEL, and modified GEL models of client change using two different predictors: the traditional session variables versus a novel variable of time (weeks) in treatment. This study aimed to address gaps in the literature by comparing the model fit across two dose types (sessions and weeks in treatment); assessing multiple model effects (i.e., linear, log-linear, quadratic, and cubic); ascertaining the predictive accuracy of sessions and weeks in within the models; utilizing minimal exclusion and inclusion criteria (i.e., removing the requirement of a clinical symptom score at intake) to increase generalizability; considering interaction effects; and comparing client characteristics and treatment perspectives among clients to determine factors that may be significantly associated with treatment success or deterioration.

Support for the DE, traditional GEL, and Modified GEL models

The present study compared the fit of the DE, traditional GEL, and modified GEL models to empirical data using both session and week variables as predictors. Findings indicated that the traditional GEL model had a significantly better relative fit than the other models across predictors. However, like previous studies, this study found some evidentiary support for all models (i.e., DE, traditional GEL, and modified GEL).

Consistent with the DE model, there was a positive and negatively accelerating curve present for aggregate data and individual response curves between dose (session or week) and effect (percentage of improved patients or individual improvement).

Additionally, a negatively accelerating curve was present for aggregate data when including and excluding early responders (Figures 5.3 and 6.3).

Conversely, many of the findings strongly aligned with the traditional GEL model. First, there was a large, positive relationship between clients' total improvement during treatment and their rate of change (r = .62), and faster responses to treatment were associated with a greater likelihood of clinically significant improvement. Second, there was a significant, negative association between individuals' rates of change and their total number of sessions [$F(1, 309) = 34$, $p < .001$] and total weeks in treatment [$F(2, 308) = 38$, $p < .001$]. This suggests those who improved quickly attended significantly fewer sessions across fewer weeks, which is consistent with the GEL model's theory that progress determines therapy length. Third, there was a large, negative relationship between slope and intercept across both models, which suggests that those who have higher initial symptom scores have faster rates of change (r = .73-.74). Further, upon entry of clinical interaction terms (i.e., depression, hopelessness, and suicidal thoughts) to the models, model fit significantly improved, suggesting that individual presenting problems have a significant impact on treatment. Fourth, there were large, inverse relationships between a client's initial OQ score and their total sessions and frequency. Fifth, there were insignificant relationships between clients' total change in their OQ scores and their total number of sessions/weeks and session frequency, further supporting the GEL model's theory that change rates determine improvement instead of total sessions. Sixth and most importantly, the traditional GEL model had a significantly better fit to the data than the other MLMs across both weeks and session predictors.

Slightly consistent with Reece et al.'s (2011) modified GEL model, there was a slightly significant interaction and significant correlations between total sessions/weeks and frequency across models ($r_{\text{Sessions}} = 0.60$; $r_{\text{Weeks}} = 0.67$; $p < .05$). However, this interaction term did not significantly improve the traditional GEL model fit, which was deemed to better fit the data. There were different findings for the interaction of predictor variables with OQ scores across weeks and sessions. Specifically, in the week's models, there were large, inverse relationships between a client's initial OQ score and their total sessions and frequency, and there was a positive, large relationship between session frequency and total sessions. However, in the sessions model, the total number of sessions and frequency of sessions (e.g., weekly, biweekly, etc.) had an insignificant impact on the client OQ score and on the likelihood of participants achieving RSI.

Inconsistent with both models, the present study found atypical effects for some variables. Although the DE posits that baseline severity has an insignificant effect while the GEL model posits that it has a positive effect, the present study found that baseline severity had a significant, inverse effect on total number of sessions (i.e., those with higher baseline symptom scores tended to attend fewer sessions; $r = -.55$) which is inconsistent with previous reviews and likely due to lesser exclusion criteria in the present study (Bone et al., 2021; Robinson et al., 2019). Further, the present study's results suggest that baseline symptom score has a significant impact on many additional psychotherapy factors for clients. Specifically, baseline symptom score had a large, inverse relationship with total weeks in treatment ($r = -.74$) and frequency of treatment ($r = -.87$), so more symptomatic patients at intake were more likely to attend less frequently and over a shorter number of sessions and weeks.

Comparing the Predictive-Accuracy of Session- and Week-based Predictors

The present study compared the MLMs of both weeks and sessions and then created multi-level models based on data partitions to assess their predictive model fit and accuracy. Findings indicate that the week variable had greater predictive accuracy and a better model fit, which suggests that the time between sessions and time spent in treatment are meaningful. However, potential inaccuracies in the recording of session data may have influenced these results.

Interaction Effects

Further, this study investigated interaction effects within the models. Some preliminary, non-diagnostic data provided by participants were found to have a significant effect on model fit and other patient variables. Specifically, a history of a suicide attempt and client endorsement of current depression, hopelessness, suicidal thoughts, and attention problems at intake had a significant effect ($p < .05$) on either or both the weeks and session models. The inclusion of these interactions greatly improved overall model fit and helped to explain 98.7% of the variance in OQ scores. However, of all the interactions, only depression, hopelessness, and suicidal thoughts had a highly significant impact ($p < .001$) on both session and weeks models. Further, two of the interactions, depression and hopelessness, had large associations with baseline scores [(Depression: $r_{\text{B0-Wks}} = .65$ $r_{\text{B0 - Sess}} = .62$) & (Hopeless: $r_{\text{B0-Wks}} = .64$, $r_{\text{B0 - Sess}} = .61$)] and with each other ($r = .75$). As such, individuals struggling with depression and hopelessness may be more symptomatic than clients who are not struggling with these symptoms.

Variables Significantly Associated with Improvement and Deterioration

Compared to clients who were unable to achieve reliably significant improvement (RSI), clients who achieved RSI had significantly higher initial symptom scores, lower final scores, and faster change rates per session and week. Further, participants who achieved RSI were significantly less emotionally stable at baseline (which was likely a reflection of anxious distress), and patients who achieved RSI had a significantly greater improvement in their emotional stability, quality of life, gratitude, and hope throughout treatment. Further, patients who achieved RSI had significantly higher levels of emotional stability and hope at discharge compared to other clients. Notably, when comparing those who experienced RSI at any point in time (i.e., not strictly at discharge) to all other clients, only the lower baseline emotional stability and significant improvements in emotional stability and quality of life are present.

Further, when comparing individuals who get significantly worse during treatment (or those with RSD) to those achieving RSI or to all other clients, most demographic and clinical variables were insignificant. However, patients with RSD had a significantly lower appraisal of their initial working alliance with their therapist relative to all other clients. This is consistent with the literature, which also suggests that an unsatisfactory therapeutic alliance is associated with poorer treatment outcomes.

Clinical Implications

The present study has many clinical implications. First, consistent with many other DE and GEL comparison studies, evidentiary support was found for both models. Many effects found in this study are consistent with previous studies (i.e., consistent with the GEL model, there are often significant differences in rates of change across clients and consistent with the DE model, individual curves are often non-linear). Therefore, researchers should consider creating a new model with the most replicated aspects of the DE and GEL models. Thus, instead of researchers finding and then ignoring mixed support across models in favor of a binary conclusion, a new model can improve the level of accuracy and fit to the data and better inform studies on optimal treatment conditions.

Second, since the present study suggests that many factors significantly impact theoretical model fit and treatment outcomes (i.e., depressive symptoms, hopelessness, appraisal of therapeutic alliance, etc.), it is important to consider that any model (including the current DE and GEL models) may be limited in its utility for individual clients. Therefore, researchers may want to consider the benefits and costs of including many of these factors or considering an individualized approach to determining optimal session variables for each client for effective treatment.

Third, since the present study suggests that both clients' weeks in treatment and sessions are important, both in-session and out-of-session time should be used wisely, as homework assignments given to clients outside of sessions may have a significant additive benefit to treatment outcomes. Fourth, since the present study supports the notion that individual client rates of change differ significantly and are rarely linear, it is important for therapists to remember that symptom scores may vary over time due to individual differences and are not necessarily reflective of an ineffective therapist.

Fifth, comparisons among those who achieved RSI, failed to achieve RSI, and experienced RSD provides important clinical insight. While therapists should create collaborative goals focused on decreasing symptoms with clients, treatment goals should also include improving patient functioning and well-being, as those who achieved RSI had significantly greater improvements in their emotional stability, quality of life, gratitude, and hope during treatment compared to those without RSI. Additionally, consistent with previous literature, the present study suggests that patients experiencing significant deterioration tend to have a significantly poorer appraisal of their working alliance with their therapist in early treatment, so therapists should consider utilizing routine outcome measures (ROMs) to regularly assess this and actively work to improve the relationship if the alliance is poor. Further, in the present study, participants who experienced RSD at discharge met the criteria for a statistically significant worsening of symptoms much earlier than treatment termination, as they stayed in treatment for an average of 33 sessions and 79 total weeks but experienced a significant deterioration during session 7 and week 16. As such, ROMs should be used regularly for treatment, and a trend toward possible deterioration should be monitored and swiftly addressed.

School Implications

Further, the present study has implications for school-based counseling through practice and policy, but the rigidity and laws regarding the provision of school counseling services and the lack of funding may impede implementing changes consistent with the findings. For example, although the present study found that session frequency, total session number, and total weeks in treatment are important, the regularity of school counseling services is not easily changeable, as the number of counseling sessions utilized per student can be limited by the number of weeks in a school year (i.e., typically 36), the length of sessions can be limited by a school district's policy (e.g., 30 to 45 minutes), and the frequency of sessions and overall access to mental health providers may limit mental-health services. Although 18% to 22% of school-age children experience mental health problems, only 7% to 16% of those children receive mental health services which are largely provided by schools (Maag & Katsiyannis, 2010). This discrepancy is likely related to understaffing, as only 17% and 17.8% of school districts across the United States meet the recommended psychologist or counselor to students' ratios determined by their national organizations (Farmer et al., 2021; Gagnon & Mattingly, 2016). Limited providers and limited access to counseling services likely further impede changes in the provision of services (i.e., changing the frequency, number of sessions, etc.), especially for students who are not guaranteed mental health services through special education law.

Also, while the present study suggests that baseline severity has significant correlations with specific symptoms (e.g., depression, hopelessness, etc.) which can meaningfully inform counseling parameters, schools are often limited in the information they can collect via standardized or written measure from students without extensive parental consent. Therefore, although baseline severity would likely inform appropriate counseling support, standardized measures are likely infrequently given.

However, there are many important implications that can be utilized by providers despite the limitations associated with a school system. First, since the present study deems the traditional good-enough level model to be the best fitting model, these findings suggest that individual clients are best positioned to decide when a good-enough level of counseling has been achieved. Although the present study includes adults and findings from children may not be replicated, it is likely that some students still have a sense of a good enough level of support that is meaningful. As such, when making decisions about the amount of counseling a student may receive [i.e., especially when writing an individualized education plan (IEP)], student input regarding the provision of related services should be sought and considered. Importantly, if a student is given counseling services through an IEP, the student cannot stop counseling when they feel they have had "good enough benefits," as IEPs are legal documents that must be followed with fidelity. So, the student should be a member of the Committee on Special Education (CSE) when developmentally appropriate, and their opinion on counseling services should be considered when making a collaborative decision on the provision of counseling services.

Second, when providers develop rapport with students, it is important they develop a strong, initial therapeutic alliance that is closely monitored, as a poor alliance has an increased risk of leading to negative outcomes like deterioration. If providers notice deterioration in a student who receives mandated services through an IEP, counseling providers should first consider changing their approach. If deterioration continues, providers should consider holding a CSE meeting to make changes to the students' IEP if the service is no longer required or if the student is highly resistant to the service despite motivational techniques and changes are needed. As noted in this study, a poor alliance likely worsens student outcomes more than if treatment is withheld, as it led to a deterioration in functioning. As such, providing counseling services against a student's wishes likely does actual harm to the student. So, a discontinuation of services should be considered if students indicate a poor therapeutic relationship or belief that they have received a good-enough level of counseling service, as overriding their decisions can negatively impact their functioning.

Third, as shown in the present study, an individual's time in treatment is important regardless of their total number of sessions, and this should also be considered when determining the provision of all services for students. For example, some students, such as students with autism spectrum disorder who aim to improve social skills, may benefit from fewer counseling sessions with greater support generalizing their skills in the classroom. Similarly, students with internalizing disorders may benefit from time to practice their new skills in vivo or engage in interventions outside of session such as behavioral activation. As such, school counselors should remember that both time and sessions are important and should be intentional.

Fourth, the present study suggests that current suicidal thoughts, hopelessness, and depression affect theoretical model fit and may impact the level of service warranted for individual students. Since students may be hesitant to discuss these topics due to limits to confidentiality, counselors should be hyper-aware to notice depressive symptoms given their significant importance. If students endorse these symptoms or appear to be at an increased risk of these symptoms, school providers should consider referring the student and their parent to an additional outpatient provider.

Limitations

There are many limitations of the present study. First, a significant limitation of the present study is the unknown extent of inaccuracies recording session data. Earlier in data collection (i.e., prior to 2022), many sessions were not recorded unless an outcome measure was given. As such, it is likely the session variable is inaccurate for many clients. If the sessions variable was accurate, it may have impacted the predicted session model and overall value of sessions as a predictor compared to weeks. Although some data points may be missing for weeks, weeks are calculated automatically based on times, so the weeks data are less impacted by these data collection problems. Second, at this clinic, many clients began treatment as children before transitioning into adult clients. Since only data for individuals over 18 was included in the study, client data for those who began treatment as a child is likely inaccurate, as their initial OQ score reflects their first score following adulthood, not their first score in treatment. Third, since clients supplied information regarding interaction effects instead of therapists, the interaction effects may not replicate across studies and may not represent clinical symptoms, as the client's interpretation influences them. Fourth, given the naturalistic nature of the data, there may be lower internal validity within this study than present in other studies. Fifth, given that only adults were included in the present study, it is unknown whether these findings may generalize to child clients or students.

Future Research

The results of the present study indicate the traditional, log-linear GEL model had the best relative fit to the data for both week and session predictors. When compared, week variables were more accurate predictors than session variables; however, inaccurate measurements of the session variable may have impacted our results. Depressive symptoms (i.e., hopelessness, suicidality, previous suicide attempts, and attention problems) had a significant, positive relationship with initial symptom scores and improved the traditional GEL model fit. Additionally, low emotional stability and high symptom scores at baseline and improved emotional stability, hope, gratitude, and quality of life during treatment were all associated with better treatment outcomes. Low initial client appraisals of the working alliance were associated with deterioration and drop-out.

Future studies should aim to continue to add to the robustness of the DE and GEL model literature. First, researchers should consider the impact of time in treatment within all future modeling. Second, researchers should consider including participants whose characteristics are generalizable to many settings (i.e., using community-based samples). Third, researchers should consider investigating model fit to child client samples to see if the conclusions may be extended to that population and the implications findings may have on the provision of school-based services. Importantly, to get a true understanding of the model fit for these clients, children should be permitted to initiate treatment termination. However, the child may benefit from discussing thoughts about treatment termination with their parents and provider. Although premature termination is difficult to accurately calculate, research suggests that adult clients typically prematurely terminate psychotherapy approximately 25% of the time, but the rate is highest for young clients and those treated in university clinics (Swift & Greenburg, 2012). As such, it is important to include parent and therapists in decisions of termination. Fourth, further studies should consider assessing the predictive value of sessions and weeks and identifying potential interactive effects for more robust conclusions. Fifth, future studies should consider integrating the DE and GEL models to include frequently replicated aspects to improve their accuracy.

Table 1. *Key differences between the DE and GEL Models*

DE Model	GEL Model
There is a positive and negative accelerating curve present for aggregate data and individual response curves between dose (session number) and effect (percentage of improved patients or individual improvement)	A negatively accelerating curve is only present for aggregate data, is an artifact of the aggregate, and can be attributed to differing dropout rates. Linear relationships better represent individual response curves.
The rate of change is not significantly different across patients	The rate of change is significantly different across patients, as different patients have different responses to therapy (i.e., early responders typically have the highest rates of change)
A positive, significant relationship exists between session number and the percentage of improved patients.	Improvement is not associated with total sessions, as individuals can have the same improvement with fewer sessions.
The number of sessions determines the likelihood of progress and termination.	The rate of progress determines the number of sessions (responsive model)

Table 2. *Characteristics of Studies that Directly Compare DE and GEL Models*

Authors	Setting	Sample size (n)	Mean Session # & Session Limit	Mean & Range of Time in Treatment	Exclusion Criteria for Initial Sx Score	Primarily Supported
Baldwin et al., 2009	US UCC	4,676	M = 6.46; max = 40	N/A, max = 40 weeks	Clinical Range (> 63) OQ	GEL
Falkenström et al., 2016	Swedish primary (PC) & psychiatric care	n_{PC} = 640 n_{Psych} = 284	M_{PC} = 6 & M = 9.1; max_{Psych} = 12			GEL
Lee et al., 2022	VA PC & Specialty Mental Health Clinics (SMHC)	13,647	M_{PC} = 3.7 & M_{SMHC} = 6.6; max_{PC} = 10 (1/2 hr) max_{SMHC} = 20 (1 hr)	M_{PC} = 71.8 days SD_{PC} =49.4 M_{SMHC} = 71.8 days SD_{SMHC} =49.4	Clinical Range (> 9) PHQ	GEL
Nordmo et al., 2021	Norway Outpatient clinics	362	M = 52; max = 40 in one, N/A in others	N/A	N/A	GEL
Owens et al., 2016	US UCC	13,664	M = 9.04; max = 100	N/A	Clinical Range on all scales of BHM	GEL
Reese et al., 2011	US UCC	1,207	M = 5; max = 15 for 90% of clients		N/A	GEL
Schuler et al., 2022	US Outpatient Clinic	263	M = 12.19; max = 38	N/A	Clinical Range OQ (>63)	DE
Niileksela et al., 2021	US UCC	64,319	M = 7.86; max = N/A; >13 stratified together	N/A	Clinical range on Sx Scales from CCAPS-34	DE (both)
Stultz et al., 2013	Primarily US UCC	6,375	M = 5; max = 20	$M_{frequency\ (f)}$ = 13.3 days SDf= 9.6 $Range_f$ = 7-14.5	Clinical range BHM	DE (both)

Notes: United States College Counseling Center (US CCC), Mean *(M)*, Session Limit (max), Hour (hr), Outcomes Questionnaire (OQ), Patient Health Questionnaire (PHQ), Behavioral Health Measure (BHM), Counseling Center Assessment of Psychological Symptoms (CCAPS-34), Dose Effect (DE), Good Enough Level (GEL)

Table 3. *Sample Distribution of Participants by Total Weeks and Sessions in Psychotherapy*

Total Sessions/Weeks	Total Weeks		Total Sessions	
	Number of Participants (n)	Percentage of Participants	Number of Participants (n)	Percentage of Participants
3	5	1.6	19	6.1
4	6	1.9	15	4.8
5	8	2.6	14	4.5
6	3	1	14	4.5
7	5	1.6	5	1.6
8	4	1.3	16	5.1
9	4	1.3	16	5.1
10	5	1.6	10	3.2
11	7	2.3	7	2.3
12	9	2.9	11	3.5
13	8	2.6	5	1.6
14	4	1.3	12	3.9
15-16	8	2.6	12	3.9
17-18	5	1.6	12	3.9
19-20	3	1	9	2.9
21-22	6	1.9	14	4.5
23-24	10	3.2	6	1.9
25-26	10	3.2	8	2.6
27-28	4	1.3	7	2.3
29-30	7	2.3	7	2.3
31-32	10	3.2	5	1.6
33-34	8	2.6	4	1.3
35-36	5	1.6	6	1.9
37-38	3	1	4	1.3
39-40	8	2.6	3	1
41-45	14	4.5	15	4.8
46-50	10	3.2	5	1.6
51-55	11	3.5	14	4.5
56-60	5	1.6	7	2.3
61-65	11	3.5	3	1
66-70	11	3.5	8	2.6
71-80	12	3.9	1	0.3
81-90	15	4.8	3	1
91-100	16	5.1	7	2.3
101-150	30	9.6	4	1.3
151-200	7	2.3	3	1
201-250	6	1.9	0	0
251-300	5	1.6	0	0
301-326	3	1	0	0

Table 4. *Aggregate (Dose-Effect) Linear, Quadratic, Cubic, and Log-Linear two-level multilevel models predicting changes in OQ Scores by Sessions (n=311; Study 1)*

Session Parameter	**Linear**	**Log-Linear****	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at baseline (intercept; B_{00})	67.84***	67.96***	67.96***	68.77***
Mean linear growth rate (x_{ij}/B_{100})	-0.51***		-0.53***	-0.67***
Session^2 (x^2_{ij})				0.006***
Session^3 (x^3_{ij})				
Session$_{\log}$ ($x_{\log ij}$)		-0.015***		
Random effects				
Residual variance (e_{ij})	94.47	0.048	94.51	94.04
Intercept variance (u_{0i})	464	0.168	463	466
Linear variance (u_{1i})	0.57	0.001	0.56	0.48
Correlation (B_{00j}, u_{1i})	-0.06	0.23	-0.06	-0.05
Covariance (B_{00j}, u_{1i})	-0.976	0.003	-0.963	-0.750
Model fit				
Conditional-R^2 (fixed & random effects)	0.939	0.976	0.936	0.929
REML Criterion	45987	795	45998	45971
Deviance (*-2LL*)	-22993	-389.5	-22991	-22966
ΔX^2 (*df*)	-	45206 (0)	0 (1)	50 (1)***
BIC	46039	847	46059	46041
AIC	45999	807	46012	45987

Note: **** Log-Linear values are provided using non-log, original scale
Significance codes: *** = < 0.0011; ** = < 0.01; * = < 0.05

Table 5. *Aggregate (Dose Effect), Two-Level Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear)*

	Intercept (B_{00})***
Session (x_{ij})***	-0.077

Magnitude codes: **Large, *Moderate,*** Small/Negligible correlation

Table 6. *Traditional two-level Good-Enough Level Linear, Log-Linear, Quadratic, and Cubic, multilevel models predicting changes in OQ Scores by Sessions (n=311; Study 1)*

Session Parameter	**Linear**	**Log-Linear***	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at baseline (intercept; B_{00})	68.26***	65.16***	68.60***	68.77***
Mean linear growth rate (x_{ij}/B_{100})	-0.94***	-0.97***	-1.08***	-1.14***
#Sessions (maxx_{ij})	0.006	1.00	0.069***	0.03
#Sessions * Session (x_{ij} * maxx_{ij})	0.01***	1.00***	0.011***	0.01***
Session^2 (x^2_{ij})			0.008***	-0.01***
Session^3 (x^3_{ij})				0.000
Session$_{\log}$ ($x_{\log ij}$)		-0.03***		
Random effects				
Residual variance (e_{ij})	94.35	1.049	93.78	93.75
Intercept variance (u_{0i})	466	1.19	468	468
Linear variance (u_{1i})	0.47	1.01	0.44	0.42
Correlation (B_{00j}, u_{1i})	-0.07	0.28	-0.07	-0.07
Covariance (B_{00j}, u_{1i})	-1.05	0.003	-1.00	-0.99
Model fit				
Conditional-R^2 (fixed & random effects)	0.942	0.978	0.930	0.927
REML Criterion	45964	796	45946	45995
ΔX^2 (*df*)	-	0 (1)	0 (2)	2.77 (2)
Deviance (*-2LL*)	-22974	-375	-22946	-22945
BIC	46034	767	46033	46099
AIC	45981	820	45966	46019

**** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.0011; ** = <0.01; * = <0.05

Table 7. *Traditional two-level Good-Enough Level Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear) by Session*

	Intercept (B_{00})***	Session (x_{ij})***
Session (x_{ij})***	-0.011	
#Sessions (maxx_{ij})*	**-0.692**	-0.051

Note: Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation
Significance codes: *** = < 0.001; ** = < 0.01; * = < 0.05

Table 8. *Good-Enough Model Modified by Sessions and Modeled with Linear, Quadratic, Cubic, and Log-Linear two-level models predicting changes in OQ Scores (n = 311; Study 1)*

Session Parameter	**Linear**	**Log-Linear***	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at baseline (B_{00})	72.68***	68.72***	76.05***	84.58***
Mean linear growth rate (x_{ij}/B_{100})	-0.29*	-0.995*	-0.38	-0.47*
#Sessions (maxx_{ij})	-0.37	1.00	-0.35	-0.31
Frequency$_{S\ (Sessions)}$ ($freqx_{ij}$)	-1.83	-0.974	-3.59	-11.70
#Sessions * Session (x_{ij} * maxx_{ij})	-0.002	-1.00	-0.002	-0.003*
Frequency$_S$ * Session (x_{ij} * $freqx_{ij}$)	-0.24**	-0.99***	-0.24**	-0.22**
Frequency$_S$ * #Sessions ($freqx_{ij}$*maxx_{ij})	0.14	1.00*	0.16*	0.16*
($freqx_{ij}$ * x_{ij} * maxx_{ij})	0.006*	1.00**	0.006*	0.006*
Session^2 (x^2_{ij})			0.0003	0.005***
Session^3 (x^3_{ij})				0.00002***
Random effects				
Residual variance (e_{ij})	94.20	1.049	94.20	93.67
Intercept variance (u_{0i})	458	1.182	460	244
Linear variance (u_{1i})	0.48	1.00	0.48	
Session (x_{ij})				0.46
#Sessions (maxx_{ij})				0.08
Frequency$_{S\ (Sessions)}$ ($freqx_{ij}$)				0.41
Correlation (B_{00j}, u_{1i})	-0.09	0.25	-0.09	
Session (x_{ij})				-0.96
#Sessions (maxx_{ij})				1.00
Frequency$_{S\ (Sessions)}$ ($freqx_{ij}$)				-0.19
Model fit				
Conditional-R^2 (fixed & random effects)	0.937	0.974	0.938	0.932
REML Criterion	45962	815	45989	45990
Deviance (*-2LL*)	-22973	-361.35	-22963	-22942
ΔX^2 (*df*)		42506 (0)	0 (3)	44 (9)***
BIC	46033	827.06	46057	46092
AIC	45953	746.70	45957	45931

Note: **** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.0011; ** = <0.01; * = <0.05

Table 9. *Modified Good-Enough Model Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear) by Session*

	Intercept (B_{00})***	Session (x_{ij})	#Sessions (maxx_{ij})*
Session (x_{ij})	-0.08		
#Sessions (maxx_{ij})*	***-0.55***	0.001	
Frequency ($freqx_{ij}$)	**-0.83**	0.11	**0.60**

Note: Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 10. *Comparing the superior Log-Linear Models across Dose-Effect, Good-Enough Level, and Modified Good-Enough Level models Using Session Variables as Predictors*

Session Parameter	**DE Model**	**GEL model*****	Modified **GEL**
Fixed effects			
Mean OQ-45 at baseline (intercept; B_{00})	67.96***	65.94***	68.72***
Mean linear growth rate (x_{ij}/B_{100})	-0.985***	-0.97***	1.001
#Sessions (maxx_{ij})		1.00	-0.995*
Frequency$_{S\ (Sessions)}$ ($freqx_{ij}$)		1.00***	-0.974
#Sessions * Session (x_{ij} * maxx_{ij})		-1.00***	-1.00
Frequency$_S$ * Session (x_{ij} * $freqx_{ij}$)		1.00*	-0.99***
Frequency$_S$ * #Sessions ($freqx_{ij}$*maxx_{ij})		-0.03***	1.00*
($freqx_{ij}$* x_{ij} * maxx_{ij})			1.00**
Session$_{log}$ (x_{logij})	-0.015***	0.032***	
Random effects			
Residual variance (e_{ij})	1.049	1.049	1.049
Intercept variance (u_{0i})	1.18	1.19	1.182
Linear variance (u_{1i})	1.00	1.01	1.00
Correlation (B_{00}, slope | ID)	0.23	0.28	0.25
Covariance (B_{00}, slope | ID)	0.003	0.003	0.003
Model fit			
Conditional-R^2 (fixed & random effects)	0.976	0.973	0.974
Deviance (*-2LL*)	795	848	815
ΔX^2 (*df*)	-	96 (6)***	0 (9)
REML	-389.5	-342	-361.35
BIC	847	820	827.06
AIC	807	788	746.70
Notes: **** Log-Linear values are provided using non-log, original scale; Significance codes: *** = < 0.0011; ** = < 0.01; * = < 0.05			

Table 11. *Comparing Linear, Quadratic, or Cubic Best Fit Models Across Theories*

Session Parameter	**DE Model$_{CUBIC}$**	**GEL model$_{QUAD}$*****	Modifi**ed GEL$_{CUBIC}$**
Model fit			
Conditional-R^2	0.927	0.937	0.932
Deviance (*-2LL*)	-22945	-22946	-22942
ΔX^2 (*df*)		89.98 (3)***	8.55 (14)
BIC	46099	45979	46092
AIC	46019	45912	45931

Notes: Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 12. *Aggregate (Dose-Effect) Linear, Quadratic, Cubic, and Log-Linear two-level multilevel models predicting changes in OQ Scores by Weeks (n=311; Study 1)*

Week Parameter	**Linear**	**Log-Linear***	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at baseline (B_{00})	67.34***	62.15***	67.51***	67.97***
Mean linear growth rate (x_{ij}/B_{100})	-0.20***	-1.00***	-0.21***	-0.26***
Week^2 (x^2_{ij})			0.0001**	0.0009***
Week^3 (x^3_{ij})				-0.000002***
Week$_{log}$ (x_{logij})		-0.005***		
Random effects				
Residual variance (e_{ij})	94.27	1.049	94.238	94.53
Intercept variance (u_{0i})	454.41	1.19	455.83	430.18
Linear variance (u_{1i})	0.09	1.00	0.09	0.08
Correlation (B_{00j}, u_{1i})	-0.07	0.22	-0.07	-0.05
Covariance (B_{00j}, u_{1i})	-0.449	0.0009	-0.443	-0.295
Model fit				
Conditional-R^2 (fixed & random effects)	0.922	0.955	0.920	0.913
REML Criterion	45959	655	45969	45980
ΔX^2 (*df*)	-	45318 (0)	0 (1)	17 (1)***
Deviance (*-2LL*)	-22978	-318.3	-22974	-22965.2
BIC	46007	689	46009	46000
AIC	45967	649	45962	45946

Notes: **** Log-Linear values are provided using non-log, original scale; Significance codes: *** = < 0.0011; ** = < 0.01; * = < 0.05

Table 13. *Aggregate (Dose Effect), Two-Level Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear) by Weeks*

	Intercept (B_{00})***
Week (x_{ij})***	0.084

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 14. *Traditional two-level Good-Enough Level Linear, Quadratic, Cubic, multilevel models predicting changes in OQ Scores by Weeks*

Week Parameter	**Linear**	**Log-Linear***	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at baseline (intercept; B_{00})	66.59**	61.98***	66.95***	67.85***
Mean linear growth rate (x_{ij}/B_{100})	-0.36***	-0.99***	-0.428***	0.023***
#Weeks (maxx_{ij})	0.019***	1.00	0.026	-0.565
#Weeks * Week (x_{ij} * maxx_{ij})	0.0018***	1.00***	0.0018***	0.002***
Week^2 (x^2_{ij})			0.0015***	-0.00002***
Week^3 (x^3_{ij})				- 0.00000005***
Week$_{\log}$ ($x_{\log ij}$)		-0.009***		
Random effects				
Residual variance (e_{ij})	94.28	1.048	93.97	93.43
Intercept variance (u_{0i})	457.23	1.19	453.35	456.07
Linear variance (u_{1i})	0.08	1.00	0.07	0.06
Correlation (B_{00j}, u_{1i})	-0.12	0.21	-0.11	-0.11
Covariance (B_{00j}, u_{1i})	-0.71	0.0008	-0.62	-0.59
Model fit				
Conditional-R^2 (fixed & random effects)	0.919	0.955	0.908	0.900
REML Criterion	45946		45950	45971
ΔX^2 (*df*)		45307 (0)	0 (2)	42 (2)***
Deviance (*-2LL*)	-22961	-307.6	-22942	-22920
BIC	45992	685	45970	45945
AIC	45938	631	45903	45865

**** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.0011; ** = <0.01; * = <0.05

Table 15. *Traditional two-level Good-Enough Level Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear) by Weeks*

	Intercept (B_{00})***	Week (x_{ij})***
Week (x_{ij})***	0.02	
#Weeks (maxx_{ij})	**-0.71**	-0.04

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 16. *Good-Enough Model Modified by Weeks and Modeled with Linear, Quadratic, Cubic, and log-linear two-level models predicting changes in OQ Scores*

Week Parameter	**Linear**	**Log-Linear***	**Quadratic**	Cubic
Fixed effects				
Mean OQ-45 at intercept (B_{00})	73.70***	69.33***	71.91***	81.24***
Mean linear growth rate (x_{ij}/B_{100})	0.42***	-0.99***	-0.42***	0.024
#Weeks ($\max x_{ij}$)	-0.089	-1.00	-0.085	0.035
$\text{Frequency}_{W\ (\text{Weeks})}$ ($freqx_{ij}$)	-2.80*	-0.96	-1.80	-9.53
#Weeks * Week (x_{ij} * $\max x_{ij}$)	0.0026**	0.99*	0.0025**	0.0022**
Weeks_S * Week (x_{ij} * $freqx_{ij}$)	0.02	-1.00	0.020	0.017
Weeks_S * #Weeks ($freqx_{ij}$*$\max x_{ij}$)	0.041*	1.00*	0.042*	-0.021
($freqx_{ij}$ * x_{ij} * $\max x_{ij}$)	-0.00028	-1.00	-0.00028	-0.00025
Week^2 (x^2_{ij})			0.00011*	0.00079***
Week^3 (x^3_{ij})				-0.000002***
$\text{Week}_{\log}$ ($x_{\log ij}$)		-0.0086***		
Random effects				
Residual variance (e_{ij})	94.24	0.048	94.21	94.11
Intercept variance (u_{0i})	448.15	0.17	449.78	218.3
Linear variance (u_{1i})	0.079	0.000093	0.078	
Week (x_{ij})				0.075
#Weeks ($\max x_{ij}$)				0.0083
$\text{Frequency}_{W\ (\text{Weeks})}$ ($freqx_{ij}$)				0.23
Correlation (B_{00j}, u_{1i})	-0.10	0.22	-0.11	
Week (x_{ij})				-0.29
#Weeks ($\max x_{ij}$)				-0.88
$\text{Frequency}_{W\ (\text{Weeks})}$ ($freqx_{ij}$)				1.00
Model fit				
Conditional-R^2 (fixed & random effects)	0.92	0.96	0.92	0.91
REML Criterion	45966	717	45996	46034.7
ΔX^2 (*df*)	-	45307 (0)	4.74 (3)	15.74 (10)***
Deviance (*-2LL*)	-22973	-301.4	-22955	-22948
BIC	46020	712	46041	46112
AIC	45939	632	45941	45945

**** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.001; ** = <0.01; * = <0.05

Table 17. *Modified Good-Enough Model Fixed Effects Correlation Matrix using the Model of Best Fit (Log-Linear)*

	Intercept (B_{00})***	Week (x_{ij})	#Weeks ($\max x_{ij}$)*
Week (x_{ij})	-0.003		
#Weeks ($\max x_{ij}$)*	**-0.74**	-0.03	
Frequency ($freqx_{ij}$)	**-0.87**	0.004	**0.67**

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 18. *Comparing the superior Log-Linear models across Dose-Effect, Good-Enough Level, and Modified Good-Enough Level models Using Weeks Variables as Predictors*

Session Parameter	**DE Model**	**GEL model***	Modified **GEL**
Fixed effects			
Mean OQ-45 at baseline (intercept; B_{00})	62.15***	61.98***	69.33***
Mean linear growth rate (x_{ij}/B_{100})	-1.00***	-0.99***	-0.99***
#Week (maxx_{ij})		1.00	-1.00
Frequency$_{S\ (Sessions)}$ ($freqx_{ij}$)		1.00***	-0.96
#Weeks * Weeks (x_{ij} * maxx_{ij})			0.99*
Frequency$_S$ * Week (x_{ij} * $freqx_{ij}$)			-1.00
Frequency$_S$ * #Week ($freqx_{ij}$*maxx_{ij})			1.00*
($freqx_{ij}$* x_{ij} * maxx_{ij})			-1.00
Session$_{log}$ (x_{logij})	-0.005***	-0.009***	-0.0086***
Random effects			
Residual variance (e_{ij})	1.049	1.048	0.048
Intercept variance (u_{0i})	1.19	1.19	0.17
Linear variance (u_{1i})	1.00	1.00	0.000093
Correlation (B_{00}, slope | ID)	0.22	0.21	0.22
Model fit			
Conditional-R^2 (fixed & random effects)	0.955	0.955	0.96
Deviance (*-2LL*)	-318	-308	-301
ΔX^2 (*df*)	45318 (0)	45307 (1)	45307 (1)
REML	655		717.4
BIC	689	685	712
AIC	649	631	632

**** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.0011; ** = <0.01; * = <0.05

Table 19. *Comparing Linear, Quadratic, or Cubic Best Fit Models Across Theories*

Session Parameter	**DE Model**$_{CUBIC}$	**GEL model**$_{CUBIC}$	Modified **GEL**$_{CUBIC}$
Model fit			
Conditional-R^2	0.932	0.900	0.91
Deviance (*-2LL*)	-22942	-22920	-22948
ΔX^2 (*df*)	44 (9)	42 (2)	15.74 (10)
BIC	46092	45945	46112
AIC	45931	45865	45945

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 20. *Comparing the Superior, traditional GEL Model (Log-Linear) for Weeks and Session with each other and their respective predicted models*

Week Parameter	$\text{Session}_{\text{Actual}}$	$\text{Weeks}_{\text{Actual}}$	$\text{Session}_{\text{Predicted}}$	$\text{Weeks}_{\text{Predicted}}$
Fixed effects				
Mean OQ-45 at baseline (intercept; B_{00})	65.16***	61.98***	58.754***	56.90***
Mean linear growth rate (x_{ij}/B_{100})	-0.97***	-0.99***	-0.99***	-1.00**
#Weeks (maxx_{ij})	1.00	1.00	1.00	1.00
#Weeks * Week (x_{ij} * maxx_{ij})	1.00***	1.00***	1.00*	1.00*
Week_{log} ($x_{\text{log}ij}$)	-0.03***	-0.009***	-0.005***	-0.002**
Random effects				
Residual variance (e_{ij})	1.049	1.048	1.069	1.065
Intercept variance (u_{0i})	1.19	1.19	1.30	1.29
Linear variance (u_{1i})	1.01	1.00	1.00	1.00
Correlation (B_{00j}, u_{1i})	0.28	0.21	-0.52	-0.49
Model fit				
Conditional-R^2 (fixed & random effects)	0.973	0.955	0.7979	0.7932
REML Criterion	796		707	677.9
ΔX^2 (*df*)		137 (0)	-	6.7 (0) ***
Deviance (*-2LL*)	-375	-307.6	-353	-357
BIC	767	685	761	746
AIC	820	631	723	729
Pseudo R^2			0.695	0.686
Real Mean Square			63.9955	63.9948

**** Log-Linear values are provided using non-log, original scale
Significance codes: *** = <0.0011; ** = <0.01; * = <0.05

Table 21. *Predicted Fixed Effects of the traditional GEL Log-Linear Model by Session*

	Intercept (B_{00})***	Session (x_{ij})
Session (x_{ij})	-0.27	
#Sessions (maxx_{ij})*	**-0.73**	0.05

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 22. *Predicted Fixed Effects of the traditional GEL Modified Model (Log-Linear) by Weeks*

	Intercept (B_{00})***	Week (x_{ij})
Week (x_{ij})	-0.26	
#Weeks (maxx_{ij})*	**-0.74**	0.05

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

Table 23. *Interactions impacting the superior GEL models by Weeks and Sessions*

Current Symptoms & Psychiatric History (Hx)	**Weeks as a Predictor** Estimate	t-Score	p-value	**Sessions as a predictor** Estimate	t-Score	p-value
Emergent Problem	0.13	0.36	0.72	0.006	0.17	0.87
Current Treatment	1.07	0.58	0.56	0.49	0.70	0.48
Hx of Treatment	1.06	-0.009	0.99	0.0008	0.04	0.97
Medication	-0.95	-0.73	0.47	-0.58	-0.86	0.39
Trauma	-0.99	0.022	0.98	-0.008	-0.16	0.87
Anxiety	0.98	-0.152	0.879	-0.019	-0.33	0.74
Obsessive Thoughts	1.00	0.123	0.9	0.004	0.10	0.92
Depression***	1.21***	3.56***	0.0004***	0.129**	3.34	0.0009***
Sleep Problems	-0.99	-0.234	0.815	-0.02	-0.40	0.69
Eating Problems	1.04	0.881	0.379	0.05	0.83	0.41
Social Problems	1.01	0.283	0.778	0.01	0.26	0.79
Hopeless***	1.22***	3.85***	0.0001***	0.186***	3.66***	0.0003***
Current Suicidal Thoughts***	1.29***	3.58***	0.0004***	0.31*	2.21*	0.029*
Hx Suicide Attempt*	1.42*	2.40*	0.017 *	0.18*	1.48*	0.14*
Aggressive Behavior	1.02	0.483	0.629	0.02	0.43	0.67
Academic Difficulty	-0.91	-1.153	0.25	-0.09	-1.27	0.20
Hyperactivity	-0.98	-0.547	0.58	-0.07	-1.03	0.30
Memory Problems	1.07	0.723	0.471	0.07	0.95	0.35
Attention Problems*	1.22*	1.849*	0.06*	0.17*	2.25*	0.026*
Voices	1.38	1.22	0.22	0.296	1.18	0.24
Income	-0.99	-0.88	0.378	-0.008	-0.74	0.46
Group Effects:	SD	Variance	Correlation	SD	Variance	Correlation
Supervisor	1	1	0.03	1	1	0.07
Therapist	1	1	0.03	1	1	0.07
Significance codes: *** = < 0.001; ** = < 0.01; * = < 0.05						

Table 24. *Comparing the Weeks & Sessions GEL Model with Interactions*

	Weeks	**Sessions**
Interactions Included in Best Fitting Model	Depression; Suicidal Thoughts; Hopelessness; Previous Suicide Attempt; Attention Problems	Depression; Suicidal Thoughts; Hopelessness
AIC	183.6	196.1
BIC	266.7	-266.4
R^2	0.962	-0.987

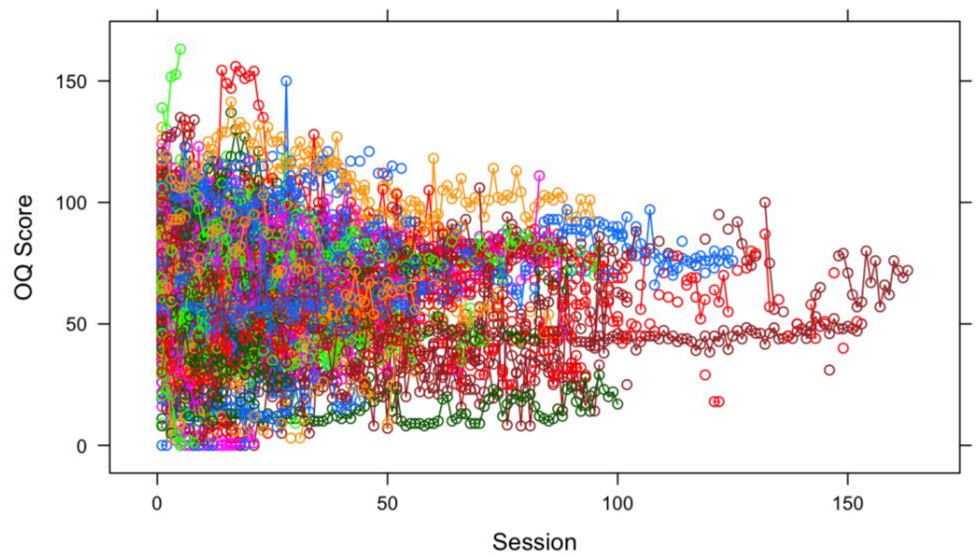
Table 25. *Correlation Matrix with Interaction Terms*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. B_{00} Weeks	1										
2. B_{00} Session	**-**	1									
3. Depression	**0.65**	**-0.62**	1								
4. SI	-0.11	-0.08	-0.23	1							
5. Hopeless	**0.64**	**-0.61**	**0.75**	-0.23	1						
6. Attn Probs	-0.3	-0.04	-0.20	0.03	-0.24	1					
7. Suic Atmpt	-0.4	-0.05	-0.02	-0.32	-0.02	0.01	1				
8. Week	-0.09	0.01	0.007	-0.002	0.008	-0.001	0.002	1			
9. #Weeks	***0.42***	-0.17	-0.10	0.11	-0.09	0.01	0.004	0.06	1		
10. Session	-0.02	-0.10	0.01	0.001	0.01	-0.001	0.01	-0.28	0.01	1	
11. #Sessions	-0.03	***-0.40***	-0.12	0.12	-0.11	0.03	0.03	-0.04	**0.90**	0.07	1

Magnitude codes: **Large,** ***Moderate,*** Small/Negligible correlation

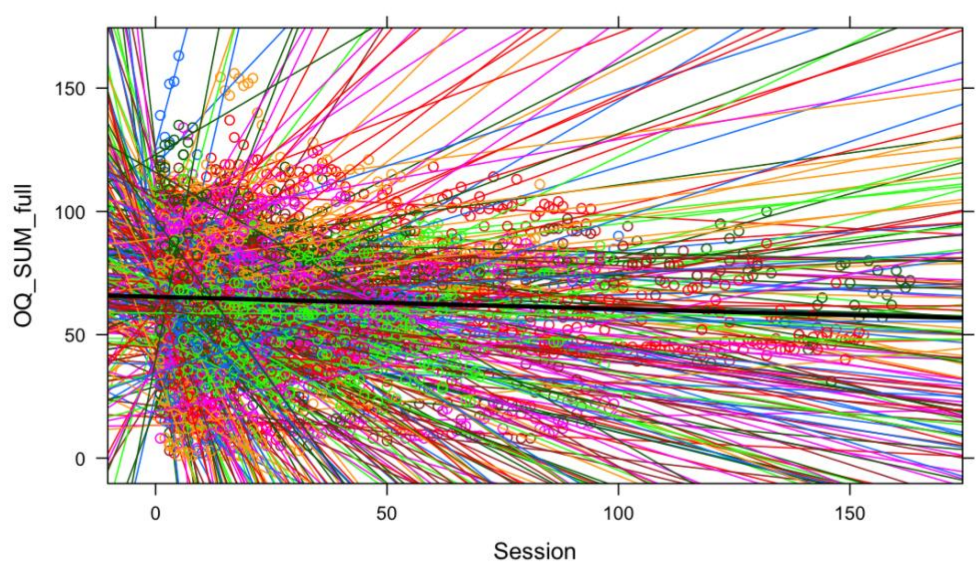
*Abbreviations: B_{00}= Intercept; *SI* = Suicidal Ideation; *Attn Probs* = Attention Problems, *Suic Atmpt* = History of Suicide Attempt; *#Weeks* = Total Weeks; *#Sessions* = Total Sessions.

Figure 1. *All Client* OQ Scores by Session (n = 311)*



*Different colors represent different participant scores

Figure 2. *All Client* OQ Score by Session with Individual Slopes and Line of Best Fit*



*Different colors represent different participant linear slopes

Figure 3. *All OQ Scores by Session with Loess Regression Line*

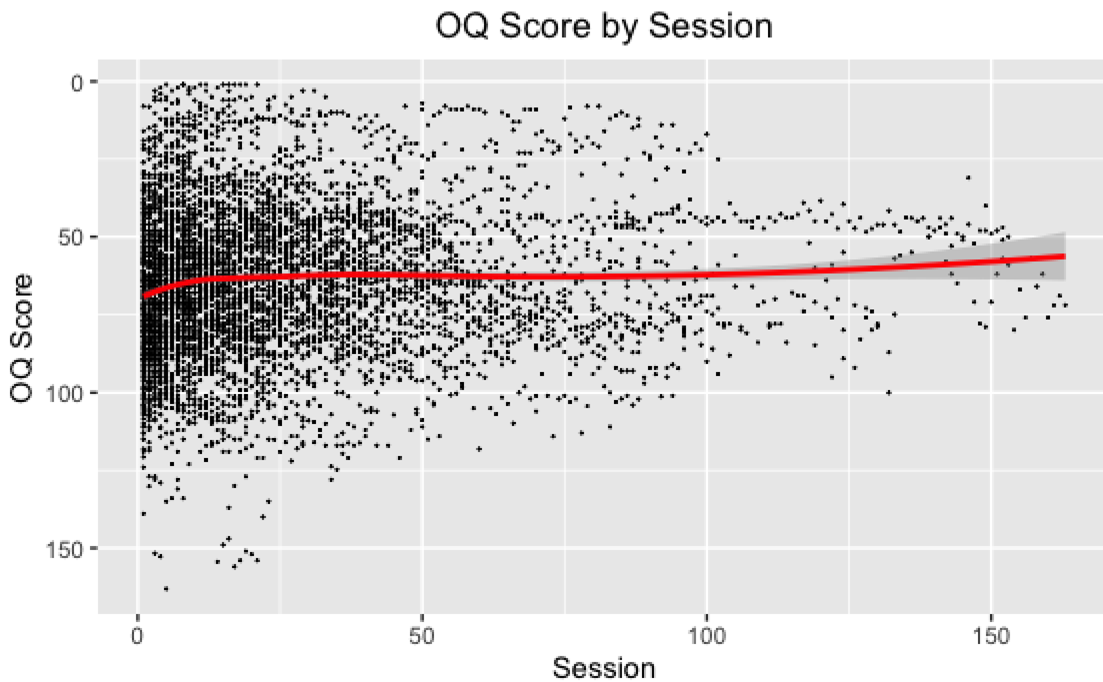


Figure 4. *All OQ Scores by Weeks with Loess Regression Line*

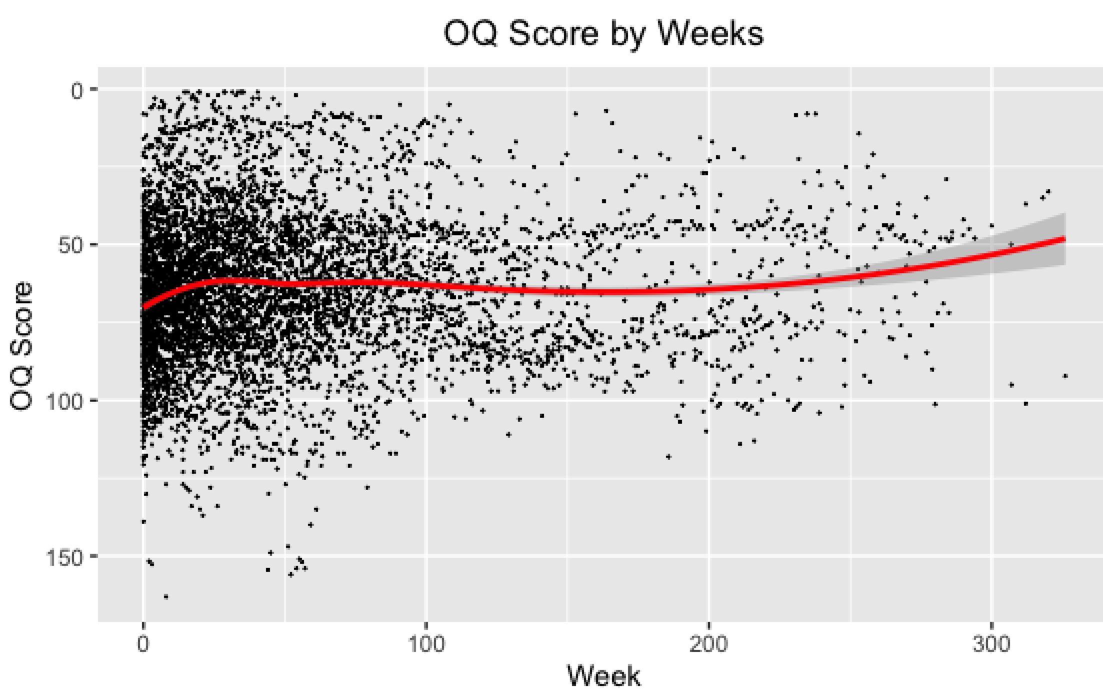


Figure 5. *All OQ Scores by Week & Session with Regression Lines Comparison*

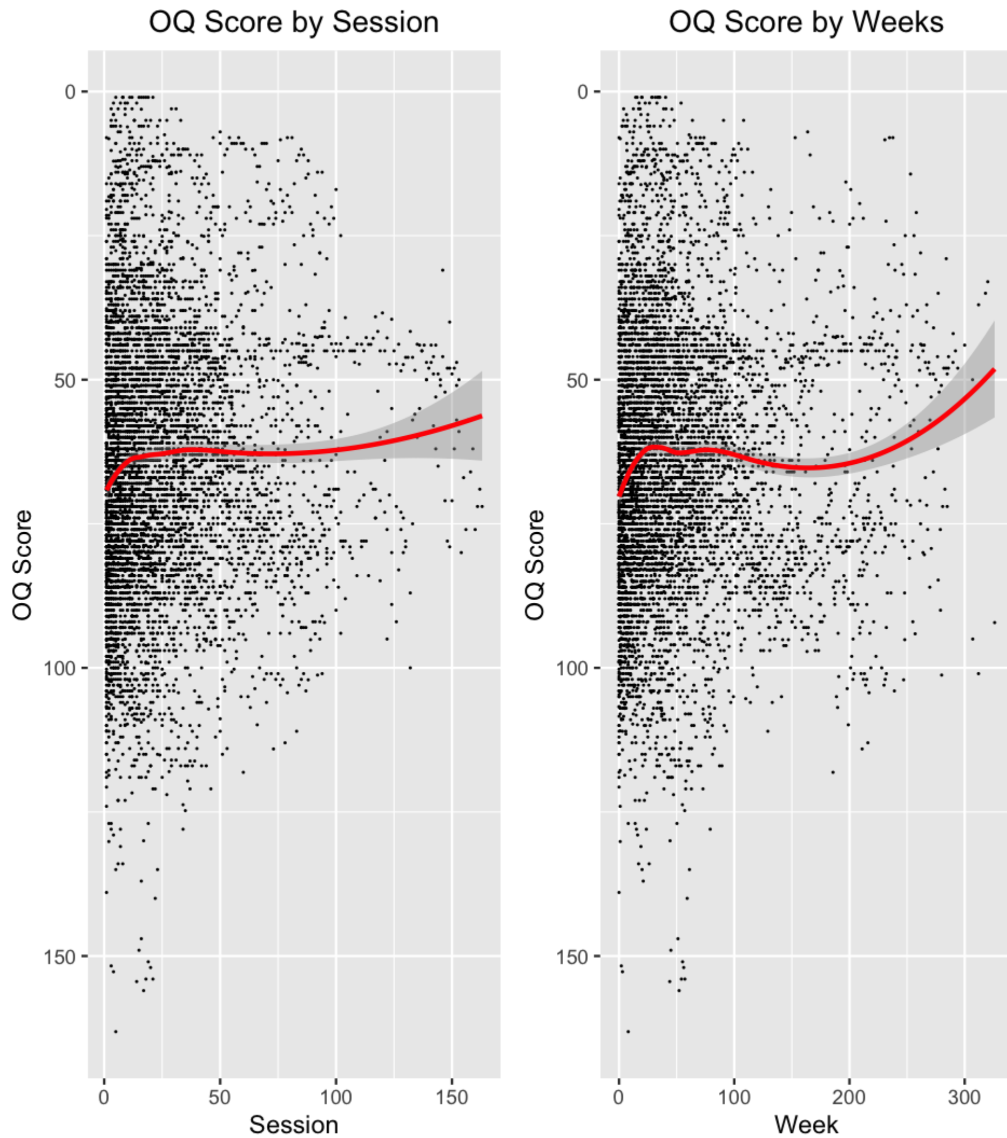


Figure 6. *Predicted OQ Scores by Weeks and Sessions Comparison*

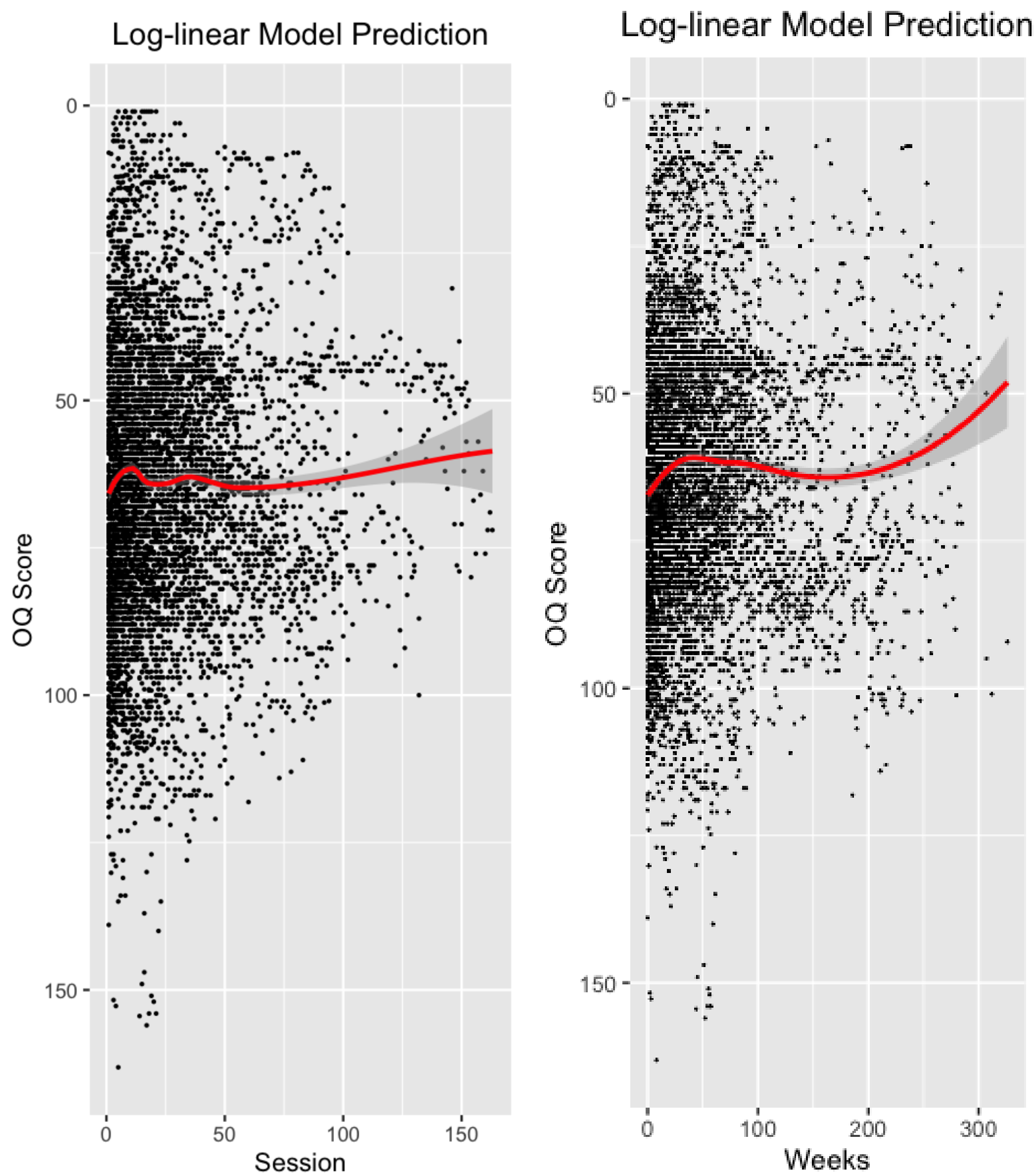


Figure 7. *OQ Scores for Participants Above & Below Median (Mdn) Total Weeks (Mdn = 40)*

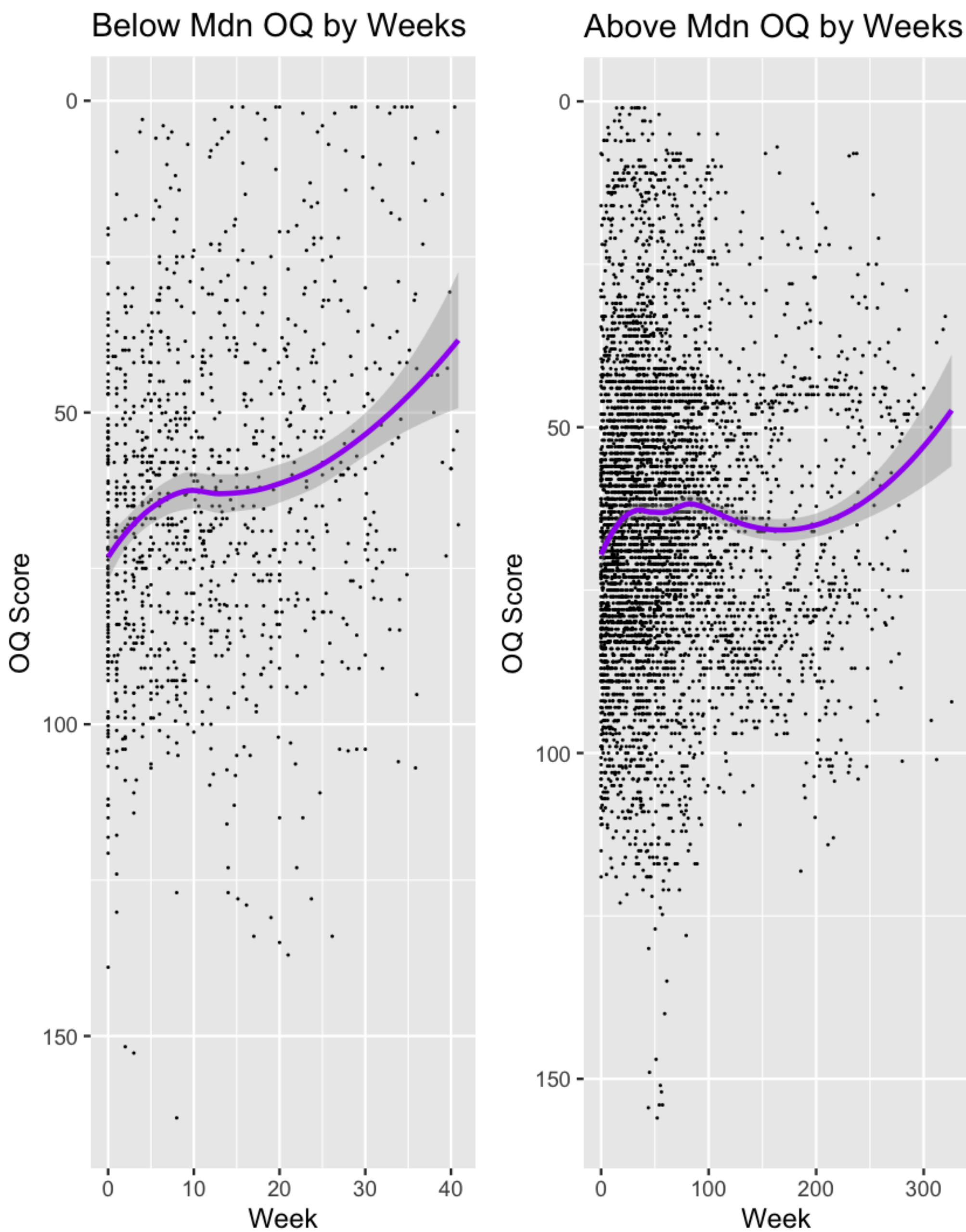


Figure 8. *OQ Score for Participants Above & Below Mdn Total Sessions (Mdn = 40)*

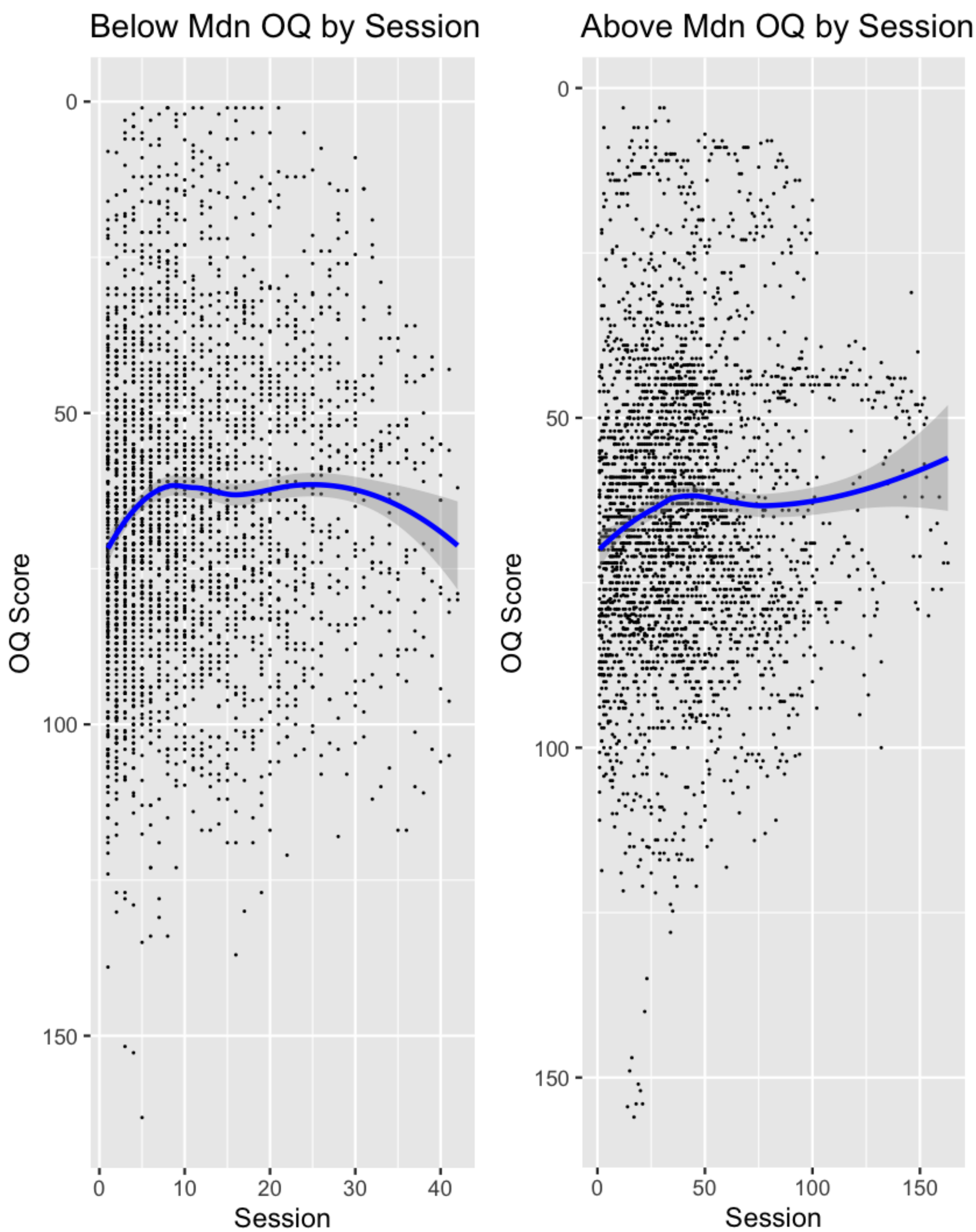


Figure 9. *OQ Score for Participants with Fast (MFast = -2.82), Moderate (MMod = -0.61) and Slow (MSlow = 0.09) Change Rates per Session*

Figure 10. *OQ Scores for Participants with Fast (MFast = -1.28), Moderate (MMod= -0.27), and Slow (MSlow = 0.04) Change Rate per Week*

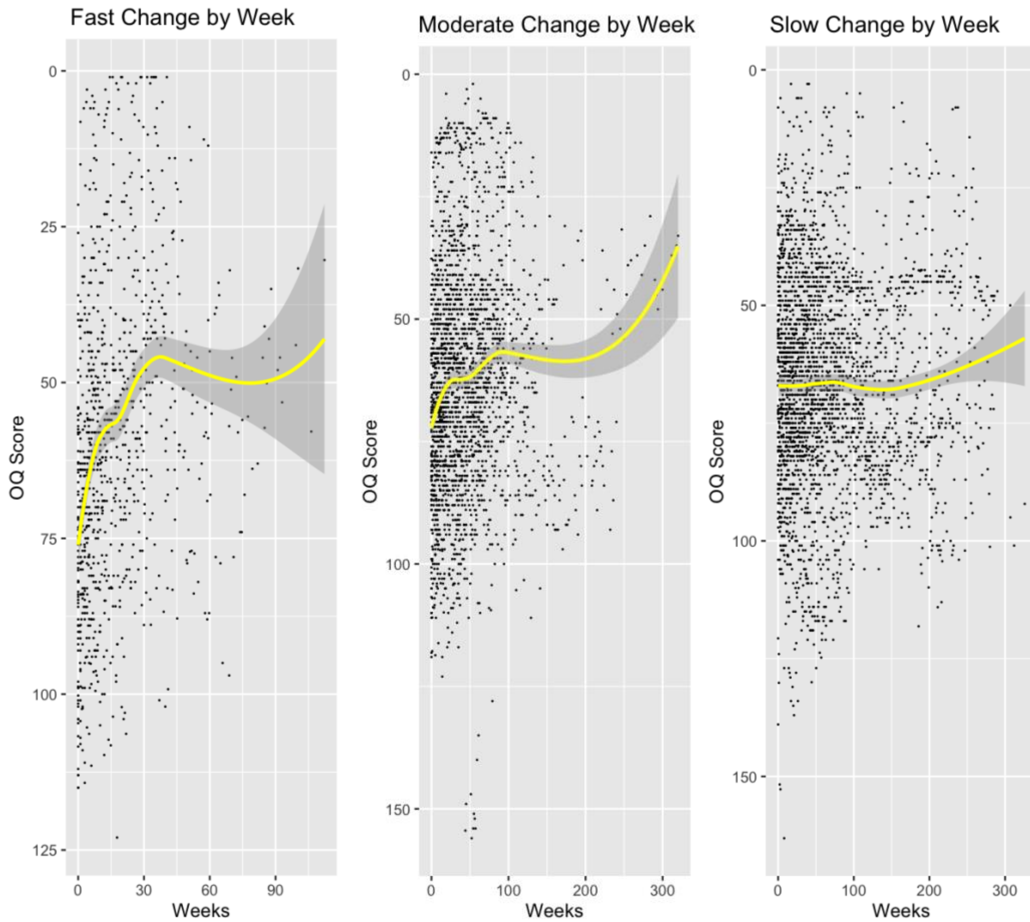


Figure 11. *Average OQ Score per Session for All Clients*

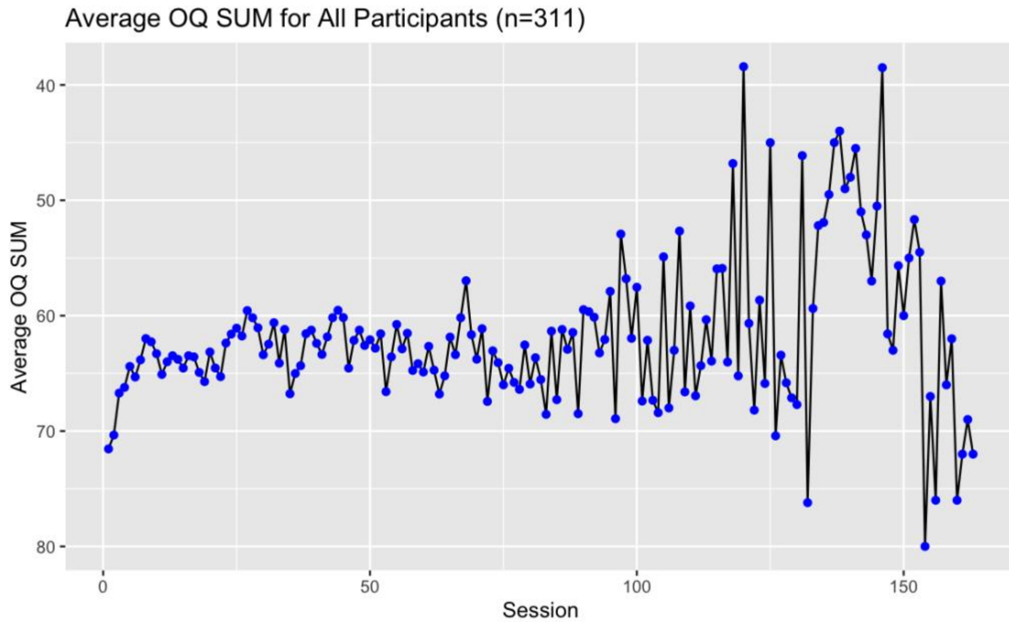


Figure 12. *Average OQ Score per Session with Regression for All Clients*

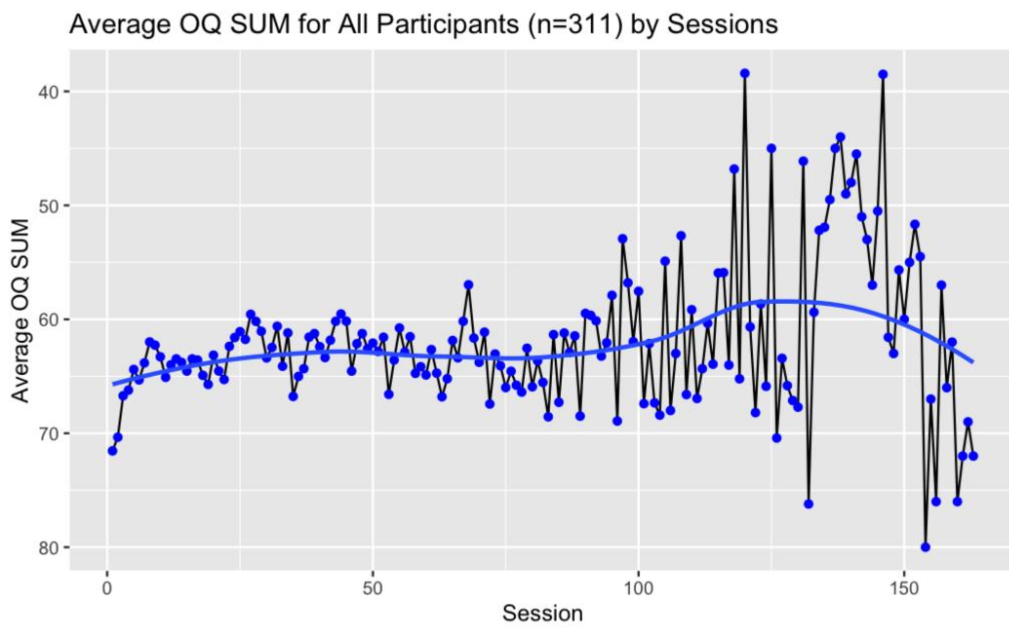


Figure 13. *Average OQ Score per Week Across All Clients*

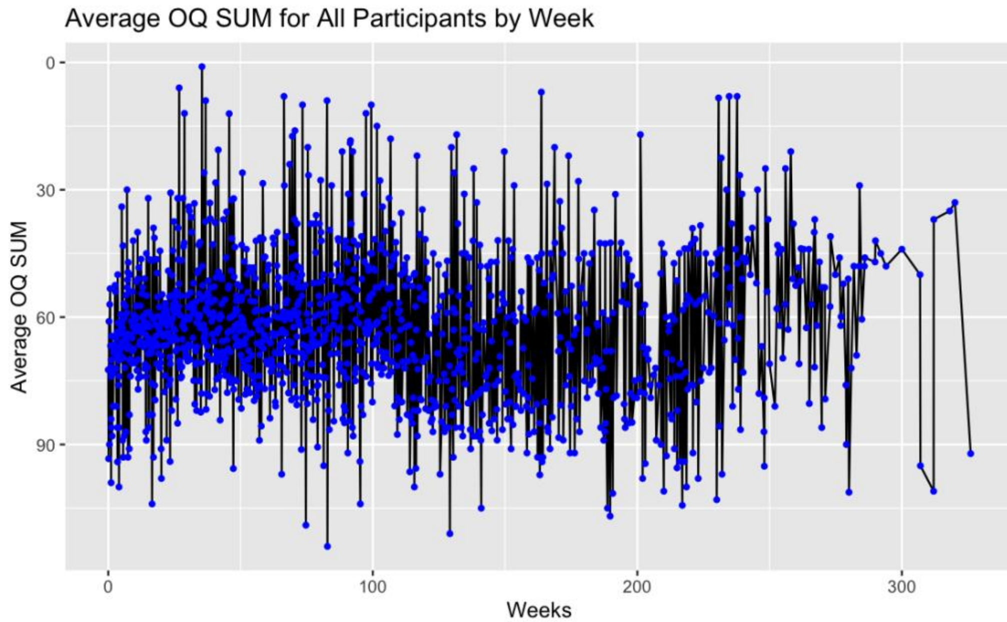


Figure 14. *Average OQ Score per Week across All Clients with Regression Line*

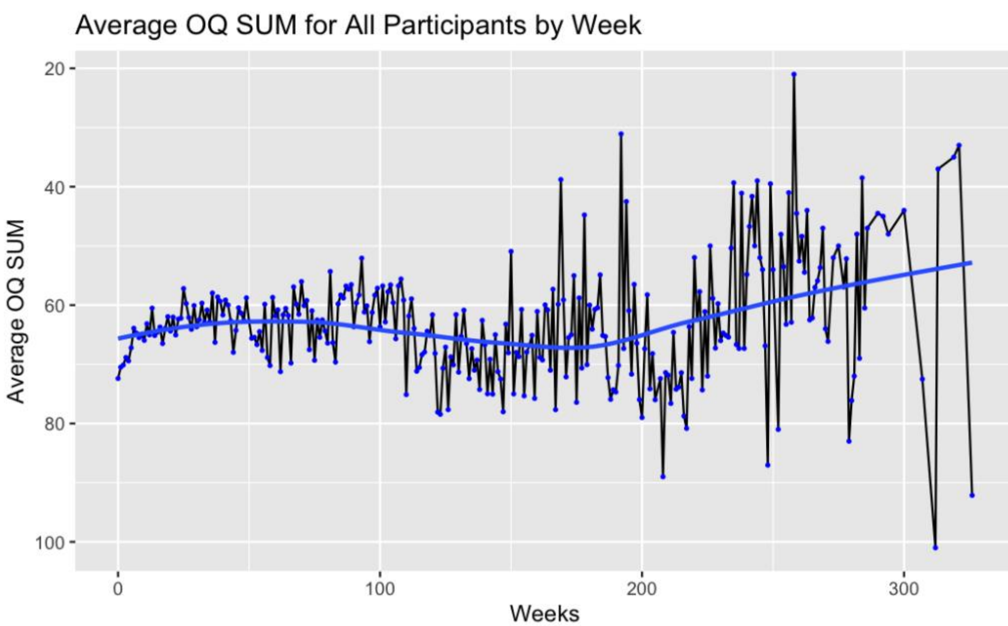


Figure 15. *Average OQ Score for Participants with Below Mdn Total Sessions (n > 17)*

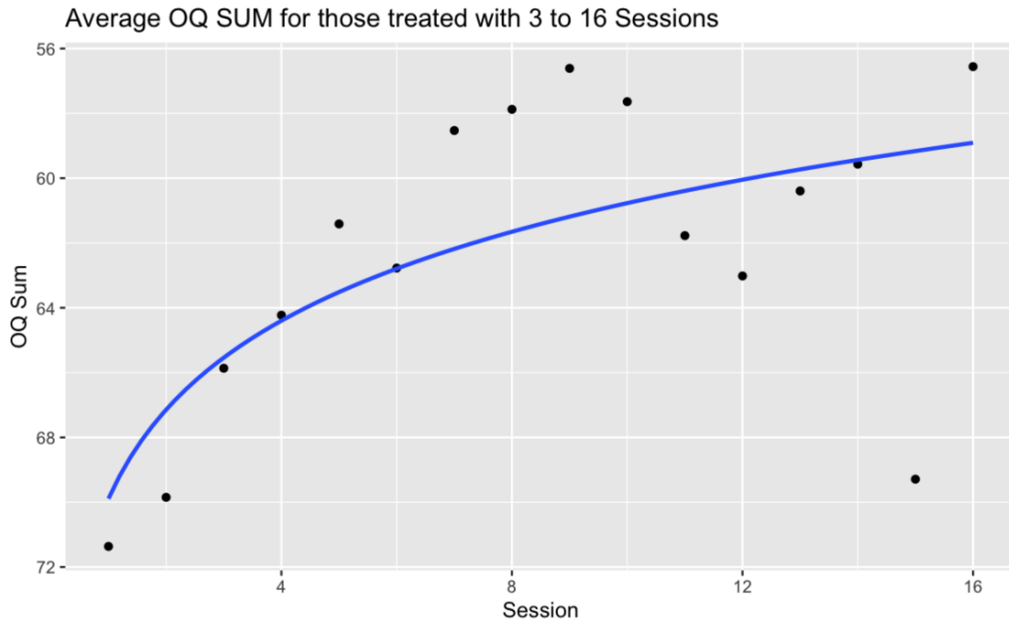


Figure 16. *Average OQ Score for Participants with Above Mdn Total Sessions (n < 17)*

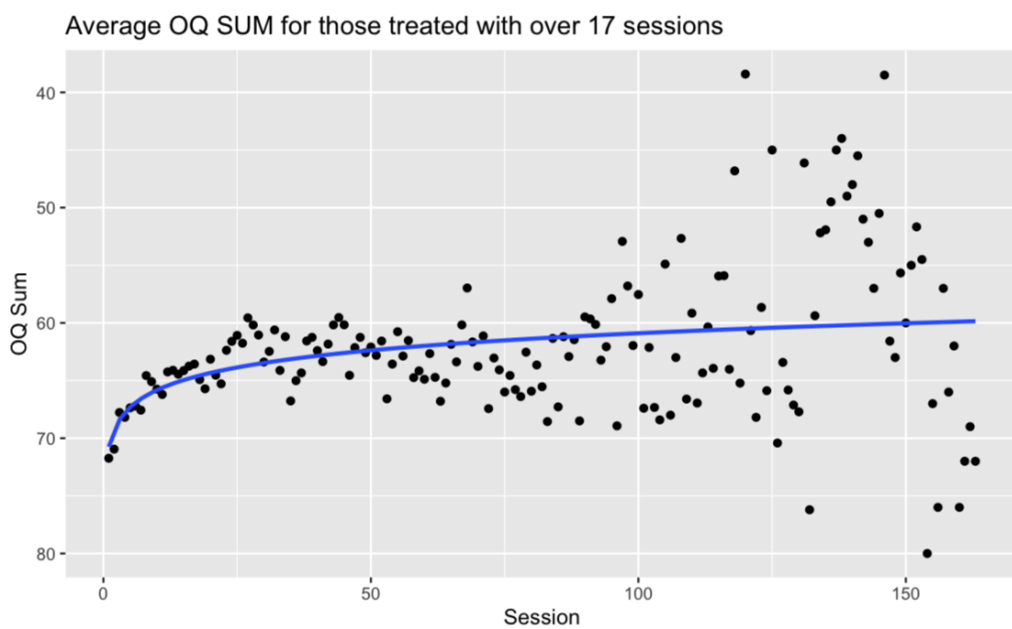


Figure 17. *Average OQ Score by Session with Linear Trend Line*

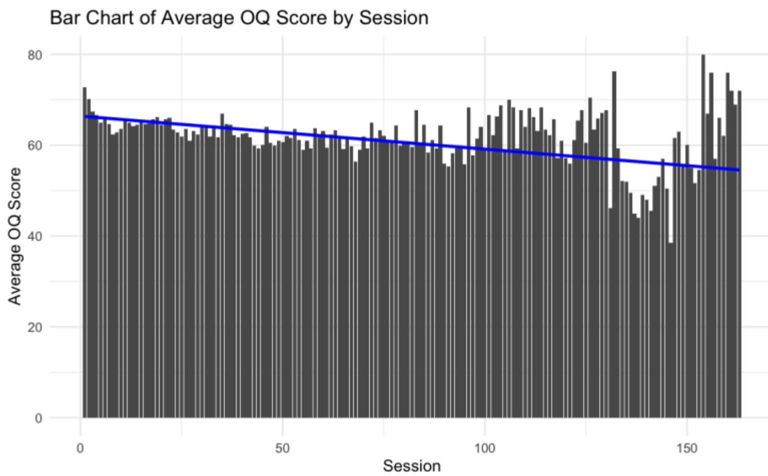


Figure 18. *Average OQ Score by Session with Quadratic Trend Line*

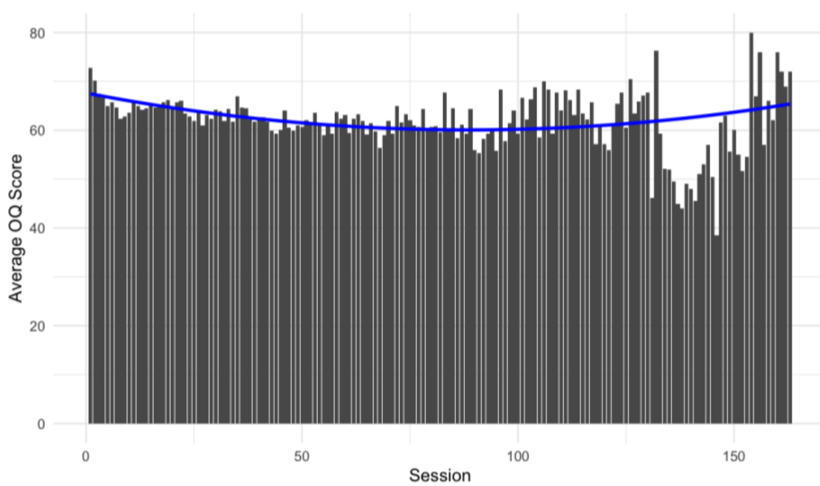


Figure 19. *Average OQ Score by Session with Cubic Trend Line*

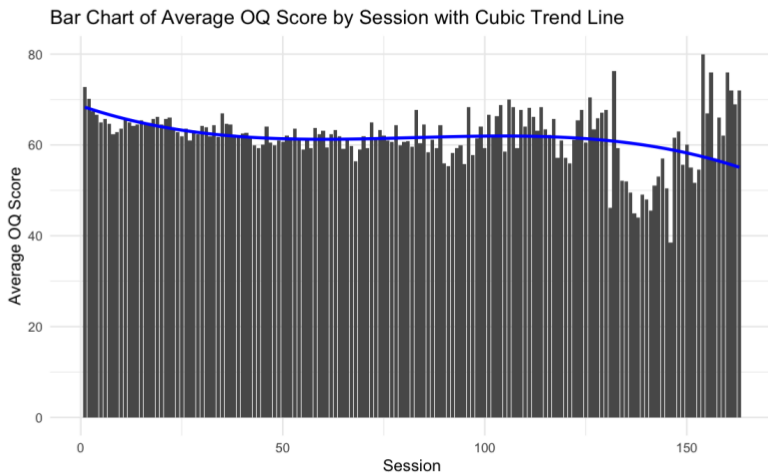


Figure 20. *Average OQ Score by Session with Log-Linear Trend Line*

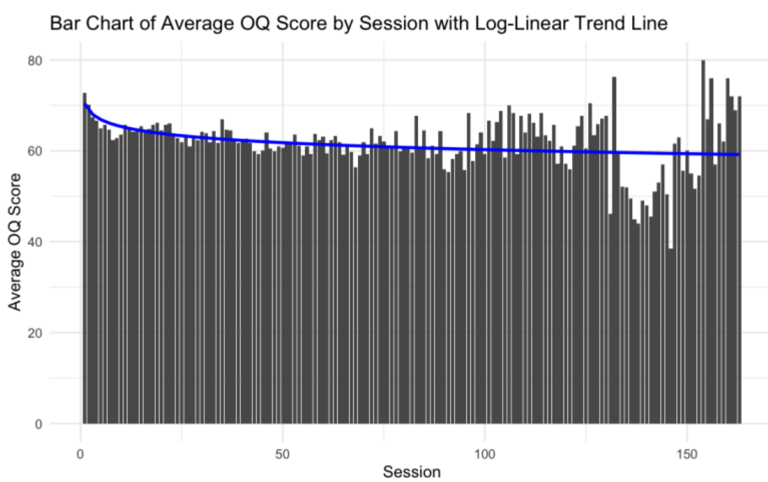


Figure 21. *Accumulation of Participants with Reliably Significant Improvement (RSI) by Session (n=203)*

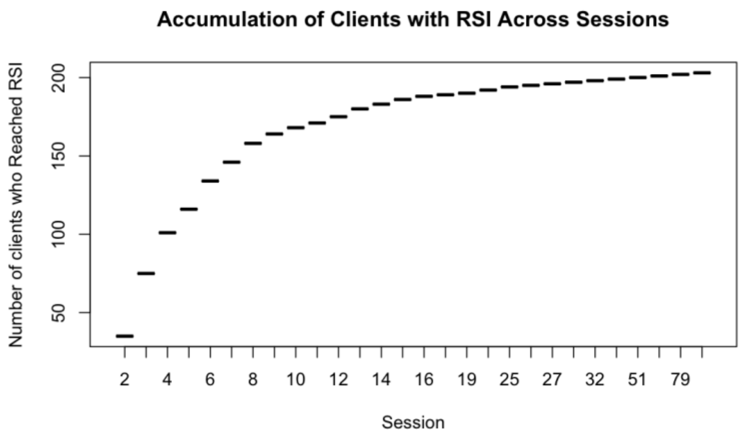


Figure 22. *Accumulation of Participants with RSI & Below Mdn Total Sessions (n = 96)*

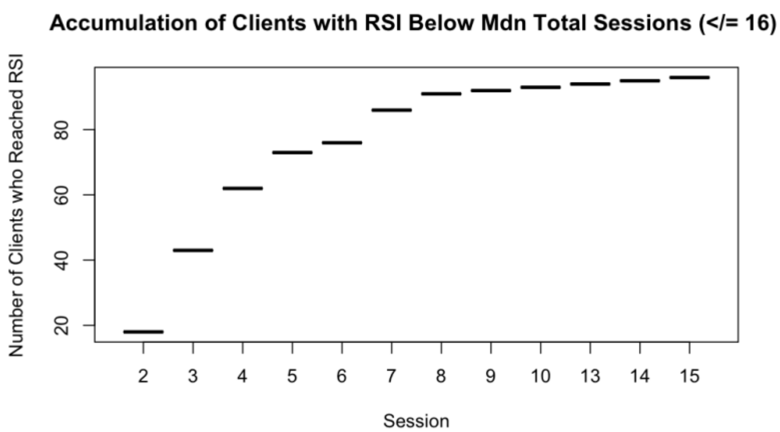


Figure 23. *Accumulation of Participants with RSI & Above Mdn Total Sessions (n = 107)*

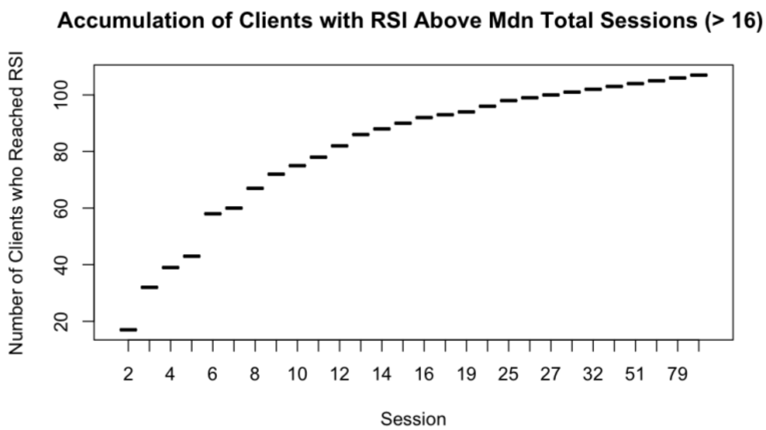


Figure 24. *Accumulation of All Participants with RSI by Total Weeks (n = 203)*

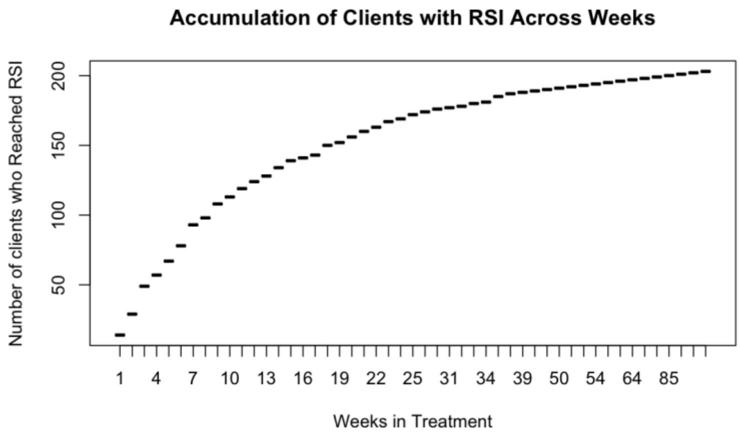


Figure 25. *Accumulation of Participants with RSI & Below Mdn Total Weeks (n = 86)*

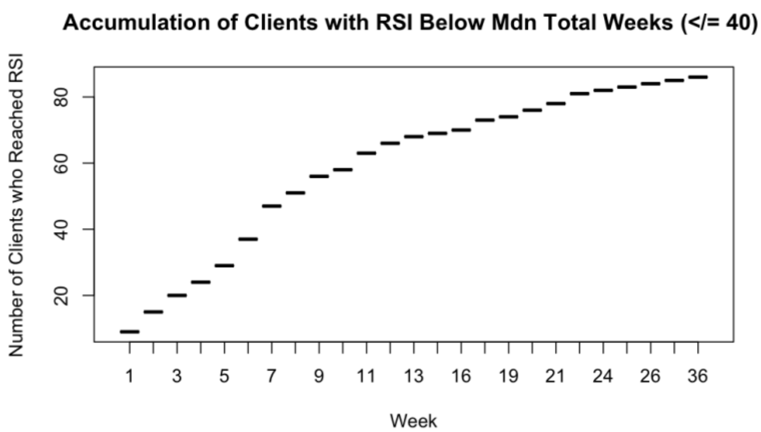


Figure 26. *Accumulation of Participants with RSI & Above Mdn Total Weeks (n = 117)*

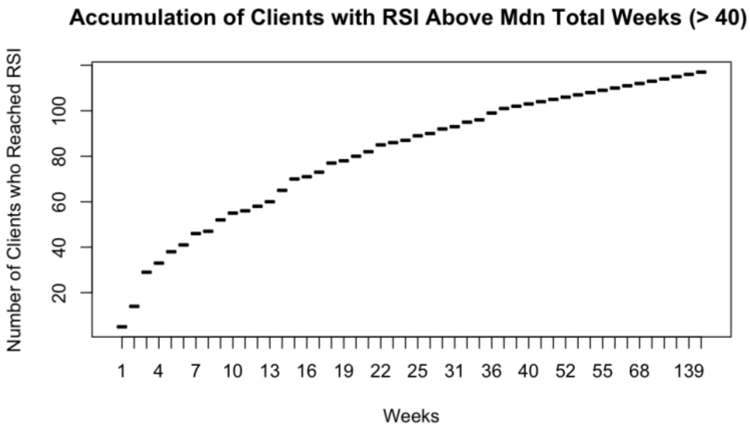


Figure 27. *Accumulation of Clients with Reliably Significant Deterioration (RSD) at Discharge by Total Sessions (n=13)*

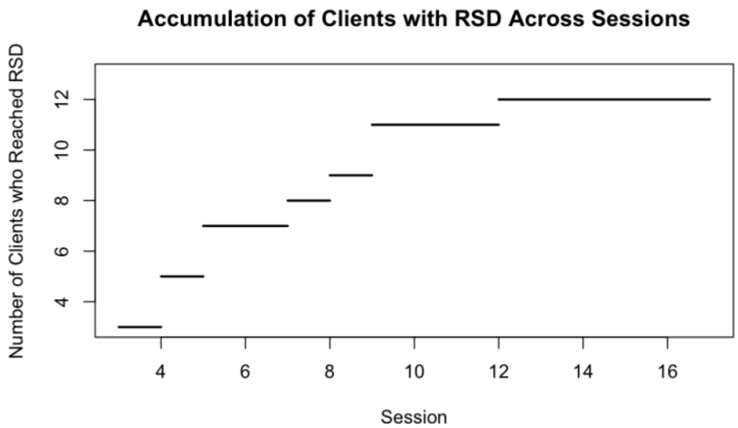
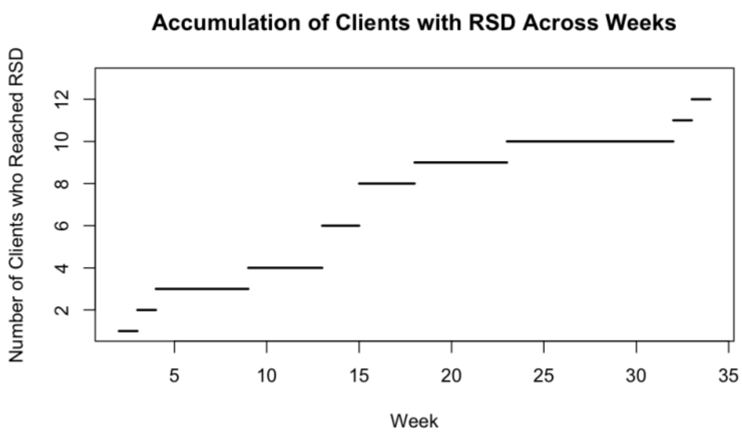


Figure 28. *Accumulation of Clients with RSD at Discharge by Total Weeks (n = 13)*



REFERENCES

Anderson, E. M., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology,* 57(7), 875-888.

Barkaoui, K. (2014). Quantitative approaches for analyzing longitudinal data in second language research. *Annual Review of Applied Linguistics,* 34, 65-101. doi:10.1017/S0267190514000105

Barkham, M., Connell, J., Stiles, W. B., Miles, J. N., Margison, F., Evans, C., & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting and Clinical Psychology*, *74*(1), 160.

Bone, C., Delgadillo, J., and Barkham, M. (2021) A systematic review and meta-analysis of the good-enough level (GEL) literature. *Journal of Counseling Psychology,* 68 (2). pp. 219-231. ISSN 0022-0167 https://doi.org/10.1037/cou0000521

Bruijniks, S., Lemmens, L., Hollon, S., Peeters, F., Cuijpers, P., Arntz, A., . . . Huibers, M. (2020). The Effects of Once- versus Twice-Weekly Sessions on Psychotherapy Outcomes in Depressed Patients. *The British Journal of Psychiatry, 216*(4), 222-230. doi:10.1192/bjp.2019.265

Chase, R. M., Whitton, S. W., & Pincus, D. B. (2012). Treatment of Adolescent Panic Disorder: A Nonrandomized Comparison of Intensive versus Weekly CBT. *Child & Family Behavior Therapy*, *34*(4), 305–305.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. Structural Equation Modeling: *A Multidisciplinary Journal*, 5(3), 247-266.

Cuijpers P, Huibers M, Daniel Ebert D, Koole SL, Andersson G. (2013). How Much Psychotherapy is Needed to Treat Depression? A Meta-regression Analysis. *Journal of Affective Disorders.* 2013;149(1–3): 1–13.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, 11(2), 121-13 6, DOI: 10.1080/15248371003699969

Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The session rating scale: preliminary psychometric properties of a "working" alliance measure. Journal of Brief Therapy, 3(1), 3-12.

Erekson, D. M., Lambert, M. J., & Eggett, D. L. (2015). The relationship between session frequency and psychotherapy outcome in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, 83(6), 1097.

Erekson, D. M., Bailey, R. J., Cattani, K., Klundt, J. S., Lynn, A. M., Jensen, D., ... & Worthen, V. (2022). Psychotherapy session frequency: A naturalistic examination in a university counseling center. *Journal of Counseling Psychology*, 69(4), 531.

Falkenström, F., Josefsson, A., Berggren, T., & Holmqvist, R. (2016). How much therapy is enough? Comparing dose-effect and good-enough models in two different settings. *Psychotherapy,* 53(1), 130–139. https://doi.org/10.1037/pst0000039

Farmer, R. L., Goforth, A. N., Kim, S. Y., Naser, S. C., Lockwood, A. B., & Affrunti, N. W. (2021). Status of school psychology in 2020: Part 2, professional practices in the NASP membership survey. *NASP Research Reports,* 5(3), 1-17.

Gagnon, Douglas J. and Mattingly, Marybeth J., "Most USS School Districts Have Low Access to School Counselors: Poor, Diverse, and City School Districts Exhibit Particularly High Student-to-Counselor Ratios" (2016). The Carsey School of Public Policy at the Scholars' Repository. 286. https://scholars.unh.edu/carsey/286

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the BigFive personality domains. Journal of Research in Personality, 37(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Gotaas, M. E., Stiles, T. C., Bjørngaard, J. H., Borchgrevink, P. C., & Fors, E. A. (2021). Cognitive Behavioral Therapy Improves Physical Function and Fatigue in Mild and Moderate Chronic Fatigue Syndrome: A Consecutive Randomized Controlled Trial of Standard and Short Interventions. *Frontiers in Psychiatry*, *12*, 580924–580924. https://doi.org/10.3389/fpsyt.2021.580924

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist*, *41*(2), 159.

Jané-Llopis, E. (2006). Mental health promotion and mental disorder prevention in Europe. *Journal of Public Mental Health,* Vol. 5 No. 1, pp. 5-7. https://doi.org/10.1108/17465729200600002

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. Journal of the American Statistical Association, 90(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kivlighan III, D. M., Jung, A. K., Berkowitz, A. B., Hammer, J. H., & Collins, N. M. (2019). 'To show or No-show'? Therapist Racial-ethnic Disparities in Clients' Nonattendance in Therapy. *Journal of College Student Psychotherapy*, 33(1), 1-13.

Kristjansson, S. D., Kircher, J. C., & Webb, A. K. (2007). Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. Psychophysiology, 44(5), 728-736. https://doi.org/10.1111/j.1469-8986.2007.00544.x

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. Clinical Psychology and Psychotherapy, 3, 249–258. https://doi.org/10.1002/ (SICI)1099-0879

Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. Psychotherapy, 52(4), 381.

Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., ... Reisinger, C. (1996). Administration and scoring manual for the OQ-45.2. Stevenson, MD: *American Professional Credentialing Services.*

Lee, A. A., Sripada, R. K., Hale, A. C., Ganoczy, D., Trivedi, R. B., Arnow, B., & Pfeiffer, P. N. (2021). Psychotherapy and depressive symptom trajectories among VA patients: Comparing dose-effect and good-enough level models. *Journal of Consulting and Clinical Psychology*, 89(5), 379–392. https://doi.org/10.1037/ccp0000645

Levy, H. C., Worden, B. L., Davies, C. D., Stevens, K., Katz, B. W., Mammo, L., ... & Tolin, D. F. (2020). The dose-response curve in cognitive-behavioral therapy for anxiety disorders. *Cognitive Behaviour Therapy,* 49(6), 439-454. DOI: 10.1080/16506073.2020.1771413

Lincoln, T. M., Jung, E., Wiesjahn, M., & Schlier, B. (2016). What is the Minimal Dose of Cognitive Behavior Therapy for Psychosis? An Approximation Using Repeated Assessments over 45 Sessions. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *38*, 31–39. https://doi.org/10.1016/j.eurpsy.2016.05.004

Maag, J. W., & Katsiyannis, A. (2010). School-based mental health services: Funding options and issues. *Journal of Disability Policy Studies,* 21(3), 173-180.

McCoach, D. B., & Adelson, J. L. (2010). Dealing with dependence (Part I): Understanding the effects of cluster data. *Gifted Child Quarterly,* 54, 152–155. doi:10.1177/0016986210363076

Niileksela, C. R., Ghosh, A., & Janis, R. A. (2021). The dose–effect and good enough level models of change for specific psychological concerns. *Journal of Consulting and Clinical Psychology*, 89(3), 200–213. https://doi.org/10.1037/ccp0000635

Okun, M., & Glidewell, R. (2020). Efficacy of a Single 4-hour CBT-I Workshop in a Community Sample with Self-Reported Insomnia Symptoms. *Sleep*, *43*, 203. https://doi.org/10.1093/sleep/zsaa056.528

Owen, J. J., Adelson, J., Budge, S., Kopta, S. M., & Reese, R. J. (2016). Good-enough level and dose-effect models: Variation among outcomes and therapists. *Psychotherapy Research*, *26*(1), 22-30.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology,* 25, 111–195. doi:10.2307/271063

Reese, R. J., Toland, M. D., & Hopkins, N. B. (2011). Replicating and extending the good-enough level model of change: Considering session frequency. *Psychotherapy Research,* 21(5), 608-619.

Robinson, L., Delgadillo, J. and Kellett, S. (2019) The Dose-Response Effect in Routinely Delivered Psychological Therapies: A Systematic Review. *Psychotherapy Research.* ISSN 1050-3307 https://doi.org/10.1080/10503307.2019.1566676

Schuler, K., Kilmer, E. D., Callahan, J., Dziurzynski, K., & Swift, J. (2021). The dose-effect model is good enough. *Counseling and Psychotherapy Research,* 21(3), 608-620.

Shephard, R. W., & Färe, R. (1974). The law of diminishing returns. In *Production theory* (pp. 287-318). Springer, Berlin, Heidelberg.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A (statistics in society),* 171(1), 5-19. DOI: 10.1111/j.1467-985X.2007.00509.x

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., ... & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *Journal of Consulting and Clinical Psychology*, 71(1), 14.

Stulz, N., Lutz, W., Kopta, S. M., Minami, T., & Saunders, S. M. (2013). Dose–effect relationship in routine outpatient psychotherapy: Does treatment duration matter? *Journal of Counseling Psychology,* 60(4), 593.

Wolgast, B. M., Rader, J., Roche, D., Thompson, C. P., von Zuben, F. C., & Goldberg, A. (2005). Investigation of clinically significant change by severity level in college counseling center clients. *Journal of College Counseling,* 8(2), 140-152.

Xiao, H., Hayes, J. A., Castonguay, L. G., McAleavey, A. A., & Locke, B. D. (2017). Therapist effects and the impacts of therapy nonattendance. *Psychotherapy,* 54(1), 58.

Vita

Name	*Ashley Rottkamp*
Baccalaureate Degrees	*Bachelor of Arts*, *Binghamton University, Vestal, Major: Psychology* *Bachelor of Science*, *Binghamton University, Vestal, Major: Accounting*
Date Graduated	*May, 2019*
Other Degrees and Certificates	*Master of Science, St. John's University, Jamaica, Major: School Psychology*
Date Graduated	*May, 2022*