UNIVERSITY OF TWENTE.

**Faculty of Behavioral, Management and Social Sciences**

# Predictive ability of well-being for class membership of latent classes of eating disorder pathology change

Erik Schmilgus

M.Sc. Thesis

1st Supervisor: Prof. Dr. Gerben J. Westerhof

2nd Supervisor: Dr. Erik Taal

External Supervisor: Dr. Jan A. de Vos

# Table of Contents

## Abstract

**Background.** Eating disorders (EDs) have a high prevalence globally plus ED treatment requires tailoring on a more individual level due to high variability in adaption, treatment course and outcomes. The construction of latent classes, subgroups with similar trajectories of symptom change, makes the variability visible on a more individual group level. They enable the observation of trajectories of change during the treatment and the tailoring of treatment. Predictors of membership of latent classes of change facilitates the tailoring process and prevention of expected negative change.

**Methods.** This study included 1283 participants and used a naturalistic longitudinal design. Five measurements of eating disorder pathology (EDP) and well-being (WB) were measured every three months over a duration of one year. Latent growth curve modelling was applied to identify latent classes of change, multinominal regression was used to test WB as a predictor for class membership.

**Results.** Three latent classes of change were identified: first class *high symptom baseline, slow recovery* (10.54%), the second-class *low symptom baseline, fast recovery* (54.44%) and the third-class *moderate symptom baseline, moderate recovery* (35.02%). WB was able to predict the membership to the *high symptom baseline, slow recovery* class and to the *low symptom baseline, fast recovery* class, but not the *moderate symptom baseline, moderate recovery* class.

**Conclusion.** This study validated three class model for EDP change trajectories although the quality of the classes differentiates from previous studies. The quality of the classes seems to differ from study to study which must be respected in clinical practice. WB partly predicts membership to latent classes of change, but it is not clear if it predicts EDP level or change trajectory and therefore needs further research and validation.


*Keywords*: Eating disorder pathology, well-being, latent classes, change trajectory, predictor

# Introduction

Eating disorders (EDs) are a global burden due to their prevalence around the world (Streatfeild et al., 2021). Over 3.3 million healthy life years worldwide are lost to EDs (Owens, Attia, Fitzpatrick, Phillips & Nolan, 2021). Additionally, treatment courses and outcomes are highly variable from patient to patient (Espel-Huynh et al., 2020) which is calling for more individual treatment approaches.

EDs are a group of syndromes characterized by abnormal eating behaviors and cognitions which are accompanied by overvaluation of body shape or image (Chang, Delgadillo & Waller, 2021; Qian et al., 2021). These changes significantly impair quality of life, social functioning and are commonly comorbid with mental disorders like depression, anxiety disorders or personality pathology (Jordan et al., 2008; Qian et al., 2021). Eighty to 85% of the individuals suffering from EDs are not underweight and regardless of weight, EDs bear increased health risks concerning cardiac problems or bone deterioration (Treasure, Claudino & Zucker, 2010). According to the Diagnostic and statistical Manual of Mental Disorders, 5$^{th}$ edition (DSM-V), EDs include anorexia nervosa (AN), bulimia nervosa (BN), eating disorder not otherwise specified (EDNOS) and the most recent additions binge eating disorder (BED) and avoidant/restrictive food intake disorder (ARFID) which were added in the last ten years (Hay, 2020; Qian et al., 2021). The EDs AN and BN are characterized by extreme weight control behaviors due to body image concerns. AN is a self-starving condition where individuals try to prevent weight gain to preserve underweight whereas people suffering from BN show a cycle of regular binge eating episodes and purging in order to compensate for the food intake and are not necessarily underweight (Hay, 2020; Qian et al., 2021). BED and ARFID are disorders not characterized by body image concerns. BED is defined by regular binge eating episodes without compensation and ARFID by selective food restriction and underweight but is not affected by distress about body image (Hay, 2020; Qian et al., 2021; National Eating Disorders Association, 2022). Further EDs are summarized under EDNOS.

There are different treatment options how EDs can be treated. The treatment can be delivered in an in-patient setting, in a daily hospital or out-patient setting depending on psychological, physical and social factors (Meads, Burls, Gold & Jobanputra, 1999). Usually, treatment for EDs is delivered in a treatment continuum which starts with outpatient treatment, followed by intensive outpatient treatment, daily hospital or partial residential and finally inpatient hospitalization (Peckmezian & Paxton, 2020). The treatment options are primarily psychological interventions, individual, group or family interventions, optionally a combination of psychological and pharmacological interventions (UK N. G. A., 2017). Examples for psychological interventions would be interpersonal therapy (IPT), family-based therapy (FBT) and cognitive behavioral therapy (CBT) although CBT is considered as the most established treatment for EDs (Linardon, Brennan, & de la Piedad Garcia, 2016; Linardon, Wade, De la Piedad Garcia & Brennan, 2017). Even an "enhanced" version of CBT (CBT-E) was developed which focuses especially on the treatment of EDs. The core concept of CBT-E is addressing the eating disorder pathology and mood intolerance which displays great efficacy (Atwood & Friedman, 2019). However, recommended treatment depends on the patient's type of disorder and age (Peckmezian & Paxton, 2020).

The treatment of EDs also bears limitations. Remission rates for EDs are low and dependent on the type of ED, but the rates may increase over time (Keel & Brown, 2010; Milos, Spindler, Schnyder & Fairburn, 2005). Thereby remission is defined as an abstinence of ED symptoms e.g., for BN an abstinence of 2 to 8 weeks (Eating Disorders Review, 2005). A further limitation is the high relapse rates ranging from 20 to 50% (Sperrazza, 2022), especially high for AN (Berends, Boonstra & Van Elburg, 2018; Steinglass et al., 2022). Furthermore, there is no consensus on the definition of the recovery for EDs (de Vos et al., 2017; Noordenbos, 2011). However, it is clear that recovery needs more than only a reduction of symptoms but also an improvement in qualities like positive relationships with others, self-acceptance, positive body evaluation and self-esteem (de Vos et al., 2017; Noordenbos, 2011). Since the definition of recovery is not clear, it is lacking an international standard on how to measure recovery of EDs (de Vos et al., 2017; Noordenbos, 2011) which can be considered as a limitation for disordered treatment evaluation. A further limitation is the overall rate of premature drop-out of 24% (Linardon et al., 2018) which is a problem since it produces not optimal treatment outcomes (Cooper & Conklin, 2015; Swift & Greenberg, 2014). Additionally, considerable part of the ED patients does not profit from treatment (Linardon et al., 2017) and patients react differently to treatment regarding change of symptoms (Melchior et al., 2016). That means analyzing change on a group level is not always meaningful since individual patients might differentiate in treatment changes. But by identifying latent classes, subgroups connected by same change trajectories (Wardenaar, 2020), we may be able to more accurately model how patients change during treatment on a more individual level. Latent classes allow to observe change of groups with similar trajectories of treatment change. This information can be used for tailoring treatment to an individual group of patients which could lead to more effective and efficient care, thus reducing costs (Ram & Grimm, 2009). Additionally, it can be also useful to identify characteristics of groups predicting the membership of their latent classes of change (Preacher, Wichman, MacCallum & Briggs, 2008) which makes it easier for professionals to tailor the treatment of patients.

But how do we detect differences in change during treatment between patients in a meaningful way? Generally, change can be detected with more than one method. However, a fitting method would be using latent growth mixture modeling (LGMM). This approach is suitable because the growth mixture model is used to identify multiple unobserved sub-populations, investigate differences among those groups and describing longitudinal change within each unobserved sub-population (Ram & Grimm, 2009). Furthermore, LGMM enables to differentiate groups of treatment change with previously mentioned latent classes (Wardenaar, 2020). It is a method suited to model non-linear patterns of change since it offers more flexibility (Ram & Grimm, 2007; McArdle & Epstein, 1987) and it is used to describe and test hypotheses about between-person differences and within-person change (Ram & Grimm, 2009). This method has been applied to treatments of various medical issues like breast cancer (Bower et al., 2018), multiple sclerosis (Westergaard, Skovgaard, Magyari & Kristiansen, 2022) and depression (Carter, Sperandei, Chitty & Page, 2022). Furthermore, LGMM was used multiple times before to identify trajectories in context with ED treatment (Colder et al., 2001; Hilbert et al., 2018; Espel-Huynh et al., 2020; Linardon et al., 2016).

Indeed, latent classes of change were found in ED treatment. Hilbert et al. (2018) researched early change in binge-eating disorder in a sample of 178 BED patients over a treatment period of four months consisting of CBT and internet-based guided self-help. They found that change can be described in four different classes: *Class 1: Low level binge eating stable*, *Class 2: Low level binge eating decreasing*, *Class 3: Medium level binge eating decreasing* and *Class 4: High level binge eating decreasing*. The last class was excluded and not considered in the analysis due to low number of members. Furthermore, Espel-Huynh et al. (2020) investigated latent trajectories of change in overall ED symptoms among female patients in residential care (N=360), diagnosed with AN, BN, and OSFED. The data acquisition was 5 weeks with weekly measurements of symptoms. The results presented that a three-class model was fitting best when modelling ED symptom change including the classes gradual response, rapid response and low-symptom with static response. Furthermore, it showed that most patients were assigned to the gradual response class followed by rapid response and lastly low-symptom with static response. For further research Espel-Huynh et al. (2020) suggested to conduct more research on classes of ED change trajectories to validate the three-class model with bigger sample sizes. Research approaching a similar topic to the current paper by de Vos, Radstaak, Bohlmeijer & Westerhof (2022) titled modeling trajectories of change in psychopathology and wellbeing (WB) for ED outpatients, also found the three-class model to be most characteristic for the ED pathology trajectories. The classes were found in a sample of 442 participants over a treatment duration of twelve months. The majority of the participants were assigned to high baseline EDP, slow recovery followed by moderate baseline EDP, no recovery and lastly high baseline EDP, substantial recovery (de Vos et al., 2022). The results of the study showed variability of patients and their treatment trajectories and that the change might be influenced by characteristics of the patients like early EDP change or hope for recovery. Summarizing, all previously stated papers showed that researchers agreed on a consensus for the number of classes when modelling change of EDP but the quality of classes differs from study to study. Mentionable is that the studies are not directly comparable. The studies by de Vos et al. (2022) and Espel-Huynh et al. (2020) were modeling change for transdiagnostic EDs while Hilbert et al. (2018) modeled BED symptom change. Additionally, Hilbert et al. (2018) gathered the data over a period of four months with a sample size of 178 participants, Espel-Huynh et al. (2020) modelled change with a sample of 360 participants over five weeks and de Vos et al. (2022) over one year with 442 participants. Although the studies differentiated, they highlighted the variability and the need for tailoring the treatment to the classes.

The identification of predictors of change can be crucial to optimize treatments (Kraemer, Wilson, Fairburn & Agras, 2002; Porter & Chambless, 2015) and to react to change trajectories in advance. Hilbert et al. (2018) found early change as a predictor for class membership and de Vos et al. (2022) found that ED type, general psychopathology, early change, and hope for recovery are possible predictors for class membership of EDP change. Besides a low symptom level, WB is considered a central part of mental health (de Vos et al., 2017; Noordenbos, 2011). Research has also shown a positive correlation between hope, which is predictive for EDP change, and psychological well-being (Murphy, 2023; Singh, Singh, Singh, & Srivastava, 2013). This is underlined by the two continua model that showed symptom level and well-being are two different dimensions but indeed related (Westerhof & Keyes, 2010). Although a correlation between WB and EDP have been researched and confirmed in previous research, WB was not tested as a predictor for EDP change. It can be argued that a higher level of WB at the start of the treatment can be considered a solid base for a positive treatment effect and positive influence for symptom change. Therefore, WB might be a possible predictor for EDP change due to its negative correlation to each other.

## Current study

In sum, previous research on EDP change trajectories showed three different classes of eating disorder change. This research replicates parts the study by De Vos et al. (2022) to validate the three-class model hypothesis for EDP change trajectories shown in the studies by Espel-Huynh et al. (2020) and Hilbert et al. (2018). The current study uses the same study design with and extended sample with more participants because the data collection for this study was longer and therefore the sample size is larger, which will provide increased representativity to the results of the study compared to previous studies (Andrade, 2020). Although pathology and well-being are related concepts, the predictive value for EDP symptom change is not clear. This research will provide clarity on how WB works as a predictor for class membership for EDP change. Consequently, the research questions (RQ) of this paper are 1. *What class model provides the best fit for eating disorder symptom change trajectories?* and *2. Does wellbeing predict class membership for eating disorder patients?* Regarding the first RQ, the hypothesis is that the three-class model will provide the best fit for EDP change trajectories. This hypothesis is built on previous research by Hilbert et al. (2018), de Vos et al. (2022) and Espel-Huynh et al. (2020) where three class models were found to describe the trajectories of EDP the best. Since the majority of the participants in the study by de Vos et al. (2022) were assigned to the *high baseline EDP, slow recovery* class, this is also to be expected one of the three classes in the current study. The hypothesis for the second research question is that WB will predict class membership for EDP trajectories. This hypothesis is based on WB being positively related to the identified predictor hope for recovery (Murphy, 2023; Singh et al., 2013) in addition to the negative correlation between WB and EDP (Murphy, 2023; Singh, et al., 2013). It is hypothesized that a higher level of WB will predict a lower level of EDP and faster reduction of EDP.

# Methods

## Design

The study used a naturalistic and longitudinal study design (Hua, & David, 2008). EDP, general pathology and WB of the same participants were measured every three months over a time span of one year, so overall five measurements per participant. The measurements were made for each participant consisting of the intake interview in addition to measurements every three months. After the initial measurement, the patients received therapy sessions twice a week over the period of twelve months. These sessions included psycho-education, food management, cognitive behavioral therapy, and insight-oriented psychotherapy. The process was monitored and evaluated by the multiprofessional team mentioned specialized in ED treatments (de Vos et al., 2022). This paper only used the measures on eating disorder pathology and well-being, the measures on general pathology were not needed for the research aims of this paper. The goal of the study is to identify latent groups of change based on the change of eating disorder symptoms and to test well-being as a predictor for group membership of the latent classes of change.

**Participants and treatment**

The participants were patients for outpatient treatment at Stichting Human Concern, a specialized treatment center for EDs in the Netherlands. Before the patients could participate in the study, a team of professionals specialized in EDs and their treatment assessed them (de Vos, Netten & Noordenbos, 2016). The inclusion criteria for the patients were a minimum age of 16 years and a DSM 5 (American Psychiatric Association, 2013) ED diagnosis at intake. Additionally, the patients had to go through six months of prior treatment, and they had to sign the informed consent. In total, 1283 patients were initially included in the study and the data collection was made between 2015 and 2020. More detailed information is provided in Table 1.

**Table 1**
*Characteristics of the sample (n=1283)*

| | | |
|---|---|---|
| Age at the start of treatment (in years) | | |
| | mean (m) | 27,44 |
| | < 20 | 203 |
| | 20 – 25 | 502 |
| | 26 – 30 | 239 |
| | 31 – 40 | 199 |
| | 41 – 50 | 100 |
| | 50 < | 30 |
| Gender | | |
| | Female | 1250 |
| | Male | 33 |
| Primary ED diagnosis | | |
| | AN | 404 |
| | BN | 274 |
| | BED | 99 |
| | OSFED | 506 |
| Start year of the treatment | | |
| | 2015-2016 | 408 |
| | 2017-2018 | 498 |
| | 2019-2020 | 377 |

## Procedure

The Behavioral, Management, and Social Sciences Ethics Committee of the University of Twente approved the study protocol (registration number BCE15484). The participants had to fill in a battery of questionnaires. This is part of the intake procedure and the baseline of Routine Outcome Monitoring (ROM) and was used to monitor progress during the treatment. The intake interview included information on the psychiatric history of family members in the first line, psychotropic medication, earlier treatments, prior hospitalization for the ED, duration and start year of the ED, complex trauma (sexual abuse, verbal/physical abuse, severe instability in the family, or multiple negative life events), daily activities such as work or study and the financial situation. Furthermore, based on the instructions of Stichting Benchmark GGZ, a national benchmark for treatment outcomes of mental health providers in the Netherlands, educational level and living situation were taken in (Stichting Benchmark GGZ, 2013). Lastly, the informed consent was presented to the patients and the aims of the study were explained. For the upcoming year, the patients participated in measurements of general pathology, the measurements were excluded from this study, EDP and finally well-being every three months.

## Material

### *Eating Disorder Pathology (EDP)*

In order to examine the eating disorder pathology, the global score of the 36 item Eating Disorder Examination Questionnaire (EDE-Q) was used (Fairburn & Beglin, 1994), that offered a strong internal consistency that increased to an excellent internal consistency (Taber, 2018). Coefficients are presented in Table 2. These questionnaire items were scored on a seven-point Likert scale from 0 (not 1 day) to 6 (every day) where the participants could self-report the frequency of symptoms in the last 28 days. Lower scores are associated with a lower level of EDP. The total score was calculated by adding up the scores of four scales Restrain, Eating Concern, Weight Concern and Shape Concern and dividing them by four.

### *Well-being*

Well-being was measured by the usage of the Mental Health Continuum Short Form (MHC-SF) (Keyes, 2002; Lamers et al., 2011). The questionnaire contained 14 items, rated on a six-point Likert scale ranging from 0 (never) to 5 (always), and it measured different dimensions of well-being: overall, emotional, psychological, and social well-being with higher scores on the scales indicating higher well-being of the participants. The total score was calculated by summing all item scores. The internal consistencies for the current study are presented Table 2. It is observable that Cronbach's alpha provided a strong to excellent internal consistency (Taber, 2018) and improved over the course of the study.

**Table 2**

*Cronbach's alpha of the EDE-Q and MHC-SF per measurement.*

| | EDE-Q | MHC-SF |
|---|---|---|
| Measurement 1 | .910 | .901 |
| Measurement 2 | .926 | .909 |
| Measurement 3 | .937 | .925 |
| Measurement 4 | .965 | .923 |
| Measurement 5 | .950 | .955 |

**Data analysis**

The data was mainly analyzed with R (4.2.2) with the package latent growth mixture model (3) (LGMM), and some small procedures were executed in SPSS. The first step was to get an overview of the data by analyzing the demographics of the sample. After that, a new data set was created that contained the global scores of the EDE-Q for all five points of the measurement which was needed for the main analysis LGMM and the initial total score of the MHC-SF which was needed for the second analysis regarding the predictor. Then the percentage of missing data was calculated and afterwards the missing data patterns were observed. The next step was to calculate and observe the descriptive statistics to get an idea of the basic statistical information of the five measurements of EDE-Q global scores including mean, standard deviation, skewness and kurtosis. If the skewness and kurtosis coefficients are in the range of ±2, the normality is considered acceptable (Burdenski Jr, 2000; George & Mallery, 2010) and Q-Q plots were created as a visual aid to understand the data (Appendix). The information was used to check the normality of the EDE-Q scores for all points of measurement because tests like the Shapiro-Wilk (SW) test or the Kolmogorov-Smirnov (KS) test are not reliable for a sample size this large (Razali & Wah, 2010). After the normality and the missing data for the data set was checked, it was decided to impute data for participants with one missing value to improve the power of the analysis and representable value of the analysis (Andrade, 2020) since cases with missing values are excluded from the analysis LGMM. A mean implementation was chosen as the implementation method. After the missing data was imputed and the cases with missing values > 1 were excluded from the sample, the first step was to estimate the underlying baseline model. The baseline model helps to understand whether there is change in EDP and how the change can be described. First, linear models with fixed effects on the intercept and slope, fixed effects and random intercept and fixed effect with random intercept and slope were compared to the data. Fixed effect or intercept means that slope or intercept are constant, random means that these are not constant and can vary (Fan, 2010). The quadratic effect model was also added. These baseline models were chosen based on the study by de Vos et al. (2022) and their fit to the data in the study. To find the best fitting model, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are crucial. The lower the value of AIC and BIC, the better is the fit of the model for the data (Muthén, 2003; Nylund, Asparouhov, & Muthén, 2007). Although both values are important, the BIC indicates the quality of fit for the current data, whereas AIC shows how future values can be predicted (Chakrabarti & Ghosh, 2011) which make the BIC more valuable over the AIC for the current study. After the best fitted baseline model was chosen, new models were estimated with multiple classes in a sequential order. This helped to understand individual change of EDP. The first analysis included two classes, which was compared to the baseline model analysis, where only one class exists. The Integrated Complete Likelihood (ICL) is important because it indicates the fit of model, as the BIC, but additionally includes the Entropy. The Entropy is a value between 1 and 0 that shows the classification certainty of the models (Biernacki, Celeux & Govaert, 2000) whereas a value close to 1 indicates optimal classification certainty (de Vos et al., 2022). The Entropy will start at 1 for the one class model, which included the whole sample because it is 100% certain that the data belongs to the data set, but it will decrease eventually when further classes are added. So if the two-class model is more fitting than the baseline model, a further class was added to the analysis to test if more classes would lead to lower and therefore better AIC, BIC and ICL values. This process was done until the addition of classes was not resulting in an improvement of the models fit. Lastly a multinominal regression was conducted in SPSS to get insight on a predictor for class membership. The initial general well-being measurement was chosen as the predictor variable for class membership where membership to a specific class indicated a specific pattern of EDP change.

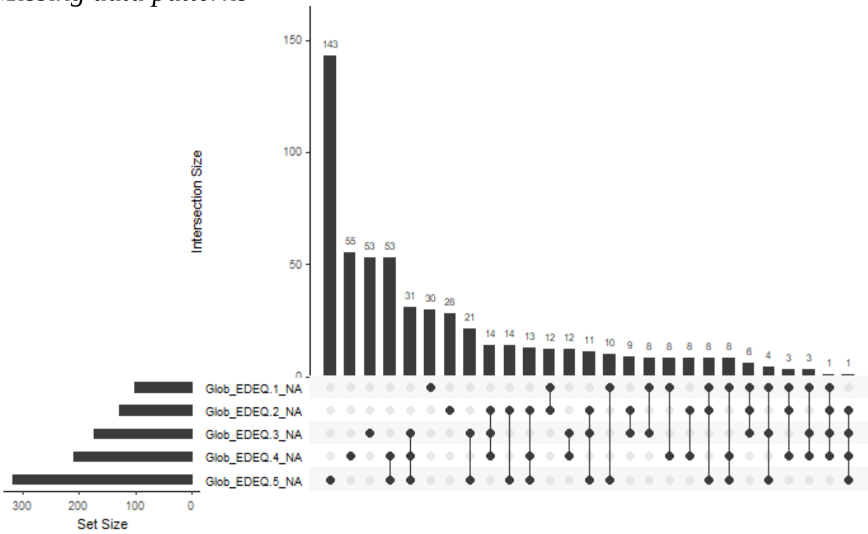
# Results

## Missing Data

The data set used for the analysis included seven variables: five measurements of the EDE-Q global score, initial total score of the MHC-SF and the patient ID. Overall, 12.08 % of the data was missing in the data set. The study design was longitudinal and included five points of measurement. Figure 1. shows how the missing data of the EDE-Q was spread across the cases in different patterns.

**Figure 1**

*Missing data patterns*




Only 716 of the cases 1283 were complete (55.81%). Figure 1 shows the non-complete cases and the pattern of missing data. The black dots show where measurements were completed and the grey dots where measurements were missing. It appears that the most participants, 309, only missed one measurement and completed all other measurements. The 85 participants missing out measurement 4 and 5 or 3, 4, and 5 et cetera did not continue the measurements at some point in the study and might have dropped out. However, there were participants missing multiple measurements but completed measurements in the further process of the treatment. This indicates that they might have missed measurements due to various reasons means they did not drop out and were continuing the study afterwards. Since the LGMM analysis only uses complete cases, data mean imputation was executed for cases missing 1 value (n=309). Consequently, 1025 cases had complete data after the imputation.

**Descriptive statistics**

Descriptive data included mean, standard deviation (sd), skewness and kurtosis. The values are presented in Table 3.

**Table 3**

*Descriptive Statistics of the five variables.*

| | mean | sd | skewness | kurtosis |
|---|---|---|---|---|
| Glob_EDEQ.1 | 3.89 | 1.16 | -0.65 | -0.27 |
| Glob_EDEQ.2 | 3.60 | 1.24 | -0.57 | -0.36 |
| Glob_EDEQ.3 | 3.27 | 1.33 | -0.33 | -0.65 |
| Glob_EDEQ.4 | 3.03 | 1.41 | -0.20 | -0.91 |
| Glob_EDEQ.5 | 2.87 | 1.41 | -0.10 | -0.91 |

The data showed an acceptable skewness and kurtosis and therefore the data was considered normal. All skewness coefficients were negative that indicated a skewed tail to the left (Burdenski Jr, 2000; George & Mallery, 2010). Table 4 Provides information on the initial well-being total scores.

**Table 4**

*Descriptive statistics for initial well-being total scores (n=1283).*

| | |
|---|---|
| Missing values | 115 |
| Mean (m) | 2.36 |
| Standard Deviation | .90 |
| Minimum | 0 |
| Maximum | 4.71 |

**Latent growth curve modeling**

***Baseline model***

The underlying growth model of the data, the baseline model, was estimated. The fits of the models can be found in Table 5. First the fits of the linear models were tested. Additionally, the quadratic model was tested because it had the best fit to the data in the study by de Vos et al. (2022) and therefore had to be considered for the replicated study.

**Table 5**

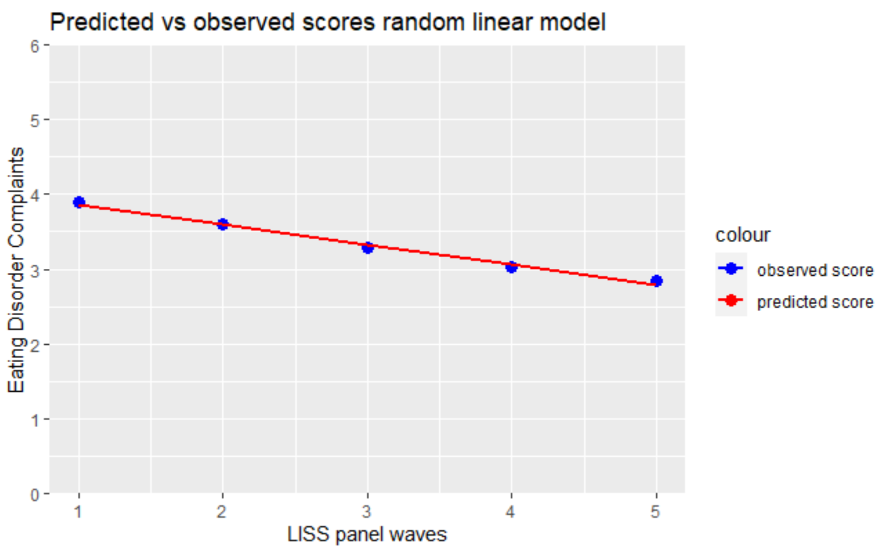
*Fitting of the model to the data.*

| | | AIC | BIC | loglikelihood |
|---|---|---|---|---|
| 1 | Linear model with fixed effects | 16917.98 | 16931.77 | -8455.988 |
| 2 | Linear model with fixed effects and random slope | 13851.00 | 13870.73 | -6921.500 |
| 3 | Linear model with random intercept and slope | 13513.47 | 14358.84 | -6750.737 |
| 4 | Quadratic model | 14324.61 | 14360.70 | -7155.303 |

Although the Linear models with random intercept and slope and the quadratic models AIC and BIC provided a rather similar fit of the models to the data, the linear model with random intercept and slope showed a slightly higher AIC but a slightly lower BIC. Since the BIC was slightly better for the Linear model with random intercept and slope, I decided that the Linear model with random intercept and slope had the best fit for the data. The fit of the selected model is visualized in Figure 4. It is observable that the EDP decreased from an EDE-Q score of approximately 4 to a score below 3. According to Aardoom, Dingemans, Op't Landt & Van Furth (2012) the normative score of the EDE-Q global score for the ED population is 4.02. That shows that this sample displayed an about average EDP level before the treatment. The change of the EDE-Q global score showed an improvement of the symptoms during the treatment.

**Figure 4**

*Fit of Linear model with random intercept and slope to the data.*



***Classes***

Loglikelihood, AIC, BIC, Entropy and ICL are reported per classes in Table 5. which indicates the fit of the model including the classes. First an additional class was added to the base model, so a two-class model, and was tested against the base model with one class. The two-class model showed lower values for all reported measures and consequently the two-class model had an improved fit of the data compared to the base model. Since adding a class improved the fit, a third class was added to the model and compared to the previous two class model. The additional class provided lower values for BIC and Loglikelihood but a higher value for ICL and Entropy. Although the two-class model displayed a slightly lower and therefore an improved ICL and entropy, the overall fit of the three-class model is better. Again, an additional class was added to the model to check if it would improve the fit. The four-class model had slightly lower values for loglikelihood, Entropy and AIC but higher values for BIC, and ICL. Those values indicate that the fit of the four-class model is not better than the fit of the three-class model. Although the scores in the table only differentiate slightly, the three-class model had the best fit for the data. This decision is based on a combination of multiple factors. The three-class model provided the slightly worse ICL compared to the two-class model however, BIC, Entropy and Loglikelihood improved by adding a third class.

**Table 6**

*Overview of the classes and their fit to the data.*

| | Classes | Loglikelihood | AIC | BIC | Entropy | ICL |
|---|---|---|---|---|---|---|
| Baseline model | 1 | -7155.303 | 14324.61 | 14360.70 | 1.0000000 | 14360.70 |
| | 2 | -6697.609 | 13415.22 | 13464.54 | 0.5942147 | 11672.64 |
| | 3 | -6673.917 | 13375.83 | 13444.89 | 0.6655464 | 11723.07 |
| | 4 | -6660.762 | 13357.52 | 13446.31 | 0.6633853 | 11808.85 |

The three classes of the selected model are illustrated in Figure 5. Intercept, slope, significance and percentages assigned to each class are presented in Table 7.

**Figure 5**

*Illustration of the course of the classes*

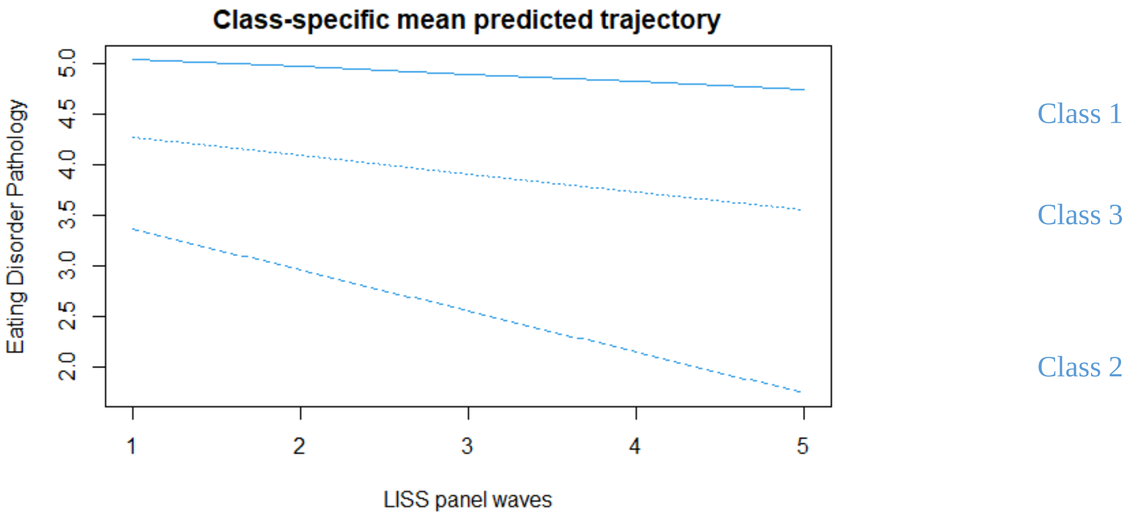



**Table 7**

*Intercept, slope, and significance of the classes.*

| | intercept | slope | p-values | percentages |
|---|---|---|---|---|
| Class 1 | 5.114 | -.066 | <.001 | 10.54% |
| Class 2 | 3.667 | -.336 | <.001 | 54.44% |
| Class 3 | 4.619 | -.215 | <.001 | 35.02% |

The first-class started with EDE-Q mean global score of 5.114 which is more than one score point higher compared to the norm values. Looking at the graph, a slow reduction in EDP is observable however, the score is still above the normative score of 4.02. Therefore, the label for class 1 is *high symptom baseline, slow recovery* (10.54%), The second-class starts with an EDE-Q score of 3.667 which is lower than the normative score and the graph shows a great improvement during the treatment to a score < 2. An EDE-Q global score < 2 comes close to the normative scores of a non-ED population of .93 (Aardoom et al., 2012). This class is labeled *low symptom baseline, fast recovery* (54.44%). AT the start of the treatment the EDE-Q score of the third-class comes closest to the normative score of 4.02. The graph shows a small but greater improvement of EDP symptoms during the treatment compared to class 1. The post treatment score was approximately 3.5 which is under the EDE-Q norm score. Consequently, the third class is labeled *moderate symptom baseline, moderate recovery* (35.02%).

## Multinominal regression

A multinominal regression was conducted to predict *class membership* based on their initial *well-being* at their first measurement. The predictor variable was *well-being* and the dependent variable was *class membership*. The values for the fitting of the model were significant with B = 403.514, $p < .001$. The Nagelkerke statistics showed that the predictor variable explained 5.4% of the variation of the dependent variable. The parameter estimates showed that the Wald test for class 2 in relation to class 1 for the predictor *well-being* is 19.026 with $p < .001$ is significant while the Wald statistic test for class 3 in relation to class 1 provided a not significant value of 0.230 with $p = .631$, $p > .001$. The null-hypothesis could be rejected for predictor *well-being* when it is predicting class 2 in relation to class 1 but is not rejected when well-being predicts class 3 in relation to class 1. Consequently, if a patient has an increased well-being score by one point, the multinominal log-odds of being member of class 2 compared to class 1 would be expected to increase by .549 units with a probability of 73% [Exp *(B)* = 1.731, 95% CI (1.353, 2.216)].

## Discussion

### What class model provides the best fit for eating disorder symptom change trajectories?

The current paper investigated symptom change trajectories during the treatment of eating disorder outpatients. The goal was to study classes of trajectories of EDP change. The results showed that the three-class linear model with random slope and intercept had the best fit to the data of the sample. The three classes were labeled with *high symptom baseline, slow recovery*, the second-class *low symptom baseline, fast recovery* and the third-class *moderate symptom baseline, moderate recovery*. Concluding, the results of the study showed variation in EDP change trajectories that could be described by three latent classes of change. Consequently, the hypothesis regarding the first RQ could be confirmed and also the hypothesis on the reoccurring *high symptom baseline, slow recovery* class could be confirmed as well.

Previous related research by Hilbert et al. (2018) on symptom change trajectories for eating disorder focused on change of binge eating symptoms. Although they found three classes of change with a significantly smaller sample, the sample displayed a smaller level of ED symptoms and the classes symptoms where decreasing in two out of three classes. This study showed at least small improvements is every class and was not limited to BE patients. However, Espel-Huynh et al. (2020) investigated trajectories of eating disorder treatment response among female patients. The research by Espel-Huynh et al. (2020) also established three classes of symptom change trajectories of eating disorder patients. The base model of data was quadratic, and the three classes were labeled gradual response, rapid response and low symptom with static response. Although the researchers found three classes as well, the quality of the classes differed in initial EDP level and change velocity to the classes of the current study. A striking difference is that the data collection was conducted weekly over five weeks and the symptom severity was measured with the Progress Monitoring Tool for Eating Disorders (PMED). This questionnaire was developed for transdiagnostic patient populations with specific sensitivity to change during treatment (Espel-Huynh et al., 2020). The EDE-Q just measures symptom severity without specific sensitivity to change (Aardoom, Dingemans, Op't Landt & Van Furth, 2012). Thus, a possible explanation for the differences is class quality could be that the measurements of symptom severity with the PMED might model EDP symptom change more accurate which resulted in better quality classes compared to this study. Additionally, the different durations of the studies might have influenced the results of the quality of the classes as well. Espel-Huynh et al. (2020) provided detailed information on early change by measuring weekly while this study modelled symptom change not that detailed but over a longer duration. Consequently, the classes by Espel-Huynh et al. (2020) rather described a shorter timeframe at the beginning of the treatment while this study classes tried to provide information on change for a duration close to traditional CBT treatment for ED patients (Murphy, Straebler, Cooper, & Fairburn, (2010).

The study by de Vos et al. (2022), who were modeling trajectories of change in EDP during eating disorder outpatient treatment, were researching change trajectories similar to the current study. The current paper replicated the study by de Vos et al. (2022) and found similar results: both studies found the three-class model as the best fitting model, but the base model and the classes differed from each other. De Vos et al. (2022) found the quadratic base model which makes the change trajectories of the three classes quadratic as well. The labels of the classes differed. Although one of the three classes stayed the same, high baseline EDP with slow recovery, de Vos et al. (2022) found one moderate baseline EDP with no recovery class and one high baseline EDP with substantial recovery whereas this paper established that all classes displayed at least a slow recovery. However, the current study found another class with low EDP symptoms which was not found in the study of de Vos et al. (2022). The difference between the studies can be explained by the large sample size, since the data collection for the current study was two years longer than the data collection of the previous study which resulted in a sample double the size. Very large samples have the tendency to ascribe statistically significance to small differences although they are not clinically significant (Faber & Fonseca, 2014). Smaller sample sizes are more sensitive to extreme values and differences between sample and population can happen just by chance (Tipton, Hallberg, Hedges, & Chan, 2017). So, the differences in quality of classes might be due to chance or false positive significance.

Additionally, the data collection took part during the COVID-19 pandemic. The pandemic had an influence on the anxiety of the eating disorder patients as well as the provided healthcare of the professionals (Termorshuizen et al., 2020) however the effect is still unclear (Fernández-Aranda et al., 2020). COVID-19 might have increased the general symptom level of the participants due to the additional anxiety and might have impacted the EDP symptom change of the patients. This might explain the higher EDP symptom level of this sample.

**Does wellbeing predict class membership for eating disorder patients?**

This study tried to identify well-being as a possible predictor for group membership of latent classes of EDP symptom change. The results showed that well-being did predict group membership for two of the three classes. A by one score point increased value in well-being at the beginning of the treatment increases the chances of being in the *low symptom baseline, fast recovery* class compared to *high symptom baseline, slow recovery* by 73%. That means a higher well-being score at the begin of the treatment predicted membership for the *low symptom baseline, fast recovery* class indicating that a higher well-being score is related to lower EDP and faster recovery. Consequently, a lower well-being score increases the chance for the membership in the *high symptom baseline, slow recovery* class. The predictive value of well-being for the *moderate symptom baseline, moderate recovery* class was not significant. Therefore, the hypothesis regarding this research question was partly fulfilled, well-being predicted two out of three class memberships and higher well-being predicted lower EDP symptom levels and faster recovery.

Although research has shown that well-being and symptom severity are related concepts (Murphy, 2023; Singh et al., 2013; Westerhof & Keyes, 2010; Weijers et al. 2022), there is a lack of research on the correlation of EDP and well-being but also the predictive ability for EDP change. De Vos et al. (2022) identified a positively related concept of wellbeing, hope for recovery, as a predictor of EDP change classes which is partly in line with the findings of this study. Furthermore, Singh et al. (2013) established a positive correlation between hope and psychological well-being and a negative correlation for the previous mentioned concepts and depression symptoms. Consequently, an increase in either hope or psychological well-being was related to decreasing depression pathology. This is supported by findings of the study by Weijers et al. (2022) where general well-being was negatively correlated to symptom severity of depression. Consequently, this study validated that higher well-being is related to lower EDP symptoms however, if well-being predicted class membership based just on the initial well-being score or if the membership prediction is based on the EDP change trajectories, is not completely clear. The results showed that all three classes differentiate in initial EDP score but also differe in recovery speed, so it is not visible why the *high symptom baseline, slow recovery* class and the *low symptom baseline, fast recovery* class was predicted by well-being and the *moderate symptom baseline, moderate recovery* was not. What is visible is that well-being predicted class membership that differentiate drastically from each other but was unable to predict the class which was in-between the extremes.

**Strengths and limitations**

A strength of the study is that it fits the lack of research on the correlation of well-being and EDP as well as the predictive ability of well-being for latent classes of EDP change. Furthermore, the large sample size and the inclusion of men into the research increased the representativeness of the sample. The sample included a further group of people that could possibly suffer from ED can be considered strengths of the study. This increases the representativity of the results (Andrade, 2020). The study was able to find a three-classes model for EDP change which is in line with the previous studies and more importantly, it provides similar results compared to the replicated study. There are several limitations for this study. Firstly, the naturalistic study design bears limitations. The researchers had no control over the treatment and consequently had no control over unexpected factors or influences. These influences could not be respected while interpreting the results. Furthermore, self-report measurements can be influenced by biases (Adams, Soumerai, Lomas & Ross-Degnan, 1999) which could have resulted impaired validity of the results (de Vos et al., 2022). Additionally, the large amount of missing data is limiting the interpretation and precision of the results (Jakobsen et al., 2017; Cummings, 2013). The missing data pattern revealed that almost one quarter of the participants missed one EDP measurement. For these values mean imputation was used to dramatically increase the sample size but this might have impacted the results. Mean imputation can influence validity or biases positively but also negatively compared to the complete cases which depends on the data set and important to keep in mind (White & Carlin, 2010; Cummings, 2013). Furthermore, the selection of the class model might be considered a limitation. Although the three-class model was chosen as the model with the best fit, the four-class model provided some indicators indicating that it might have a better fit than the three-class model. Additionally, the linear model with random intercept and slope had the best fit for the data and was consequently chosen as the base model. However, the chosen linear base model implicates that all graphs of the classes are displayed linear as well although they might all be linear. So, the individual classes might be different from the linear base model.

### Implications for future research

Further research should focus on possible predictors of treatment trajectories for larger samples. Especially, the predictive value of well-being for ED treatment trajectories needs to be validated because of the existing lack of research. Then there is a lack of research on gender differences when it comes to classes of change for ED patients. Most studies used female samples, but this study included a small percentage of male participants to increase the representable value, however differences were not investigated yet. Additionally, the effectiveness of personalized treatments for ED patients according to the three classes of change could be tested in future research by e.g. using an increases amount of interventions focused on increasing well-being during the treatment. This can be compared to a control group. The results showed that the treatment trajectories of ED patients can be clustered in three distinct trajectories with different EDP levels and recovery speeds and consequently a need for more tailored ED treatment. Although the researchers agreed on three classes, the quality of those classes differs due to various factors like treatment duration, sample size and measurement instruments. This must be respected in clinical practice when tailoring ED treatments to the classes and additionally monitoring the trajectories of the patients is recommended to react to unexpected change. Furthermore, the assessment of the patient's well-being at the start of the treatment might predict not only the EDP level due to its correlation, but also the course of the treatment and the recovery.

## Conclusion

This research showed that EDP change during CBT treatment for transdiagnostic ED patients is described the best by a linear model with three latent classes of change. Additionally, this study identified a predictive value of well-being for class membership. Higher well-being at the start of the treatment is associated with lower EDP and faster recovery. Furthermore, well-being was able to predict class membership for two classes. The predictive value of well-being should be considered carefully since it is not clear yet if well-being predicts the EDP levels or the trajectories of the treatment. However, it might be useful to measure well-being at the start of the treatment to get in impression of the patient's needs.

## References

Aardoom, J. J., Dingemans, A. E., Op't Landt, M. C. S., & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating behaviors*, *13(4)*, 305-309. https://doi.org/10.1016/j.eatbeh.2012.09.002

Adams, A. S., Soumerai, S. B., Lomas, J., & Ross-Degnan, D. (1999). Evidence of self-report bias in assessing adherence to guidelines. *International Journal for Quality in Health Care*, *11(3)*, 187-192. https://doi.org/10.1093/intqhc/11.3.187

American Psychiatric Association, DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). American Psychiatric Publishing, Inc.. https://doi.org/10.1176/appi.books.9780890425596

Andrade, C. (2020). Sample size and its importance in research. *Indian journal of psychological medicine*, *42(1)*, 102-103. https://doi.org/10.4103/IJPSYM.IJPSYM_504_19

Atwood, M. E., & Friedman, A. (2019). A systematic review of enhanced cognitive behavioral therapy (CBT-E) for eating disorders. *International Journal of Eating Disorders*. https://doi.org/10.1002/eat.23206

Bell, C., Waller, G., Shafran, R., & Delgadillo, J. (2017). Is there an optimal length of psychological treatment for eating disorder pathology? *International Journal of Eating Disorders*, *50*, 687–692. https://doi.org/10.1002/eat.22660.

Berends, T., Boonstra, N., & Van Elburg, A. (2018). Relapse in anorexia nervosa: a systematic review and meta-analysis. *Current opinion in psychiatry*, *31(6)*, 445-455. https://doi.org/10.1097/YCO.0000000000000453

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, *22(7)*, 719-725. https://doi.org/10.1109/34.865189

Bower, J. E., Wiley, J., Petersen, L., Irwin, M. R., Cole, S. W., & Ganz, P. A. (2018). Fatigue after breast cancer treatment: Biobehavioral predictors of fatigue trajectories. *Health Psychology*, *37(11)*, 1025. https://doi.org/10.1037/hea0000652

Boswell, D. L., White, J. K., Sims, W. D., Harrist, R. S., & Romans, J. S. (2013). Reliability and validity of the Outcome Questionnaire–45.2. *Psychological Reports*, *112(3)*, 689-693. https://doi.org/10.2466/02.08.PR0.112.3.689-693

Burdenski Jr, T. K. (2000). Evaluating Univariate, Bivariate, and Multivariate Normality Using Graphical Procedures.

Carter, G., Sperandei, S., Chitty, K. M., & Page, A. (2022). Antidepressant treatment trajectories and suicide attempt among Australians aged 45 years and older: A population study using individual prescription data. *Suicide and Life-Threatening Behavior*, *52(1)*, 121-131. https://doi.org/10.1111/sltb.12812

Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. *Philosophy of statistics*, 583-605. https://doi.org/10.1016/B978-0-444-51862-0.50018 -6

Chang, P. G., Delgadillo, J., & Waller, G. (2021). Early response to psychological treatment for eating disorders: a systematic review and meta-analysis. *Clinical psychology review*, *86,* 102032. https://doi.org/10.1016/j.cpr.2021.10203

Colder, C. R., Mehta, P., Balanda, K., Campbell, R. T., Mayhew, K., Stanton, W. R., ... & Flay, B. R. (2001). Identifying trajectories of adolescent smoking: an application of latent growth mixture modeling. *Health Psychology*, *20(2)*, 127. https://doi.org/10.1037/0278 -6133.20.2.127

Cooper, A. A., & Conklin, L. R. (2015). Dropout from individual psycho-therapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review*, *40*, 57–65. https://doi.org/10.1016/j.cpr.2015.05.001

Cummings, P. (2013). Missing data and multiple imputation. *JAMA pediatrics*, *167(7)*, 656 -661. https://doi.org/10.1001/jamapediatrics.2013.1329

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of cognition and development*, *11(2)*, 121-136. https://doi.org/10.1080/15248371003699969

de Vos, J. A., Radstaak, M., Bohlmeijer, E. T., & Westerhof, G. J. (2022). Modelling trajectories of change in psychopathology and well-being during eating disorder outpatient treatment. *Psychotherapy Research*, 1-13. https://doi.org/10.1080/10503307.2022.2139647

de Vos, J. A., LaMarre, A., Radstaak, M., Bijkerk, C. A., Bohlmeijer, E. T., & Westerhof, G. J. (2017). Identifying fundamental criteria for eating disorder recovery: a systematic review and qualitative meta-analysis. *Journal of eating disorders, 5(1),* 1-14. https://doi.org/10.1186/s40337-017-0164-0

Eating Disorders Review (2005). Bulimia Nervosa: Working Towards Better Definitions of Remission and Relapse. *Eating Disorder Review, 16(5).* Retrieved from: https://eatingdisordersreview.com/bulimia-nervosa-working-toward-better-definitions-of-remission-and-relapse/

Espel-Huynh, H. M., Zhang, F., Boswell, J. F., Thomas, J. G., Thompson-Brenner, H., Juarascio, A. S., & Lowe, M. R. (2020). Latent trajectories of eating disorder treatment response among female patients in residential care. *International Journal of Eating Disorders, 53(10),* 1647-1656. https://doi.org/10.1002/eat.23369

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental press journal of orthodontics, 19,* 27-29. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire?. *International journal of eating disorders, 16*(4), 363-370.

Fan, S. (Ed.) (2010).. (Vols. 1-0). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781412961288

Fernández-Aranda, F., Casas, M., Claes, L., Bryan, D. C., Favaro, A., Granero, R., Gudiol, C., Jiménez-Murcia, S., Karwautz, A., Le Grange, D., Menchón, J. M., Tchanturia, K., & Treasure, J. (2020). COVID-19 and implications for eating disorders. *European eating disorders review: the journal of the Eating Disorders Association, 28(3),* 239–245. https://doi.org/10.1002/erv.2738

Geller, J., & Drab, D. L. (1999). The readiness and motivation interview: A symptom-specific measure of readiness for change in the eating disorders. *European Eating Disorders Review, 7,* 259–278. https://doi.org/10.1002/(SICI)1099-0968(199908)7:4<259::AID-ERV295>3.0.CO;2-6

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference,* 17.0 update (10a ed.) Boston: Pearson.

Hay, P. (2020). Current approach to eating disorders: a clinical update. *Internal medicine journal, 50(1),* 24-29. https://doi.org/10.1111/imj.14691

Hilbert, A., Herpertz, S., Zipfel, S., Tuschen-Caffier, B., Friederich, H.-C., Mayr, A., … de Zwaan, M. (2018). Early Change Trajectories in Cognitive-Behavioral Therapy for Binge-Eating Disorder. *Behavior Therapy*. doi:10.1016/j.beth.2018.03.013

Hua, Z., & David, A. (2008). Study design: Cross-sectional, longitudinal, case, and group. *The Blackwell guide to research methods in bilingualism and multilingualism*, 88-107.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology, 17(1),* 1-10. https://doi.org/10.1186/s12874-017-0442-1

Jordan, J., Joyce, P. R., Carter, F. A., Horn, J., McIntosh, V. V., Luty, S. E., ... Bulik, C. M. (2008). Specific and nonspecific comorbidity in anorexia nervosa. *International Journal of Eating Disorders*, *41(1),* 47–56. https://doi.org/10.1002/eat.20463

Keel, P. K., & Brown, T. A. (2010). Update on course and outcome in eating disorders. *International Journal of Eating Disorders*, NA–NA. doi:10.1002/eat.20810

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of health and social behavior*, 207-222.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of general psychiatry*, *59(10)*, 877-883. https://doi.org/10.1001/archpsyc.59.10.877

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *Journal of clinical psychology*, *67*(1), 99-110.

Linardon, J., Brennan, L., & De la Piedad Garcia, X. (2016). Rapid response to eating disorder treatment: A systematic review and meta-analysis. *International Journal of Eating Disorders*, *49(10),* 905-919 https://doi.org/10.1002/eat.22595

Linardon, J., Hindle, A., & Brennan, L. (2018). Dropout from cognitive-behavioral therapy for eating disorders: A meta-analysis of randomized, controlled trials. *International Journal of Eating Disorders*, *51(5),* 381-391. https://doi.org/10.1002/eat.22850

Linardon, J., Wade, T. D., De la Piedad Garcia, X., & Brennan, L. (2017). The efficacy of cognitive-behavioral therapy for eating disorders: A systematic review and meta analysis. *Journal of consulting and clinical psychology, 85(11),* 1080. https://doi.org/10.1037/ccp0000245

Luijten, C. C., Kuppens, S., van de Bongardt, D., & Nieboer, A. P. (2019). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF) in Dutch adolescents. *Health and quality of life outcomes*, *17(1),* 1-10. https://doi.org/10.1186/s12955-019-1221-y

McArdle, J.J., & Epstein, D.B. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133. https://doi.org/10.2307/1130295

Meads, C., Burls, A., Gold, L., & Jobanputra, P. (1999). In-patient versus out-patient care for eating disorders. *Department of Public Health and Epidemiology, West Midlands Development and Evaluation Service Report, (17).*

Melchior, H., Schulz, H., Kriston, L., Hergert, A., Hofreuter-Gätgens, K., Bergelt, C., Morfeld, M., Koch, U., & Watzke, B. (2016). Symptom change trajectories during inpatient psychotherapy in routine care and their associations with long-term outcomes. Psychiatry Research, 238, 228–235. https://doi.org/10.1016/j.psychres.2016.02.046

Milos, G., Spindler, A., Schnyder, U., & Fairburn, C. G. (2005). Instability of eating disorder diagnoses: prospective study. *The British Journal of Psychiatry, 187(6),* 573-578. https://doi.org/10.1192/bjp.187.6.573

Murphy, E. R. (2023). Hope and Well-being. *Current Opinion in Psychology*, 101558. https://doi.org/10.1016/j.copsyc.2023.101558

Murphy, R., Straebler, S., Cooper, Z., & Fairburn, C. G. (2010). Cognitive behavioral therapy for eating disorders. *Psychiatric Clinics*, *33(3),* 611-627. https://doi.org/10.1016/j.psc.2010.04.004

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, *8,* 369–377. https://doi.org/10.1037/1082-989X.8.3.369

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, *24(6),* 882-891. https://doi.org/10.1111/j.1530 0277.2000.tb02070.x

National Eating Disorders Association (2022). Avoidant restrictive food intake disorder (AFRID). Retrieved from Avoidant Restrictive Food Intake Disorder (ARFID) | National Eating Disorders Association on 15.11.2022.

Noordenbos, G. (2011). Which Criteria for Recovery Are Relevant According to Eating Disorder Patients and Therapists? *Eating Disorders*, *19(5),* 441-451. doi:10.1080/10640266.2011.618738

Nylund, K.L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569. https://doi.org/10.1080/10705510701575396

Owens, R. A., Attia, E., Fitzpatrick, J. J., Phillips, K., & Nolan, S. (2021). Eating Disorders: Identification and Management in General Medical and Psychiatric Settings. Journal of the *American Psychiatric Nurses Association*, 107839032199971. https://doi.org/10.1177/1078390321999713

Porter, E., & Chambless, D. L. (2015). A systematic review of predictors and moderators of improvement in cognitive-behavioral therapy for panic disorder and agoraphobia. *Clinical Psychology Review*, *42*, 179-192. https://doi.org/10.1016/j.cpr.2015.09.004

Peckmezian, T., & Paxton, S. J. (2020). A systematic review of outcomes following residential treatment for eating disorders. *European Eating Disorders Review*, *28(3)*, 246-259. https://doi.org/10.1002/erv.2733

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling* (No. 157). Sage.

Qian, J., Wu, Y., Liu, F., Zhu, Y., Jin, H., Zhang, H., ... & Yu, D. (2021). An update on the prevalence of eating disorders in the general population: a systematic review and meta analysis. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 1-14.* https://doi.org/10.1007/s40519-021-01162-z

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development, 31,* 303–316. https://doi.org/10.1177/0165025407077751

Ram, N., & Grimm, K. J. (2009). Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33(6),* 565–576. https://doi.org/10.1177/0165025409343765

Razali, N. M., & Wah, Y. B. (2010, June). Power comparisons of some selected normality tests. *In proceedings of the regional conference on statistical sciences* (pp. 126-38).

Shi, D., DiStefano, C., Zheng, X., Liu, R., & Jiang, Z. (2021). Fitting Latent Growth Models with Small Sample Sizes and Non-normal Missing Data. *International journal of behavioral development*, *45(2),* 179–192. https://doi.org/10.1177/0165025420979365

Singh, A. K., Singh, S., Singh, A. P., & Srivastava, A. (2013). Hope and well-being among students of professional courses. *Indian Journal of Community Psychology*, *9(1),* 109 -119.

Sperrazza, N. (2022). *The Genetic, Environmental, and Sociocultural Factors in Eating Disorder Relapse, and Opportunities for Prevention* (Doctoral dissertation).

Steinglass, J. E., Attia, E., Glasofer, D. R., Wang, Y., Ruggiero, J., Walsh, B. T., & Thomas, J. G. (2022). Optimizing relapse prevention and changing habits (REACH+) in anorexia nervosa. *International Journal of Eating Disorders*. https://doi.org/10.1002/eat.23724

Stichting Benchmark GGZ. (2013). *SBG Dataprotocol 20130101 Bijlage A Minimale Dataset*. Bilthoven.

Streatfeild, J., Hickson, J., Austin, S. B., Hutcheson, R., Kandel, J. S., Lampert, J. G., … Pezzullo, L. (2021). Social and economic cost of eating disorders in the United States: Evidence to inform policy action. *International Journal of Eating Disorders*, *54(5),* 851–868. https://doi.org/10.1002/eat.23486

Swift, J. K., & Greenberg, R. P. (2014). A treatment by disorder meta-analysis of dropout from psychotherapy. *Journal of Psychotherapy Integration*, *24,* 193–207. https://doi.org/10.1037/a0037512

Taber, K. S. (2018). The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in science education*, 48, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Termorshuizen, J. D., Watson, H. J., Thornton, L. M., Borg, S., Flatt, R. E., MacDermod, C. M., ... & Bulik, C. M. (2020). Early impact of COVID-19 on individuals with self reported eating disorders: A survey of~ 1,000 individuals in the United States and the Netherlands. *International Journal of Eating Disorders*, *53(11),* 1780-1790. https://doi.org/10.1002/eat.23353

Treasure, J., Claudino, A. M., & Zucker, N. (2010). Eating disorders. *Lancet (London, England),* 375(9714), 583–593. https://doi.org/10.1016/S0140-6736(09)61748-7

Tipton, E., Hallberg, K., Hedges, L. V., & Chan, W. (2017). Implications of small samples for generalization: Adjustments and rules of thumb. *Evaluation review, 41(5),* 472-505. https://doi.org/10.1177/0193841X16655665

UK, N. G. A. (2017). Eating disorders: recognition and treatment.

Wardenaar, K. (2020). Latent Class Growth Analysis and Growth Mixture Modeling using R: A tutorial for two R-packages and a comparison with Mplus.

Weijers, A., Rasing, S., Creemers, D., Vermulst, A., Schellekens, A. F., & Westerhof, G. J. (2022). Deconstructing recovery: A prospective study on well-being, symptom severity and acceptance in patients with major depressive disorders. *Journal of Affective Disorders*, *296*, 653-659. https://doi.org/10.1016/j.jad.2021.09.048

Westergaard, K., Skovgaard, L., Magyari, M., & Kristiansen, M. (2022). Patient perspectives on patient-reported outcomes in multiple sclerosis treatment trajectories: A qualitative study of why, what, and how?. *Multiple Sclerosis and Related Disorders*, *58*, 103475. https://doi.org/10.1016/j.msard.2021.103475

Westerhof, G. J., & Keyes, C. L. (2010). Mental illness and mental health: The two continua model across the lifespan. *Journal of adult development, 17,* 110-119. https://doi.org/10.1007/s10804-009-9082-y

White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine, 29(28),* 2920-2931. https://doi.org/10.1002/sim.3944

# Appendix

Appendix 1. *Q-Q plots of the five measurements chronologically Glob_EDEQ.1 to Glob_EDEQ.5.*