A Systematic Review of Anorexia Nervosa, Bulimia Nervosa, and Avoidant Restrictive Food Intake Disorder Research:

Taking A Behavioural Approach

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Applied Psychology – Applied Behaviour Analysis
at
The University of Waikato
by
ALYSSA ERASMUS

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Abstract

Restrictive eating disorders (EDs) are a category of disorders, defined broadly here by including any range of repeated behaviours which result in an individual significantly restricting their nutritional intake. Although the rate of diagnosed eating disorders in the general population is relatively low (Demmler et al., 2020), restrictive EDs have the highest mortality rate of any other mental disorder (Fichter & Quadflieg, 2016, Smirk et al., 2012, Steinhausen, 2002), and while EDs typically develop during the transition to adulthood, they can persist into adulthood if they go untreated (Potterton et al., 2020). The purpose of this study was to conduct a systematic review of this literature using PRISMA guidelines, on the treatment of anorexia nervosa, bulimia nervosa, and avoidant restrictive food intake disorder using three established psychological interventions (cognitive behavioural therapy, acceptance and commitment therapy and behavioural therapy).

This thesis reconceptualised understandings of restrictive EDs in behavioural analytical terms by applying a functional, individual-focused lens to evaluating research in this space. Eight articles published since 2013 were identified for analysis by the search. The methodology, scientific rigor and social validity of research in this space was considered in analysis, and this thesis applied the ABA lens to understanding where and how treatments may be failing to produce socially significant, lasting results.

Preface and Acknowledgments

I would firstly like to thank my supervisor, Dr Angelika Anderson for sharing her knowledge and time with me throughout the process of writing this thesis, my work would not be what it is without your support and guidance, thank you.

To my parents, Natalie and Bruce, for supporting my being a career-student for the past six years, I cannot thank you enough. To my partner Connor, thank you for taking on everything you have to carry us while I finish this. I could not have done this without all of you pushing me forward, and I will always appreciate that more deeply than you know.

Table of Contents

Abbreviations Shortlist

ABA: Applied Behaviour Analysis

AN: Anorexia Nervosa

ARFID: Avoidant Restrictive Food Intake Disorder

BITE: the Bulimic Investigatory Test, Edinburgh

BMI: Body Mass Index

BN: Bulimia Nervosa

BSI: Brief Symptom Inventory

BT: Behavioural Therapy

CBT-E: Enhanced Cognitive Behavioural Therapy

CBT: Cognitive Behavioural Therapy

CCM: Catastrophic Causal Misinterpretation

CDI-2: Child Depression Inventory 2

CGI: Clinical Global Impression Scale

CIA: Clinical Impairment Assessment

CSQ: Client Satisfaction Questionnaire

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

ED: Eating Disorder

EDE 16D: Eating Disorder Examination Edition 16.0D

EDE-Q: the Eating Disorders Examination Questionnaire

EOT: End of Treatment

FF-TAM: The Technology Acceptance Model Instrument-Fast Form)

FNS: Food Neophobia Scale

GAD-7: Generalized Anxiety Disorder scale; (7 point)

IOA: Inter-observer agreement

PARDI-AR-Q: Nine Item ARFID Screen, Pica, ARFID, and Rumination Disorder Interview-Questionnaire

PARDI: Pica, ARFID, and Rumination Disorder Interview

PHQ-9: Patient Health Questionnaire (9 item)

SCED: Single-case Experimental Design

STAIC: State-Trait Anxiety Inventory – Child version trait subscale

TAU: Treatment as usual

VAS: Visual Analog Scale

Restrictive eating disorders (EDs) are a category of DSM-5 (Diagnostic and Statistical Manual of Mental Disorders: 5th edition) (American Psychiatric Association, 2013) disorders, defined broadly here by including any range of repeated behaviours which result in an individual significantly restricting their nutritional intake. This can take a range of topographies, understood here functionally as an individual not consuming a typically accepted amount of food or range of nutrients, to consuming enough or too much, but later purging, thus functionally restricting nutritional intake to a point at which individual nutritional needs are not met. Although the rate of diagnosed eating disorders in the general population is relatively low (Demmler et al., 2020), restrictive EDs have the highest mortality rate of any other mental disorder (Fichter & Quadflieg, 2016, Smirk et al., 2012, Steinhausen, 2002), and while EDs typically develop during the transition to adulthood they can become lifetime persistent well into adulthood if they go untreated (Potterton et al., 2020).

Psychological treatments have broadly taken the approach that, although multifactorial, EDs are some external manifestation of internal cognitions about one's self and one's body, as evidenced by their respective DSM-5 diagnostic criteria (American Psychiatric Association, 2013). This thesis initially aimed to reconceptualise understanding of the development and maintenance of restrictive EDs in behavioural analytical terms by applying a functional, individual-focused lens to evaluating treatment as research. This functional definition allowed for the investigation of treatment-based studies of three major ED (Anorexia Nervosa (AN), Bulimia Nervosa (BN) and Avoidant Restrictive Food Intake Disorder (ARFID)) using a systematic search which followed PRISMA guidelines. It then applied the ABA (Applied Behaviour Analysis) lens to understanding where and how treatments may be failing to produce socially significant, lasting results.

Chapter One: Behaviourism as an Approach

This chapter serves to illustrate the lens that this thesis takes to understanding the topic and analyse the data through. It briefly outlines Behaviourism as an approach to understanding individual behaviour, explains the rationale for single-case experimental designs, and highlights the importance of objective outcome measures to evaluate and inform effective interventions.

Behaviourism

Behaviourism is the philosophical underpinning of the science of behaviour, and is rooted in seven core principles: empiricism, pragmatism, parsimony, experimentation, observable measurement, environmental determinism, and a rejection of mentalisms.

Baum's (2017) work succinctly explains that these core principles can be understood as such: Pragmatism emphasises that the power of scientific inquiry lies in the ability to make sense of experiences and phenomena, while parsimony encourages scientific inquiries to favour the most straightforward explanations for these experiences. Experimentation allows for the systematic study of behaviour and observable measurement ensures that conclusions are drawn only from variables which can be held constant. Environmental determinism works to facilitate the rejection of mentalisms and hypothetical constructs as an explanation for causes of behaviour, as it states that only the observable environment can be studied and manipulated, providing causal explanations for behaviour.

Environmental determinism is a key factor in understanding what a behaviourist might attribute as the cause of a given behaviour. Environmental explanations of behaviour, at a very basic level, posit that the selection of a response (behaviour) is determined by stimuli within an organism's environment, and are informed by that individual's history of consequences in a similar environment. This history of repeated response-consequence contingencies constitutes what is referred to throughout this thesis as learned behaviour. The understanding that all behaviour is learned is a key tenant of behaviourism. The behaviours which arise as a product of the environment are operant responses – first identified by Skinner (1938) – and are actions performed voluntarily as a result of some stimuli in the individual's environment. Respondent behaviour, on the other hand, is behaviour that is reflexive or involuntary, although it is also elicited by some specific stimulus in the individual's environment. Salivation in the presence of food, for example, may be reflexive, while eating is voluntary. Importantly, not only does the environment here include the world around the individual, but also physiological and biological occurrences within the individual. For a science of behaviour, physiological occurrences usually constitute private events (events which only the individual has access to) however the world around the individual is what can be measured empirically. For example, eating is a public behaviour, observable and objective, cognitions and biological processes about eating, such as appetite and hunger, are private events.

It follows from this why behaviourism rejects hypothetical constructs as explanations for behaviour. The view is that terms which do not reflect objectively measurable phenomena or attempt to group measurable behaviours with inaccessible explanations for behaviour cannot be studied in an empirical manner. Diagnostic labels, for instance, are a hypothetical construct which do exactly this. They are inherently inconsistent with the science of behaviour in that adherence to these labels in understanding behaviour and its maintenance can lead to employing circular reasoning.

Circular Reasoning in Food Aversion

Food aversion, for example, illustrates how using a hypothetical construct as opposed to objective variables to explain behaviour might impede interventions from creating positive behavioural change. Consider this example:

X does not eat enough. X is diagnosed with Anorexia Nervosa because they do not eat enough. X does not eat enough because they have Anorexia Nervosa.

This reasoning is circular; X does not eat enough so they have AN, X has AN so they do not eat enough. This tells an observer nothing about how one might move forward with getting X to start eating enough, because AN is a hypothetical construct being used to explain X's insufficient rate or frequency of eating behaviour. Changing what a person does here relies on hypothetical constructs being employed in treatment, an approach which is internally fallacious as still, there is not objective explanation for X's behaviour. A behavioural approach, however, would be focused on identifying measurable and alterable variables in X's environment, such as a punishing response-consequence contingency in X's history which has resulted in the observed deficiency of eating behaviours, or identifying that eating is not rewarding to X and therefore does not act to increase the rate of behaviour. A behaviourist might be interested in identifying if X's behaviour has generalised to not eating/eating very little amounts of certain foods or all food. They may also be interested in finding out if X will eat more at school but not at home (i,e, if behaviour is consistent across different physical and social environments). They might be interested in finding out if X will eat more if the form that food is presented in is changed, or if X will eat more when immediately reinforced with a reward. The behaviourist is identifying objective variables in X's environment and manipulating them to understand and change behaviour, as opposed to appealing to some diagnostic label of X's and trying to manipulate that. These investigations would be carried out by measuring X's responses in controlled environments and drawing conclusions based only on these findings.

This is not to say that diagnostic labels and terms do not have a practical place in science and treatment. They serve as a way for people to communicate something about a person and their behaviour quickly, without the need to define everything that person may or may not do. They serve a purpose in defining where and how people might receive treatment or funding for their behaviour and support, and they signal to practitioners what approach to treatment may be most effective before an intervention commences. However, this explanation serves to delineate the key distinction of a behavioural lens when considering, diagnosing, and altering behaviour.

On this, behaviourism tends to favour the single-case experimental research design (SCED) as BA is interested in the behaviour of *individuals*, not groups, and on measuring *observable* behaviour. As in the example above, a range of individual variables are considered in the approach to measuring and changing X's behaviour. As per the What Works Clearinghouse (2020) (WWC) single-case experimental design documentation, repeated measures of the case before and during the intervention, as well as the individual case serving as its own control are the key design tactics of SCED. Here, the subject is only ever being compared to themselves, as opposed to some other group of people, as in a between-subjects design. This is consistent with the behavioural lens in that if changing an individual's behaviour is the goal of a treatment, an individual's behaviour is what should be assessed and serve as a control to measure change. Primarily, SCEDs allow researchers to experimentally evaluate the efficacy of a treatment on behaviour directly. To use other individuals with entirely different learning histories and environments as controls may appear to allow a study to make broader claims about the effects of a treatment to varied populations, this may fail establish whether an intervention is effective for the individual across conditions and time. It also means that treatment cannot be individualised to the extent that it can be in a SCED, meaning that therapeutic effects might not be achieved as efficiently – a concern for interventions aiming to treat behaviour which is harmful or life threatening (such as restrictive EDs).

A defining characteristic of ABA relevant here is *social validity*. As Cooper, Heron & Heward (2020) assert, principals of behaviourism are applied in practice to improve socially significant behaviour, with systematic experimentation being applied to identify variables which improve behaviour. First identified by Wolf (1978), social validity provides a framework for assessing the degree to which an intervention is acceptable to the participant or those around them. It is predicated on the idea that nonacceptance of ABA programmes by consumers leads to low treatment adherence and therefore, unfavourable treatment outcomes. Wolf (1978) argues that treatment acceptability is an important part of establishing an effective intervention in ABA. Fawcett (1991) informs the following descriptions of the three key areas of social validity that should be established when developing interventions. Firstly, *social significance of goals* relates to the relevance of the goals targeted by an intervention and ensures that goals of interventions are valuable to both the participant and society. *Social appropriateness of procedures* pertains to the appropriateness of methodology used in treatment and research, ensuring that the design is respectful to the rights of the participant. Finally, *social importance of effects* is concerned with establishing that outcomes produced by treatment have meaningful and lasting effects on the participant, and work towards positive social functioning or quality of life. Thus, it is imperative that practitioners aiming to change behaviour take steps towards establishing social validity across all of these domains to ensure treatment acceptance. In order to establish social validity and work together with clients to assess the extent to which is has been established is not necessarily as simple as asking consumers of ABA programmes how they feel about the intervention. Hawkins (1991) argues that using reports of this nature measure verbal behaviour, and that the term social validity might not accurately capture what is meant here, as validity here is not solely related to social factors. Rather, they argue that the terms *consumer satisfaction* (of ABA programmes) and establishing *habilitative value* more accurately surmises what is being identified here. Habilitative value is defined here as “the extent that a consumer’s reported satisfaction predicts the benefits and costs delivered by the natural environment as a result of the particular goal set, outcomes achieved or procedures used” (Hawkins, 1991, pg. 206). This definition asserts that the appropriateness of interventions can be defined in terms of how they work to increase access to reinforcement in a naturalistic setting, and the author goes on to argue that the extent to which consumer satisfaction with treatment programmes can be evaluated relies on establishing habilitative validity.

Broadly, these concepts can be understood as the ways in which researchers and practitioners conducting interventions should consider the appropriateness of treatments in the context of the individual they are treating. Ensuring that social validity and, specifically the habilitative validity criterion are accounted for in this process works towards ensuring consumer buy-in and treatment adherence. This may be best achieved through the use of SCEDs, as this design allows for greater individualisation of treatment. This can (and should) be considered in cohort designs, however due to the nature of generalised goals and effects specifically being applied to a range of fundamentally different people in these designs, this may not be possible here to the same extent as in SCEDs.

Taken together, this works towards developing one of the underlying arguments of this thesis: that SCED rationale should be utilised in the treatment of restrictive EDs due to the highly variable nature of ED related behaviours, and the ability for the practitioner to establish an effective intervention which accounts for social validity and the habilitative value criterion, thus increasing the likelihood of treatment adherence.

Chapter Two: Normal and Abnormal Eating Behaviours

In order to establish what might constitute abnormal eating behaviours, a range of biological, social, cultural and developmental factors need to be considered. What might constitute an acceptable nutritional or caloric intake is highly dependent on both the social and cultural circumstances of the individual, and within this context, biological necessity as a result of variable individual energy expenditure and specific nutritional restrictions must be addressed. This chapter serves to briefly illustrate the key physiological and psychosocial factors of appetite, malnutrition and eating as a behaviour, and establishes that what constitutes normal eating should be considered at the individual level as opposed to taking a "one size fits" all approach.

Physiological Considerations

Malnutrition

There is no one universally agreed upon definition of malnutrition. For example, Marions (2017) presents fifteen different, accepted definitions of the term. These range from "malnutrition is the condition that occurs when your body does not get enough nutrients" (US National Library of Medicine, n.d, para. 1) to "a term used to refer to any condition in which the body does not receive enough nutrients for proper function" (MedlinePlus, n.d, para. 1). The author notes that many of the definitions conflate malnutrition with undernutrition, which is what is commonly observed in patients with EDs, however this is not consistent across all patients with EDs, and to define which term is correct here is outside of the scope of this thesis. This thesis takes the view that malnutrition is a state in which an individual is not receiving adequate nutrition to achieve normal bodily function (outside of another other health condition), and is state which is significantly harmful to the individual.

Effects of malnutrition which impair normal bodily function include an impaired immune response, increased fatigue as a result of loss of muscle tissues, loss of respiratory tissue leading to poor cough pressure and poor recovery from chest infections as well as general healing being impaired (Turner, Alison, & Nightingale, 2023). Functionally, a malnourished person will see a decrease in leptin, which acts as a signal for the size of body fat reserves. Leptin, secreted in adipose tissue, is sensitive to nutritional deprivation, and in women, low levels of leptin result in cessation of ovulation and menstruation, as commonly reported in women with AN (Chou, & Mantzoros, 2014).

As noted in the working definition of malnutrition above, malnutrition does not only occur as a result of a failure to eat enough or enough of a variety of nutrients. Cachexia, for example, is a state of malnutrition which occurs as a result of chronic diseases, such as cancer or heart and renal failure. This form of malnutrition does not respond to nutritional supplementation and is not caused by failure to consume adequate caloric intake. Malnutrition as a result of conditions such as cachexia is distinct from that associated with restrictive EDs in this context. Importantly, the argument of this thesis does not attempt to stipulate an ideal or insufficient calorie intake. Calories are a measurement of energy intake. While inadequate caloric intake can result in malnutrition, excess caloric intake is associated with increased weight gain as well as the development of insulin resistance and oxidative stress (Boden et al., 2015), however, variance in individual activity levels, health conditions and thus, nutritional needs, cannot be overstated here either. Body weight can be maintained by balancing energy consumption with energy expenditure; however, certain diseases can make this extremely difficult, and as established above, an individual’s diet and lifestyle alone are not sufficient explanation for body weight in many cases. As I have thus far made clear, there is no one right way to eat, and there is no one right amount to consume, and to reduce dietary needs to only one arbitrary metric is misguided. (For more on this, see: Hargrove, 2006). Personalised dietary interventions can be developed to treat the effects of malnutrition by considering specific factors such as the metabolic profile, background, and environment of an individual, but the range of variables associated with establishing what factors constitutes ‘adequate nutrition” is not within the scope of this thesis. Adequate nutrition must be defined by the induvial needs at a specific moment in time by a qualified medical professional, and should consider the ability of an individual to consume sufficient calories, hydration and varied nutrients to maintain nutritional needs. Again, this thesis considers malnutrition to be a state in which an individual is not receiving adequate nutrition to achieve normal bodily function (outside of another other health condition), and the behavioural lens allows this to be understood as a consequence of the rate, frequency and variance of which individual eating behaviours do or do not occur. This state is, importantly, considered harmful to the participant across both physiological and psychosocial dimensions of health.

As noted, the disorders of interest in this study (AN, BN and ARFID) are all functionally restrictive, meaning that allowing restrictive behaviours to endure can lead one to become malnourished. This chapter served to briefly illustrate that malnutrition is a life-threatening condition at worst, and a quality-of-life-threatening condition at best.

Appetite as a Consequence of the Socioeconomic Environment

It is important to note that the failure to consume a diet that satisfies an individual’s metabolic needs sufficiently may not be as the result of an underlying behavioural, medical, or psychological disorder, but rather is a result of one’s circumstances. Access to a varied diet is often a relatively privileged experience. Manufactured scarcity such as food desserts, droughts, wars and floods all effect food availability. Eating disorders, specifically those which are restrictive, are only considered here in those who have access to sufficient food but are still found to be malnourished. The DSM-5 explicitly states that individuals who present with physiological symptoms of an ED (usually being significantly underweight) should not be diagnosed if their condition could be better explained by a lack of access to sufficient nutritional intake (American Psychiatric Association, 2013). As such, only studies which used the DSM-5 criteria for diagnosis of patients are included in this review moving forward.

Hunger and Appetite

At a physiological level, hunger and appetite are a complex interaction between environmental stimuli, sensory receptors, the primary sensory cortices of the brain and the processing of hormones secreted in the digestive system. At a very basic level, food functions as an evocative stimulus, and the eyes, ears, nose and oral cavity work to process these stimuli. Hunger (and satiety) cues occur in the hypothalamus, and ghrelin, a fast-acting hormone secreted in the stomach, acts to increase appetite (Gallon et al., 2022). Nutritional intake, as a source of energy for mediating homeostasis, accounting for physical expenditure, and maintaining mental functioning is a biological impetrative for all organisms. In humans, only about one third of energy expenditure occurs through voluntary movement, while the rest is used for maintaining normal physiological functioning. The brain and nervous system generally comprise 2% of body weight but consume 25% of resting energy expenditure (Magistretti & Allaman, 2022).

Consummatory behaviour is both rewarding in and of itself due to it being required to maintain homeostasis, and as a result of reinforcement contingencies. Appetite control centres, found in the amygdala, control learnt food behaviour (ie. if something is food or not), while the nucleus acumbens - a part of the reward centre of the brain - is concerned with the pleasure of eating and rewards from food. Pleasure plays no small role in the development of eating behaviour and preferences. Marty et al. (2018) details three dimensions of pleasure derived from eating; pleasure from sensory sensations, pleasure from the social context of food, and pleasure from cognitive representations of food. They explain that understanding the pleasure and rewards of consuming foods can inform effective dietary interventions from a very early age. For example, there is evidence that newborns and infants show a preference for sweet foods, as the high glucose content offers high food energy density vital to infant development (Schwartz et al., 2009). Similarly, salt has been shown to activate reward centres in the brain (Morris et al., 2008).

Importantly, appetite control centres also have connections to the cortex and other higher brain centres, which means psychological factors may “override” physiological control of appetite. Appetite centres respond to different patterns of metabolic fuels in the bloodstream as well as hormones such as insulin and, glucagon, and as Suzuki et al. (2012) concur, “environmental cues, cognitive, reward, and emotional factors play an important role in food intake which may override homeostatic requirements” (pg. 9) Functionally, this means that the development of abnormal eating behaviours, such as those commonly associated with restrictive EDs, may involve a complex interaction between higher brain centres, hunger and satiety signals, hormone regulation, and environment, suggesting that psychological factors alone are not a sufficient explanation for eating behaviours.

Appetite as a Social Phenomenon

Although it has been established that appetite can be understood at the physiological level, hunger and satiety are not purely a biological phenomenon. Psycho-social influences on appetite are largely determined by a range of psychological and social factors. Meals serve important social functions for humans. As Papies et al. (2022) found, situational cues affect our desire for food, and underlying factors play an important role in food desire. For example, people who eat alone to have less stimulus to appetite, explaining the loss of appetite and reduced eating behaviour observed in elderly people (Mikami et al., 2022). Nutritional problems among older adults are well established (Törma et al., 2013, Mamhidir et al., 2006), with older women being more likely to be undernourished (Strube-Lahmann et al., 2021) and depressed (Velázquez-Alva et al., 2020). This is not surprising, as older adults often live alone, and food sharing is an important non-verbal indicator of positive relationships and friendship (Miller et al., 1998). As such, food offering can be a means to increase positive affect in both the recipient and in the provider (Hamburg et al., 2014), an idea further supported by Woolley and Fishbach (2017) who showed that people who ate the same food felt closer to each other and were more trusting. Moreover, food sharing can change food cue reactivity, and Shienle and Zorjan (2022) found that the mental imagery of receiving a social reward from food sharing alone produces this effect.

In social groups too, food plays an important role. The perceived value of a specific food can change based on the perception that an in-group does not enjoy the food (Robinson and Higgs, 2012), and there is evidence that children may consume more when in larger groups (Lumeng, & Hillman, 2007). Early varied exposure to food, as well, usually mediated by a parent or caregiver, has been shown to lead to increased preference and increased acceptance of foods (Birch and Marlin, 1982), demonstrating that there is a learning history effect associated with food preference. The exposure to food removes the novelty of foods, and it has been established that initial contact with new foods that does not result in negative consequences means that those foods are more likely to be preferred in future (Kalat & Rozin, 1973).

This, along with the evidence that restrictive behaviours can develop very early in childhood (Pinhas et al., 2011), and that EDs can become lifetime persistent, suggests that the social environment of a child can have a lasting effect on consummatory behaviour. Importantly for the lens of this thesis, social modelling can affect the amount of food an individual consumes. A meta-analysis by Vartanian et al. (2015) that included 38 studies revealed that participants ate more when a model ate more and ate less when a model ate less. Interestingly for the heavily gendered prevalence of EDS, the social modelling effects were greater in female participants.

It follows then, why patients with restrictive EDs might exhibit comorbid symptoms of low-mood disorders such as depression and anxiety. While these may be causal in some cases, the effect of being isolated from key social environments – mealtimes – cannot be overstated. Importantly, this also suggests that how and what people eat is largely determined by their social environment, a variable which as established, aligns with the behavioural lens, in that an individual's environment can explain and maintain their behaviour.

The interaction between social and eating behaviours is significant. How people eat, what they eat, and importantly, how much they eat, is determined significantly by their social environment and individual learning history. This complex relationship may appear to be a "life-sentence", specifically in regards to the treatment of adults with maladaptive behavioural relationships with food, but importantly, these are factors which can be manipulated. Take, for example, body image concerns. Widely accepted as a risk factor in the development and maintenance of restrictive EDs, body dissatisfaction and extreme weight control behaviours are considered to be the strongest contributing risk factors to the development of ED behaviours (Maloney et al., 1989; Munkholm et al., 2016; Ricciardelli & McCabe, 2001; Ricciardelli et al., 2003 and Stice & Whitenton, 2002). However, this is highly susceptible to change given the specific culture, social class or group, and even time period in which an individual exists. Considering the range of emotional and psychological phenomena inherent in the development of eating behaviours is important and should be considered, but I posit that restrictive diets and disordered eating should primarily be categorised functionally by the restriction of adequate nutritional intake. Internal causes/experiences of EDs may be accessible to the individual, but due to how varied individual experiences and aetiology of disordered eating can be, the best approach might not be to focus solely on these inner causes for maladaptive eating behaviours.

Attempting to tease apart the psychological and the societal factors which may contribute to an individual developing body image concerns is not within the scope of this study. Rather, this is raised to establish that how one eats is not purely determined by internal cognitions – but rather by a range of malleable factors, such one's social environment. This can, and should, be considered in treatment.

Chapter Three: Anorexia Nervosa, Bulimia Nervosa and Avoidant Restrictive Food Intake Disorder

The development of EDs is often observed in late childhood, with body dissatisfaction, a key consideration in the clinical diagnosis of EDs, being reported in 50% of children aged 6–12 years (Dion et al., 2016; Tatangelo et al., 2016). Both girls and boys have been shown to be significantly invested in a desire for thinness and muscular ideals by between 5–9 years old (Dohnt & Tiggemann, 2006; McLean, Wertheim, & Paxton, 2018; Ricciardelli, McCabe, Mussap, & Holt, 2009; Slater & Tiggemann, 2016), and McVey, Tweed, and Blackmore (2004) found that 29% of girls aged 10–14 years had reported attempting to lose weight. Although not usually concerned with body dissatisfaction or a desire for thinness, ARFID also onsets during childhood, generally much earlier than AN or BN. EDs often resolve in late adolescence, but can continue into adulthood if unsuccessfully treated or entirely untreated. High mortality rates (relative to other mental disorders) may explain why fewer adults are diagnosed with EDs, however there may also be little incentive for adults to seek help (i.e. being outside of the parental home/not having caregivers who encourage seeking care), thus leading to underdiagnosis.

Anorexia Nervosa

AN typically onsets in late adolescence (Carter, Kelly and Norwood, 2012), and can become lifetime persistent if untreated (Speciani et al., 2021). Relapse rates of up to 40% in the first year following treatment have been observed (Croce et al., 2024) with up to 10% of people becoming chronic (Strober, Freeman and Morrell, 1997).

In order to meet DSM-5 criteria for an anorexia nervosa (AN) diagnosis, a patient must present with a refusal to maintain normal body weight (>85% of expected body weight), have an intense fear of gaining weight, have disturbances in the way in which they experience their own body weight or shape, and in postmenarcheal females, amenorrhea (American

Psychiatric Association, 2013). The prevalence of AN in the general population is generally low and higher rates are observed in young women (Van Eeden, Van Hoeken & Hoek, 2021). AN has the highest mortality rate of any mental illness (Birmingham et al., 2005).

AN is typically associated with depression, panic-disorder, obsessive compulsive disorder, anxiety and depression. Clinically, mental anorexia has typically been considered to be not primarily driven by a lack of appetite, but rather by a manifestation of an impulse to be thin.

Bulimia Nervosa

The DSM-5 criteria for a Bulimia Nervosa (BN) diagnosis requires patients to present with; binge and purge behaviours (recurrent episodes of binge eating (topographically defined as eating, within a discrete period, an amount of food larger than most people would consume in a similar time period under similar circumstances) and recurrent compensatory behaviour to maintain or prevent weight-gain such as vomiting, use of laxatives, excessive exercise or abuse of other forms of medication) occurring more than twice a week for three months, as well as reporting self-evaluation to be unduly influenced by body shape. Importantly, these disturbances should not occur during episodes of AN (American Psychiatric Association, 2013). The peak age of incidence of BN ranges between 15 and 29 years, and although not uncommon, a significant decline in the global incidence rates of bulimia nervosa has been observed between 2004 and 2014 in Western countries, with significantly lower rates of incidences observed in males (Van Eeden, Van Hoeken & Hoek, 2021).

Avoidant Restrictive Food Intake Disorder

ARFID presents as an eating or feeding disturbance which results in an ongoing failure to meet appropriate nutritional or energy needs, and according to the DSM-5 diagnostic criteria, must be accompanied by significant weight loss, a dependency on nutritional supplementation, and/or a marked disturbance in psychosocial functioning. The disturbance should not be better explained by lack of available food or cultural practice, and should not occur during the course of AN or BN. Importantly, an ARFID diagnosis should not be given if an individual presents with a disturbance in which their own body weight or shape is experienced, thus drawing the distinction between ARFID and AN or BN. Further, the eating disturbance should not be attributable to a concurrent medical or mental disorder (American Psychiatric Association, 2013). ARFID typically onsets much earlier than AN, and children with ARFID often present with impairments in several executive function domains (Basile et al., 2021). There is only a small body of research on ARFID, possibly as a result of its new status as a distinct disorder, only being introduced with the release of the DSM-5. Previously, young children exhibiting aversion, fear and avoidance to novel foodstuff was referred to as neophobia (Ogden, 2007). There are only 7 studies on the prevalence of ARFID in the broader community, which suggests that accurate prevalence rates are not yet established. As Dinkler et al. (2022) note, a problem with establishing a prevalence rate for ARFID diagnoses in the general population is whether physical consequences of ARFID (such as reduced weight, for example) are necessary for diagnosis. They note that the ambiguous wording of the DSM-5 criteria for diagnosis do not make it clear whether both physical and psychosocial impairments need to be present for a patient to receive a diagnosis (Dinkler et al., 2022).

Taken together, this thesis has chosen to consider these three disorders together for the purpose of this review as the behaviours associated with them all achieve the same function – caloric or nutritional restriction – as evidenced by the diagnostic criteria for each disorder. However, it is crucial to note that there is variance between what restrictive behaviour might look like both between and within diagnoses. A patient receiving a diagnosis of BN, for example, may purge by misusing laxatives, while another might induce vomiting.

Similarly, a patient receiving a diagnosis for AN may significantly restrict intake by simply not eating a sufficient amount, while another may consume slightly more, but compulsively exercise to mitigate this. Similarly, excessive exercise may also be identified in people presenting with BN as a form of purging. Transition between AN and BN over the course of an individual's life is also possible and speaks not only to the overlap in diagnostic criteria inherent in these diagnoses but to how behaviour can change in different spatial, social and temporal environments. This highly variable individual difference, even between people with the same diagnoses, only further demonstrates why cohort studies may not be solely appropriate for assessing treatment effects of ED treatments, specifically if they do not aim to change *specific* problem behaviours. As established above, single-case designs may be more appropriate as they allow intervention to change individual behaviour as establish individualised social validity of effects.

Treatment for AN, BN and ARFID

Although EDs typically receive clinical or outpatient care, hospitalisation may be required for severe cases, with rates of hospitalisation increasing in Western countries, specifically during the COVID-19 pandemic (see: Gilsbach et al., 2022, Hansen, Stephan & Menkes, 2021; Rikani et al., 2013,). As this is a narrow review, this limits the scope of the review to three established psychological treatments of eating disorders – Cognitive Behavioural Therapy (CBT), Acceptance and Commitment Therapy (ACT) and Behavioural Therapy. Behavioural Therapy (BT) is defined here more broadly so as not to exclude the types of treatments this study is primarily interested in reviewing. All of these treatments have roots in behaviourism, but only ABA takes a purely conceptually consistent approach. As this review primarily aimed to discuss the scientific rigor of treatment-based studies in the space as established by the core philosophy of behaviourism, it is important to note what this thesis did not aim to accomplish. It is not the purpose of this thesis, to establish which treatment is most effective for each diagnosis. Similarly, this thesis does not make claims about which approach should be used for each diagnosis. A brief understanding of ABA, CBT and ACT in practice is detailed for the sake of clarity, but no one approach is favoured here.

Cognitive Behavioural Therapy (CBT) is one of the most widely used therapies in the treatment of eating disorders, with Enhanced CBT (CBT-E) emerging as a promising treatment for both adolescent and adult AN in both outpatient and inpatient populations. The cognitive behavioural viewpoint of AN is that symptoms are primarily maintained by concerns about body weight and shape, and that once established, these beliefs cause the individual to act in accordance with these beliefs by engaging in specific eating and purging behaviours (Dalle Grave et al., 2016). The model of CBT-E has been extended by Fairburn (2008) to include the treatment of BN. The CBT model of BN posits that dieting begins as a product of body dissatisfaction, primarily related to weight or shape, and binging occurs as a result of the deprivation of dieting, leading to purging to control weight gain, thus further reinforcing a cycle which maintains BN. CBT based interventions usually start by addressing harmful behavioural symptomology, and then move to addressing the cognitive distortions considered to be maintaining restrictive behaviour (see Hagan & Walsh, 2021). In conceptualising ARFID, CBT posits that the drivers of restrictive eating behaviour underlie ARFID symptomology are avoidance based on sensory characteristics of food, lack of apatite, and/or concerns about aversive consequences of eating. It is understood that treatment should be multi-disciplinary (Bryant-Waugh, et al., 2021) and focus on exposure to novel foods (Kambanis et al., 2023).

The core approach and conceptual framework for ABA in practice has been well established in Chapter One. However, ACT interventions also have firm roots in behaviourism. ACT's emphasis on acceptance and mindfulness in reducing harmful behaviours and has been shown to be effective in both single-case (Heffner et al., 2002) and cohort studies (Juarascio et al., 2013) on reducing AN symptoms. It has also been shown to reduce symptom severity in patients with both AN and BN (Juarascio et al., 2013).

Approaches to Measuring Symptoms of AN, BN and ARFID in Treatment

A range of approaches are currently used to measure and assess ED symptoms. Although discussed conceptually, the key tools identified as relevant to this review are outlined very briefly here for posterity.

Self-Report Tools

In this context, self-report tools act to assess symptomology based on asking a patient or caregiver structured questions and generating a score or impression of overall symptoms. Closed question self-report tools, such as the EDE-Q (the Eating Disorders Examination Questionnaire) might be used solely or alongside a structured interview tool such as the PARDI-AR-Q (Nine Item ARFID Screen, Pica, ARFID, and Rumination Disorder Interview-Questionnaire). More structured interview tools, such as the PARDI (Pica, ARFID, and Rumination Disorder Interview) and the derivative nine item PARDI-AR-Q (Pica, ARFID, and Rumination Disorder Interview- ARFID Questionnaire), are ARFID specific psychopathology assessment tools (Bryant-Waugh et al., 2022), specifically concerned with treatment as assessment of ARFID symptoms. Structured interview tools of this kind can be used for both assessment of severity at baseline and EOT (end of treatment), and as a means to form diagnoses, and generally allow for more open ended and individualised responses to questions. In assessment and treatment of ARFID, the Food Neophobia Scale (FNS) (Pliner & Hobden, 1992) identifies key factors in what is now considered to be food avoidance. The FNS uses a total of ten statements (both positively and negatively worded items) which are measured on a 7-point scale, from strongly disagree to strongly agree. This tool, although applicable to ARFID, is also often used for market research in the food industry.

Eating disorder specific, closed question tools include the BITE (Bulimic Investigatory Test, Edinburgh), a 33-item scale which measures both symptoms and severity of BN (Henrson et al., 1987). It has been shown to have good sensitivity for identifying cases of BN, but can falsely identify participants with AN, specifically AN binging-purging type well (Orlandi, Mannucci, & Cuzzolaro, 2005). More broadly, the EDE-Q (Eating Disorders Examination Questionnaire) and the derivative EDE-16D (Eating Disorder Examination Edition 16.0D) (Fairburn & Beglin, 1994) assess ED symptomology across four sub-scales related to cognitive factors associated with eating disorders. These are identified by Berg, Peterson, Frazier and Crow (2012) as "restraint, eating concern, shape concern and weight concern" (pg. 428). Upon evaluating the psychometric merits of the scales, they found that the instrument is reliable specifically in identifying cases and non-cases of BN. Mond et al., (2008) found that a EDE-Q score of 2.80 or higher identified possible eating disorder cases with reliability of 80%. This suggested that the EDE-Q can be used as a somewhat reliable brief measure. As with the BITE, studies suggest that brief screening tools can be effective in detecting cases versus non cases eating disorders (Anstine & Grinenko, 2000; Cotton et al., 2003; Johnston et al., 2007).

The literature in closed-question tools generally supports their used in this field, and although both open and closed question tools may perform similarly with respect to self-reported behavioural symptoms, they can provide conflicting assessments of severity and complex features of the disorder, if they do not fail to identify the case at all (Fairburn, & Beglin, 1994; House et al., 2008).

A large proportion of the tools relevant for explanation here, however, aim to measure comorbid symptoms of psychological distress, such as the The Children's Depression Inventory - 2 (CDI-2). This is a 12 items self-report assessment tool which assesses cognitive, affective and behavioural symptoms of depression in children and adolescents aged 7–17 years. It is concerned with feelings over the past two weeks originally developed by (Kovacs, 2004 as cited in Houghton et al., 2022). Houghton (2022) conducted a Rasch analysis on the CDI 2 and found good reliability, however, note that sole reliance on self-report as a measure of psychological distress presents problems due to bias and poor recall. Similarly, the PHQ-9 (Nine Item Patient Health Questionnaire) used to screen for depression in psychiatric setting, th**e** Brief Symptom Inventory, a 53 or 18 item questionnaire tool which assesses psychological distress (Franke et al., 2017), and the STAIC (State-Trait Anxiety Inventory – Child version trait subscale), which measures trait anxiety in children (Spielberger, 1973) are also relevant here. These tools, as noted, do not measure ED specific symptoms. Although evaluating psychological distress broadly is very important in the treatment of any psychological disorder, evaluating these tools individually is not within the field of interest of this study as they do not pertain directly to the behavioural symptoms identified by the work's functional definition of restrictive EDs. They also all measure hypothetical constructs, such as anxiety or depression, which again, is inconsistent with the lens of this thesis.

Tools which aim to measure dimensions of treatment acceptance include measures such as the CSQ (Client Satisfaction Questionnaire) (Larsen et al, 1979) and the FF-TAM (The Technology Acceptance Model Instrument-Fast Form) (Chin, Johnson, & Schwarz, 2008), which is used to assess attitude towards acceptability of technology. Tools of this nature aim to quantify how participant feel about an intervention or psychotherapy, or how they respond to the method in which it is delivered. The use of these types of tools move toward a more comprehensive intervention as they give the practitioner some idea about how treatment and procedural acceptability. In measuring how ED symptoms effect subjective ratings of overall functioning, the CIA (Clinical Impairment Assessment Questionnaire) is relevant here. This is a 16-item self-report instrument which was designed to measure secondary impairment as a result of EDs across three categories – personal, cognitive and social (Bohn et al., 2008). The CIA is often used to establish clinical norms, and has been found to have good criterion validity, specifically for women (Bohn et al., 2008; Vannucci et al., 2012).

As a whole, self-report tools are both helpful and restrictive in this case. While they do provide a standardised measure across time to assess change and overall, all of those discussed have well validated psychometric properties, they are inherently restrictive. Closed-question scales and questionaries do not allow for nuance or individualised assessment outside of the defined sub-factors, and can only measure verbal behaviour about an entirely different set of behaviours. They report change in hypothetical constructs, not observable behaviour. This can be unreliable, and present a range of internal and external validity issues. Finally, they are an indirect measure of behaviour, relying solely on the participant or patient to be able to accurately report their own behaviour. They are inconsistent with a behavioural lens in that it is not possible to empirically measure the behaviours a treatment is aiming to change. If a treatment is solely focused on changing how an individual thinks about their body, self-report measures alone may be sufficient to establish treatment effects. However, in the case of restrictive EDs, where the patient's life is in danger if their eating behaviour does not change, self-report alone cannot be considered a sufficient metric to determine whether a treatment is effective or enduring.

Observable Measures

Observables measures then, should at the very least be combined with self-report measures. Physiological measures such as change in weight and BMI across treatment are used in this space, and while more consistent with the behavioural approach, as has been established in Chapter Two, body weight is often not a sufficient marker of nutritional status. It also does not provide an immediate measure of change. While it may be useful to record at baseline and EOT to demonstrate treatment effects in underweight participants, it cannot provide sufficient data alone for informing how treatment progresses. On this, in terms of treating behavioural symptoms of restrictive EDs, body weight as a sole objective outcome measure is still an indirect measure of treatment effects as it does not directly report behavioural change, but rather physiological changes that may or may not be as a result of behavioural change. In order to be consistent with a behavioural lens, studies in this space should measure some dimension of behaviour (rate/frequency, duration or latency, for example) empirically, outside of just self-report and increased body weight.

Chapter Four: Systematic Review

This study has chosen to identify three disorders and examine the literature on the treatment of them as assess its scientific rigor and provide a basis for evidence-based practice. Evidence-based practice (EBP) underpins Applied Behaviour Analysis (ABA). EBP is a model of decision-making which implores practitioners to integrate the best available evidence with their practice when conceptualising, evaluating and developing interventions (Slocum et al., 2014). This review is consistent with this approach in that it systematically evaluated the literature base on restrictive EDs and draws conclusions about their merits for informing practice.

So far, it has been established that the core understanding of restrictive EDs here is as a category defined by their function. This thesis has reconceptualised the DSM-5 criteria of restrictive eating disorders in behaviour analytic terms by identifying the behavioural mechanisms which underly common symptoms and diagnostic criteria of AN, BN, and ARFID, and it has identified that these disorders require effective treatment which result in socially significant results due to the severity of the disorders and their effect on an individual's health and quality of life. This thesis posits disordered eating, particularly restrictive EDs as a set of clinically significant maladaptive behaviours which typically, if allowed to maintain, result in malnutrition and its described symptoms as these behaviours all functionally restrict adequate energy intake.

Although the criteria for AN and BN stipulate some desire for thinness or cognitive distortions about body satisfaction, this thesis takes the view that because this is the only significant difference between an AN and an ARFID diagnosis, these disorders can be grouped together. Similarly, although BN does not always result in a marked decrease in body weight as AN and ARFID does, the same deficiencies in nutritional intake are observed. As such, these three disorders have been identified for investigation. Further, it is clear that the development of restrictive behaviours can affect people of all ages and genders. This thesis was interested in including participants of all demographics in an effort to includes populations who are often overlooked in research as presenting with EDs, such as the elderly, who often are not diagnosed with, or treated for ED's, although their symptoms are often identical, barring the previously noted/removed stipulation that an individual display a motivation to achieve or maintain a very low body weight and distorted body image.

The physiological and psychosocial effects of restrictive ED behaviours and the harmful consequential effects of malnutrition explained in Chapter Two highlight why this is the area of interested for this thesis. The importance of good research in this space which informs the evidence base for treatment is substantiated by the relapse and mortality rates observed in those with restrictive EDs. As such, this study chose to conduct a systematic review which analysed the treatment methodology and outcome measures used in the published literature on restrictive EDs following PRISMA methodology. It investigated how and what objective progress/outcome measures are being used in current studies on the three key diagnoses, and how these might be adjusted using a behaviour analytical approach. The key research questions of the search were as follows:

1. What measurement systems and research designs are being used in research in the restrictive ED treatment space?
2. Are these systems in line with certain accepted scientific rigor criteria?

And more broadly:

3. Taken together, what do these systems and designs mean for the state of social validity in the research currently being done on restrictive EDs as a whole?

Method

Search

This study used PRISMA 2020 guidelines for conducting the search and screening process (Page et al., 2021). The systematic search was conducted on 14 September 2023 on Psychnet, PubMed and CINAHL, using the same search string across databases. This string and fields selected is as follows:

"Feeding and Eating disorders" OR "Restrictive eating disorder" OR "Anorexia nervosa" OR "bulimia nervosa" OR "avoidant restrictive food intake disorder" OR ARFID (IN TITLE)

AND

Treatment OR intervention OR experiment* *OR therapy (ALL FIELDS)*

AND

CBT OR "cognitive behavioural therapy" OR ACT OR "Acceptance and commitment therapy" OR ABA OR "Applied Behaviour Analysis" (ALL FIELDS)

NOT

binge eating disorder or binge eating or compulsive eating (ALL FIELDS)

At the time of the search, limits applied using the database's inbuilt exclusion tools were to studies: published in English after August 2013, and to human-only studies. All other limits were established with search terms and applied throughout the data screening phase.

Screening Process

At the screening phase, studies were assessed by the primary observer as well as by a trained independent observer (20% of the sample at each phase) in order to establish IOA. Studies sought for retrieval were assessed at three phases: title only, abstract, and full text. At each stage, observers either chose to include or exclude studies in the next stage based on the criteria below. If at the title or abstract phase there was insufficient information available for if studies met the established criteria, they were included in order to allow for further analysis at the next stage.

Inclusion and Exclusion Criteria

As noted in the search string, studies were limited to three treatment interventions: ACT, CBT (any variant), OR BT. Primarily medical, drug, nursing and nutritional studies were excluded. Multi-faceted treatments, (for example CBT with nutrition counselling, drug therapy and art therapy), were not excluded, however none were identified for inclusion after the abstract screening stage. Between-subjects designs were excluded, and both single-subject and within-subject designs were included. Importantly, studies needed to be primarily concerned with treating symptoms of EDs in patients with active EDs, meaning that studies which primarily aimed to treat comorbid symptoms of anxiety or depression, for example, were excluded. Studies which aimed to treat these as well as observable ED factors were included. Studies which aimed to determine the effectiveness of two different treatments by employing a between-subjects design were excluded. In order to be included, studies needed to report some form of objective outcome measure and treatment effect score. Treatment had to be focussed on the patient with ED (as opposed to primarily on caregivers or dependants of patients with EDs, for example). Published case studies (with no objective outcome measures) or descriptive studies of TAU were excluded. Reviews or meta-analyses were excluded.

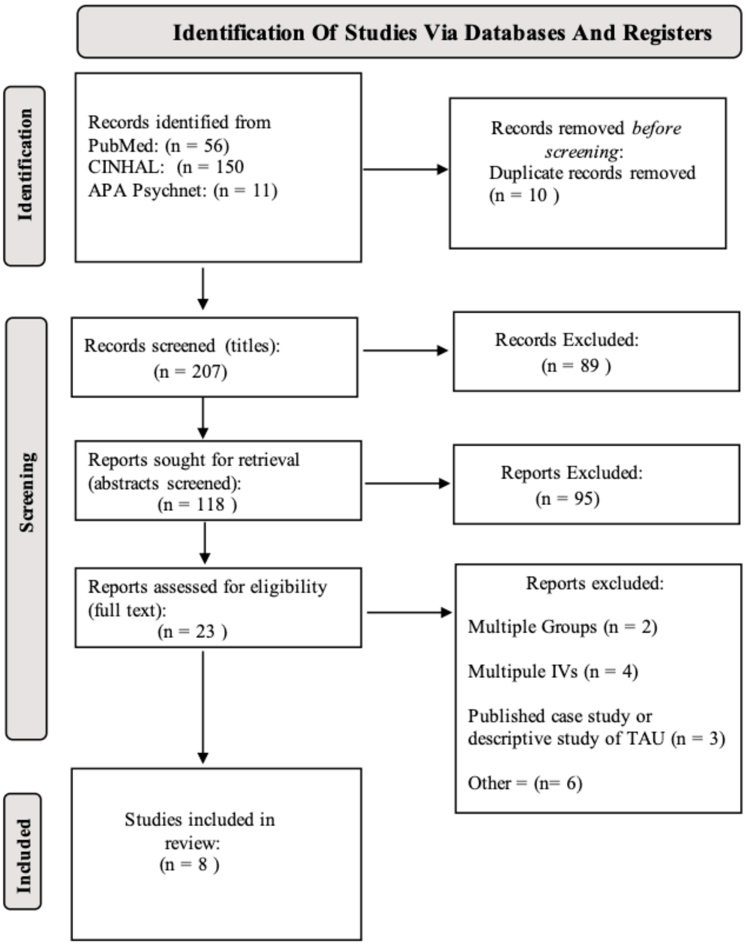
Qualitative studies were included at the title screening phase to be evaluated at the abstract screening stage, but as objective outcome measures had to be reported for inclusion, studies which provided thematic conclusions alone were excluded. No qualitative studies were included in the final review as a result.

Finally, studies primarily concerned with determining predictors of treatment adherence or predictors of favourable treatment outcomes were excluded.

Figure 1. depicts the flow of eligible and candidate studies at each phase of screening. A total of 217 articles were identified for retrieval, with 207 remaining after automatic duplicate screening. These article's titles were screen for inclusion and exclusion criteria, and 87 were removed. Abstracts of the remaining 118 articles were screened, with 95 being excluded at this phase. Finally, 23 candidates were assessed for eligibility based on a full text screen. Reasons for exclusion at this stage are identified in *Figure 1.* Eight studies were identified which met the inclusion criteria for further analysis. These studies are identified below.

Figure 1.

Prisma Flow Chart of Screening Process



Articles Identified for Analysis:

1. Burton Murray, H., Becker, K. R., Breithaupt, L., Dreier, M. J., Eddy, K. T., & Thomas, J. J. (2022). Food neophobia as a mechanism of change in video-delivered cognitive-behavioral therapy for avoidant/restrictive food intake disorder: A case study. *International Journal of Eating Disorders*, *55*(8), 1156–1161. https://doi.org/10.1002/eat.23761
2. Calugi, S., Sartirana, M., Frostad, S., & Dalle Grave, R. (2021). Enhanced cognitive behavior therapy for severe and extreme anorexia nervosa: An outpatient case series.

International Journal of Eating Disorders, *54*(3), 305–312. https://doi.org/10.1002/eat.23428

3. Fischer, A. J., Luiselli, J. K., & Dove, M. B. (2015). Effects of clinic and in-home treatment on consumption and feeding-associated anxiety in an adolescent with avoidant/restrictive food intake disorder. *Clinical Practice in Pediatric Psychology*, *3*(2), 154–166. https://doi.org/10.1037/cpp0000090
4. Dalle Grave, R., Conti, M., & Calugi, S. (2020). Effectiveness of intensive cognitive behavioral therapy in adolescents and adults with anorexia nervosa. *International Journal of Eating Disorders*, *53*(9), 1428–1438. https://doi.org/10.1002/eat.23337
5. Jansen, A., Dumont, E., Mulkens, S., Kroes, D., & Haan, E. (2019). A new cognitive behavior therapy for adolescents with avoidant/restrictive food intake disorder in a day treatment setting: A clinical case series. *International Journal of Eating Disorders*, *52*(4), 447–458. https://doi.org/10.1002/eat.23053
6. King, H., Howarth, R., Choi, S. R., & Fischer, A. J. (2022). Using a Teleconsultation-Enhanced Treatment for Avoidant/Restrictive Food Intake Disorder in an Adolescent Male. *Child & Family Behavior Therapy*, *44*(1), 35–59. https://doi.org/10.1080/07317107.2021.2024716
7. Setsu, R., Asano, K., Numata, N., Tanaka, M., Ibuki, H., Yamamoto, T., Uragami, R., Matsumoto, J., Hirano, Y., Iyo, M., Shimizu, E., & Nakazato, M. (2018). A single-arm pilot study of guided self-help treatment based cognitive behavioral therapy for bulimia nervosa in Japanese clinical settings. *BMC Research Notes*, *11*(1), 257. https://doi.org/10.1186/s13104-018-3373-y
8. Thomas, J. J., Becker, K. R., Kuhnle, M. C., Jo, J. H., Harshman, S. G., Wons, O. B., Keshishian, A. C., Hauser, K., Breithaupt, L., Liebman, R. E., Misra, M., Wilhelm, S., Lawson, E. A., & Eddy, K. T. (2020). Cognitive-behavioral therapy for

avoidant/restrictive food intake disorder: Feasibility, acceptability, and proof-of-concept for children and adolescents. *International Journal of Eating Disorders*, *53*(10), 1636–1646. https://doi.org/10.1002/eat.23355

Inter-Observer Agreement

Inter-observer agreement (IOA) was obtained by a trained observer with a 20% sample of each phase of screening. Total IOA showed total agreement of 80.88% (*n* = 68). At title screening, agreement was 70.73%(*n = 41)*. At abstract screening, agreement was 95.65%(*n=23)*, and at full text screening, agreement was 100%. (*n = 4)*. It was established that the observer was more stringent at the title screening stage and more consistently rejected qualitative studies, which resulted in lower agreement at this stage. During abstract screening, however, qualitative studies were rejected by both the primary and secondary observer. Acceptable IOA was obtained overall and at the abstract and full text screening stages. Full IOA agreement tables can be found in Appendix B.

Analysis

Scientific Merit Rating Scale

In order to assess the rigor of the studies included in the review, the Scientific Merit Rating Scale (SMRS) (National Autism Centre., 2015) was used. This scale was designed to objectively determine whether or not an intervention is effective for participants, and acts as a tool to determine whether variables within a study are controlled sufficiently to draw firm conclusions from a study’s findings. Although this tool was primarily designed to assess studies on the effectiveness of treatments for participants with Autism Spectrum Disorder, it was selected for this thesis for a number of reasons.

Firstly, this tool was selected as it weights single-subject designs as as effective as group designs at establishing experimental control. Further, the SMRS applies ABA standards to assessing the rigor of RCTs (randomised control trials), and accounts for both generalisation and maintenance effects, which the other tools considered for this purpose did not. As mentioned, restrictive EDs have a relatively high rate of relapse, meaning that treatment effects should be evaluated beyond just EOT to ensure treatment adherence, thus further supporting the use of the SMRS in this field. The SMRS also places a firm emphasis on obtaining observation-based objective data both during and after treatment, IOA, and sufficient data points during each stage of the intervention.

Scoring the SMRS

The SMRS is scored by assigning a score between zero and five to five different dimensions of the study. These dimensions are research design, measurement of the dependant variable, measurement of the independent variable, participant ascertainment and generalisation and maintenance effects. A score of five is a strong score, with zero representing a poor score. Dimensions scores are then totalled and averaged to the closest round number to produce the SMRS score (see Appendix B for formula for combing scores). The final SMRS score can be used to draw conclusions about the scientific rigor of the study. A score between three and five indicates that sufficient scientific rigor has been applied, and that should another study use similar or better methods, similar results would likely be obtained. A score of two indicates that initial evidence of treatment effectiveness might be established by the study, but that more rigorous research should be conducted. A score of zero or one indicates that there is insufficient evidence to suggest if a treatment has been effective.

Results

Table 1. shows the SMRS scores for studies included in the final review for each dimension identified by the tool. Total SMRS scores are presented, with scores ranging from 3 to 5. Most studies received a score of at least one for all dimensions, except for studies seven and eight which did not meet the criteria for a score on the generalisation and maintenance dimension. The SMRS assessment found that four studies received an overall scientific rigor score of 3 (one, two, four and seven), three received a score of four (studies three, five and eight). Only study six received a score of five.

Table 1.

Quality Assessment Score Breakdown by Study (Scientific Merit Rating Scale)

Study #	Research Design	Measurement of DV*	Procedural Integratory	Participant Ascertainment	Generalisation and Maintenance Effects	Overall SMRS Score
1	3 (0.9)	2 (0.5)	1 (0.15)	4 (0.8)	2 (0.2)	3
2	4 (1.2)	3 (0.75)	1 (0.15)	4 (0.8)	2 (0.2)	3
3	5 (1.5)	4 (1)	5 (0.75)	3 (0.6)	4 (0.4)	4
4	3 (0.9)	3 (0.75)	2 (0.3)	5 (1)	4 (0.4)	3
5	4 (1.2)	4 (1)	2 (0.3)	3 (0.6)	5 (0.5)	4
6	5 (1.5)	5 (1.25)	2 (0.3)	5 (1)	5 (0.5)	5
7	4 (1.2)	2 (0.5)	2 (0.3)	5 (1)	0	3
8	4 (.1.2)	3 (0.75)	5 (0.75)	5 (1)	0	4

**Note.* Number in parentheses is the raw score ascertained by applying the formula for combining dimensions. Dimensions are not equally weighted in final score calculation. See Appendix A. For studies using both self-report and direct observational measures, each variable's measurement system was scored. Total scores than added together and averaged to obtain final score. For single subject designs, requirements for full research design score indicated >3 participants. This was excluded from consideration for studies where participant *n* = 1.

Table 2 shows demographic variables and treatment inventions and designs. Female participants were over represented in this sample. The total number of participants included was 236, of which 9% of participants were male (n = 22). Study four included 150 female patients, but even removed, males still only made up 25% (male n = 22; female n = 64) of total participants. Mean age of participants ranged from 13.2 to 30.6 years (M = 19.02 years). Six studies used some form of CBT (studies one, two, five, seven and eight) while only two studies used BT (three and six). No studies which used ACT were identified, despite its inclusion in the search. Three studies which used single-subject designs (one, three, six) were included, with the other five being within-subjects designs, showing an over representation of CBT and group designs. The two studies which used BT were both single-subject and both aimed to treat eating behaviour in adolescent males with ARFID. One of these studies identified that the patient also had a diagnoses of Autism Spectrum Disorder (ASD) (study six), while the other did not, but noted the patient had a significant feeding anxiety (study three). Feeding anxiety was also measured in this study. The majority of studies were on ARFID, (studies one, three, five, six and eight), with two studies AN (two and four), and one study on BN (study seven).

Table 2.

Demographic and Research Design Variables Across Studies

Study #	Title	Author	Diagnoses of Patients	Treatment used	Design	Participant (*n*)	Participant Gender	Age (years)
1	Food Neophobia As A Mechanism Of Change In Video-Delivered Cognitive-Behavioural Therapy For Avoidant/Restrictive Food Intake Disorder: A Case Study	Burton Murray et al. (2022)	ARFID	CBT-AR	Single-Subject	1	Male= 1	16
2	Enhanced Cognitive Behavioural Therapy For Sever ANd Extreme Anorexia Nervosa: An Outpatient Case Series	Calugi et al. (2020)	AN	CBT-E	Within-Subject Group Design	30, 20 to completion, 12 to 60-week FU	Female = 26 Male = 1	*M* = 22.4
3	Effects Of Clinic And In-Home Treatment On Consumption And Feeding-Associated Anxiety In An Adolescent With Avoidant/Restrictive Food Intake Disorder	Fisher, Luiselli & Dove. (2015)	ARFID	BT (Positive Reinforcement) + CBT	Single-Subject Changing Criterion Design (Multiple Baseline)	1	Male =1	16
4	Effectiveness If Intensive Cognitive Behavioural Therapy In Adolescents And Adults With Anorexia Nervosa	Grave et al. (2020)	AN	CBT-E	Within-Subject Group Design	150. 70 completed treatment; 36 to final FU	Female = 150	74 Adolescent (*M* =16.5) 81 Adult (*M* =30.6)
5	A New Cognitive Behaviour Therapy For Adolescents With Avoidant/Restrictive Food Intake Disorder In A Day Treatment Setting: A Clinical Case Series	Jansen et al. (2019)	ARFID	Exposure Based CBT	Within-Subject Nonconcurrent Multiple Baseline (AB)	11	Female = 4 Male = 7	*M* = 13.9
6	Using A Teleconsultation-Enhanced Treatment For Avoidant/Restrictive Food Intake Disorder In An Adolescent Male.	King et al. (2022)	ARFID	Telecommunication BT (DRA+CM)	Single-Subject Concurrent Multiple Baseline	1	Male =1	17
7	A Single Arm Pilot Study Of Guided Self-Help Treatment Based Cognitive Behavioural Therapy For Bulimia Nervosa In Japanese Clinical Setting	Setsu et al. (2018)	BN	CBT-GSH	Within-Subject Group Design	25	Female = 25	*M* = 25.6
8	Cognitive-Behavioural Therapy For Avoidant/Restrictive Food Intake Disorder: Feasibility, Acceptability, And Proof Of Concept For Children And Adolescents	Thomas et al. (2020)	ARFID	CBT-AR	Within-Subject Group Design	20 (17 to completion)	Female = 9; Male = 11	*M* = 13.2

The measurement tools and treatment length and procedures used by each study are identified in *Table 3*, with all studies using some form of self-report and some form of objective measure, such as weight, at some point in treatment. Most studies used a distress rating scale or ED specific scale of some kind, and at follow up, four studies reconducted baseline measures to assess whether participants still met the criteria for a diagnosis at EOT. Length of treatment ranged from eight weeks to 40 weeks, and interventions mostly occurred in outpatient settings, with only two being delivered in the participants home. Follow up occurred between four (one month) and 60 weeks in studies which conducted it. Two studies saw a reduction in therapeutic effect at follow up. It is not possible to determine if the two studies which did not conduct follow up measures saw significant change following EOT. Three studies observationally measured actual eating behaviour by measuring some dimension related to food consumption (studies three, five and six). Studies three and six measured frequency of bites, while weight percentage of food consumed was measured in study five, although behaviour was not directly measured. One study reported IOA of over 80% (study three) of the just two (study six) which reported it.

Table 3.

Research Methodology and Measurement Systems by Study Before, During And After Intervention

Study #	Baseline Measures	During Treatment Measures	End of Treatment	Follow Up (FU) Measures	Treatment Length	Treatment Effects reported	Treatment setting	Maintained Therapeutic Effects at Follow Up
1	Weight, height, BMI (normal – measurement discontinued), FNS, NIAS,PARDI-AR-Q	Weekly FNS, mid-treatment NIAS & PARDI-AR-Q, completed electronically before sessions	FNS, NIAS,PARDI-AR-Q	2 month FU-Weight, Height, BMI, FNS, NIAS,PARDI-AR-Q	26 Weeks total (4 pre-treatment, 10 treatment, 8 post-treatment)	FNS, NIAS, PARDI-AR-Q	Live Video Delivered – Home	✓
2	Weight, BMI,EDE-Q, CIA, BSI	Body weight (every 4 weeks) – no psychometric measures reported	Weight, BMI,EDE-Q, CIA, BSI	20 and 60 week FU – Weight, BMI,EDE-Q, CIA, BSI	40 Weeks	BMI,EDE-Q, CIA, BSI	Outpatient (Clinic)	×
3	Clinical interview, individual eating test (for food acceptance)	Topographically defined behaviour (food ingestion) – frequency recording. Self-monitoring, parent monitoring and clinician monitoring; Distress rating	Maintenance data recorded one week after 100% consumption observed	1 and 3 month single session FU	59 clinic sessions, 67 home meals (Maintenance (6), FU (9)	Cumulative bites per session in two settings per trial, subjective distress rating per trial.	Clinic + Home	✓
4	Weight, BMI,EDE-Q, CIA, BSI	Weight, BMI,EDE-Q, CIA, BSI	Weight, BMI,EDE-Q, CIA, BSI	20 and 60 week FU, all baseline tests repeated	20 Weeks	EDE-Q factor scores, BSI global score, CIA global score	13 Week inpatient, 7 weeks day-treatment	×
5	Individual Eating Test - % of overall food weight accepted, Eating Anxiety measured via VAS before and after meals (n=7), Body weight and length,	Acceptance of 50% of baseline foods, FNS, CCMs measured via VAS, Food diary	Food selectivity test, food acceptance score, DSM-5 ARFID index change	FU 3 months post treatment - All baseline tests repeated - food diary and body weight not reported	8 Weeks (4 Weeks CBT, 4 Weeks low intensity)	Food acceptance %, FNS score, Eating anxiety via VAS score (*n*=7).	Day treatment 4 Weeks, outpatient 4 weeks	✓
6	BMI, feeding disorder questionnaire (unspecified), food preference assessment, food preference survey	Food consumption frequency (topographically defined), rate of problem behaviour	Food consumption frequency (topographically defined), rate of problem behaviour	1.5 month FU	Two months (40 sessions)	% consumption per meal for 3 food types	Home	✓
7	EDE 16D, EDE-Q, BITE,PHQ-9, and the GAD-7.	Height and weight before each treatment session	Same as baseline measures	No FU	16-20 Weeks	Primary results from EDE 16D. Results for all scales in baseline presented as cohort - not individualised. No during -treatment effects reported.	In-patient	N/A
8	PARDI, FNS, STAIC, CDI-2, height, weight, BMI percentile, mensuration status	Credibility/Expectancy Questionnaire (following only session one) ,	Number of new foods tried, FNS, height, weight, BMI, menstruation status, CSQ, CGI, STAIC	No FU	33.7 Weeks (averaged)	Number of new foods tried, Credibility, GIS score, PARDI interview (patient and parent) FNS, weight, BMI,CDI-2 (non-sig), STAIC (non-sig)	Out-patient (Clinic)	N/A

Discussion

The primary focus of this systematic review was to investigate what measurement systems and research designs are being used in research in the restrictive ED treatment space, evaluate if these systems in line with certain accepted scientific rigor criteria and more broadly, consider what these findings mean for the state of social validity in the research currently being done on restrictive EDs as a whole. These key questions are addressed in this section.

Quality Assessment

The SMRS (National Autism Centre., 2015) provided a framework for evaluating the quality of research in this area. The SMRS identified five dimensions on a which study was rated. These areas included research design, measurement of dependant variable, procedural integratory, participant ascertainment, and generalisation and maintenance effects. Combined, these yielded an overall SMRS score, indicating if the research design was rigorous enough to contribute meaningfully to the evidence base and inform treatment, the scores for which can be found in *Table 1.* Notably*,* low scores on the SMRS procedural integrity dimension usually indicated that IOA for intervention fidelity and implementation accuracy were not reported at all (only two studies reported IOA, and one of these reported a quantified measure of implementation accuracy). For example, study six, met all requirements for a score of five, however treatment integrity was 71%, meaning it received a lower score of four.

Studies three, five, six and eight scored the highest overall on the SMRS. This provides evidence for a clear correlation between BT interventions, SCEDs and high scientific rigor on this metric. The SMRS stipulates that scores of three or above indicate that sufficient scientific rigor has been applied, and therefore, that strong conclusions can be drawn from the results of a study. As noted, all studies in this sample received a score above three. This may highlight a limitation of this study. The SMRS assessment tool was established for evaluating the evidence base for interventions designed for participants with ASD. While it was used as part of a broader analysis and was not the primary metric used in assessing studies, this may be why all studies received a satisfactory or higher overall score. Study eight, for example, provided no follow up data and no ongoing measurement but still received a high score. This may highlight a failure of the tool in this field specifically. However, the use of a novel scientific rigor and quality assessment tool which weighted both single-case and group designs equally is a strength here. Although originally designed to investigate studies in the Autistic Spectrum Disorder (ASD) field, the overlap of those with ASD and disorders such as AN and ARFID has been observed in the literature (see: Huke, et al., 2022). Further, as stipulated above, it was the view of this work that SCEDs are valuable in this field, and a clear bias towards RCTs and group designs was observed in the other tools considered for this purpose. The SMRS also accounted for generalisation and maintenance measurements, something observed in very few scientific rigor assessment tools. In order to account for the limitations that this tool may have presented, the WWC (What Works Clearinghouse, 2020) criteria for a study to meet acceptable evidence standards is considered here.

The WWC criteria specify that for a study to meet evidence criteria the independent variable must be systematically manipulated, each outcome variable must be measured systematically over time by more than one assessor for more than 20% of datapoints and agreement must meet minimal thresholds (usually 80% or above agreement), and that the study must include three or more attempts to demonstrate treatment effect at three distinct point in time or repetitions, and for a phase to demonstrate an effect, it must have a minimum of three data points. Only one study met these criteria in its entirety (study six).

Measurement Systems

The first part of the research question queried what measurement systems were being used in studies in this area. Most used some objective measure at both baseline and EOT, and most established follow up measures in some way. However, the majority of studies only reported weight or BMI change as an objective treatment outcome measure and as indicators of treatment efficacy. As discussed, observable measures are the only way to empirically establish that behaviour is changing. Self-report tools, at best, show that verbal behaviour is changing, but in the case of restrictive EDs, it is eating and food related behaviours that should be of interest in order to be sure that research in this space is empirically measuring and effectively treating behaviours that meet the functional definition of restrictive EDs applied throughout this work.

Scales, Questionnaires and Surveys

Closed question tools were used in the majority of studies in this review in order to diagnose EDs or quantify treatment effects, as well as to self-report ongoing symptom change over the course of treatment. These tools consist of standardised questions, and they allow researchers or practitioners to measure aspects of ED symptomology systematically. The specific tools used are discussed in Chapter Three, however the prevalence of these tools in the sample is significant (see *Table. 3)*.

Tools such as the EDE-Q demonstrate good internal reliability (Luce & Crowther, 1999) and at the very least can assist in identifying cases of active EDs (Anstine & Grinenko, 2000; Cotton et al., 2003; Johnston et al., 2007). However, they present several challenges for establishing construct validity.

Firstly, validity of self-report tools can be compromised by response bias, which refers to the phenomenon whereby people give inaccurate answers to surveys or questionaries. People may respond in a way that they feel the interviewer wants them to (social desirability bias), or the interviewer may unintentionally prompt the participant to answer inaccurately (interviewer bias). Survey style tools of this nature may include reverse coded questions to account for response bias to a degree, but there remains debate about whether reversed items should be included in closed question surveys at all (Vigil et al., 2020). From an ABA perspective, Hawkins (1991) notes that responses of this nature act to provide a confirmation of practitioners opinions, as opposed to accurately predicting of benefits or costs of a given intervention. There is evidence to suggest that people self-reporting behaviour unreliably (Fixsen, Phillips, & Wolf, 1972). Specifically in this context, people may be unreliable reporters of their own behaviour due to the physiological effects malnutrition, which, as established, is a primary function of restrictive EDs. Specifically in the case of people who are undernourished, significant impairment in memory and cognitive function has been observed (Kesari, Handa, R., & Prasad, 2010).

Secondly, even if participants answer in a way that they feel is accurate, or response bias is controlled for by the tool, self-report tools are inherently limited in their ability to capture individual variance and complexities related to actual behaviour. Closed-question tools specifically limit participant answers to a degree in which subtle nuances cannot be captured. On this note, self-report closed question tools may not be culturally sensitive. As noted, what constitutes normal eating behaviours is highly dependent on the social setting of an individual. This extends to the cultural context as well (see for example: Axelson, 1986, Rozin, 1996). There is work being done on establishing the reliability of ED specific tools in non-Western countries and in diverse populations in these areas (see for example: Mohd Taib et al., 2022; Oliveira Júnior et al., 2023), however, comprehensive culture-specific research is required before these tools can be used outside of the groups they were developed for.

Another challenge for self-report tools is the significant discrepancies between tools in their findings, such as between the EDE-Q and the EDE with respect to binge eating, and studies in this space note that this is probably largely due to the difficulty in assessing instances of binge eating using self-report (Black and Wilson, 1996, Carter, Aime and Mills, 2001, Fairburn and Beglin, 1994, Wilfley et al.,1997, Meadows et al., 1986 as cited in Mond et al., 2004). This suggests that the tool alone might produce different conclusions about a person's behaviour.

Structured interview (open question) tools were also used in this sample. These tools give a more comprehensive, qualitative impression of participants symptoms and experiences, but suffer the same problems as all self-report measures. There is evidence that they can be useful in diagnosing EDs (Sysko et al., 2015), but again, alone, should not be considered sufficient for measuring relevant behavioural change as they are indirect measures of behaviour.

This is not to say that self-report measures should not be used at all in this field. It is simply to reiterate in context that without also measuring observable behaviour, claims cannot be made about the efficacy of treatment. This is consistent with the ABA approach to self-report as a measure of behaviour more broadly (Baer, Wolf, & Risley, 1968). Although these tools have established support for their reliability, even though an instrument is reliable, it may not be valid, in that it may measure an irrelevant dimension or construct. In this case, because the data is not obtained by empirically measuring behaviours of concern, it is not clear that these tools are accurately measuring what they intend to.

Measurement of DVs - Direct Observation

Direct measurement of ED behaviours in this context should then involve the objective and systematic observation of participants eating habits or food intake in a controlled environment. As outlined in a number of works in the area (Baum 2017, Cooper, Heron & Heward 2020), this is directly in line with behaviourism and the understanding of behaviour as a science, however, only two studies in this sample reported doing this.

Although study one had the opportunity to present observation data (e.g number of foods consumed in exposure or number of foods consumed at follow up) as this was a part of its intervention, it did not, with the authors rather choosing to present ongoing FNS scores. This was the only metric presented as a quantifiable treatment effect and was self-reported by the participant at the start of each session. Study six was the only study to present a measure of a directly observed dimension of behaviour – food consumption percentage per session. This study perfectly demonstrates how direct observation can be combined with CBT and surveys (FF-TAM and a food preference survey) to provide a comprehensive view of change in observable behaviour and changes in self-reported data.

Although it is easy to idealise about what interventions should look like, it can be time and resource intensive to conduct direct observations (Bakeman, & Quera, 2011). Studies in this area might have considered using less intensive systems of measurement that still aim to capture a dimension of the behaviour outside of just self-report. These might have included employing diaries or incorporating family support (Dunlap et al., 2001) to conduct reliability checks, such as interrater checks via texts. Diaries allow individuals to record their eating behaviours, thoughts, and emotions, providing a detailed account of their experiences at the time, helping to mitigate the effects of interviewer bias and memory problems (Morrison, M. (2002). Furthermore, involving family members in the data collection process might have allowed for IOA to be established (as in study seven), helping to validate the reliability of the data. Additionally, conducting reliability checks, such as interrater checks via texts during the course of treatment would have helped to ensure consistency and agreement between observers in recording and reporting behavioural change outside of the clinical setting.

On this, ongoing measurement was observed in most studies in the sample, however study seven only measured weight before sessions, and study eight only measured client satisfaction (using the CIA) once (this was the only measure conducted between baseline and

EOT). Ongoing measurement should inform how treatment progresses and serves to illustrate treatment efficacy as soon as possible so that a practitioner might change course if an intervention is ineffective or harmful (Leaf et al., 2016).

Further, while most studies conducted some form of FU measure, two of these studies observed a regression in treatment effects. With the marked rate of relapse following treatment, specifically in individuals with BN (see Stice, Marti, & Rohde, 2013), it should be imperative to establish both ongoing and follow up data to assess treatment efficacy.

Weight and BMI

Almost all studies in this samples used weight or BMI as an indicator of treatment effect. Although rationalised in this context by the DMS-5 criteria (for AN and ARFID specifically) for a markedly low body weight as a stipulation for diagnosis, improvements in this area alone, again, are not sufficient for drawing conclusions about individuals eating behaviours or therapeutic changes in an individual's environment. Changes in body weight or BMI may signal that an individual is moving towards a pattern of favourable nutritional intake but does not inform what is causing or maintaining this effect and is not an immediate measure of treatment effectiveness. The complexities involved with using body weight as a marker of nutritional intake are outlined in Chapter Two, which establishes that a range of factors contribute to both increases and decreases on this metric. BMI specifically is controversial as it fails to account for factors such as bone structure, sex, age, or muscle mass (Rothman, 2008), and at best is an indirect measure of body fat (Hulkoti, V. S., Acharya, S., Shukla, Partapa, & Gupte, 2020). Most of the criticism of BMI as a metric relates to measures of obesity, suggesting a gap in the literature about how and if BMI should be used as a measure for underweight individuals at all. Further research should explore this.

Research Designs

The research question was also concerned with what research designs were being used. As above, the majority of the studies in this sample used group designs, indicating a trend in this area. Despite the exclusion of between-subject designs, group designs were still overrepresented, with only two single-subject designs identified. This may be as result of the trend to favour cohort RCTs in psychological research more broadly. Additionally, the analysis revealed an overrepresentation of adolescent females in the sample, persisting even after controlling for study four. While this aligns with existing literature on demographic variables (Van Eeden, Van Hoeken & Hoek, 2021), it prompts consideration of the other groups potentially slipping through therapeutic gaps, particularly adults and men who may be less likely to seek care or receive a diagnosis. Interestingly, all of the interventions on ARFID included only male participants and most made reference to comorbid developmental disorders. This observation is linked to the literature, highlighting the association between ARFID and Autism Spectrum Disorder (ASD) (Inoue et al., 2021) as well as the overrepresentation of males in ASD diagnoses (Giarelli et al., 2010.). Older adults with ED symptoms, for example, are not represented in this sample, despite the evidence that this is a pervasive problem in elderly care institutions (Fostinelli, et al., 2020; Mulchandani et al., 2021; Sulmont-Rossé, 2020). Further, CBT emerged as the primary psychological intervention, reflecting its widespread practice in this space. While CBT shows evidence for efficacy in treating EDs (Atwood & Friedman, 2020) this is by no means universally agreed upon (Solmi et al., 2021). This overrepresentation, may then, be as a result of CBTs compatibility with group designs and RCTs.

Social Validity

The final part of the research question sought to determine what, taken together, could be concluded about the social validity of research being done in the treatment of AN, BN, and

ARFID. These are considered here in terms of the three dimensions of social validity outlined by Wolf (1978).

Social significance of goals relates to the relevance of the goals targeted by an intervention and ensures that goals of interventions are valuable to both the participant and society. The goals of interventions vary across the eight studies included, but inherently hold one very broad goal in common: to improve eating disorder symptomology in some way. This was, however, measured and reported very differently, making it difficult to claim that all studies showed evidence of establishing social validity. As seen in three studies in this sample (studies two, four and eight), using measures which serve to understand participants perceptions of treatment and treatment works towards establishing relevance to stakeholders and increasing treatment adherence. Similarly, in the case of very young participants, using interview tools which evaluate caregiver perceptions of treatments and establish goals alongside caregivers also work to establish relevance and increase participant buy-in. Importantly, goals of treatment should align with societal values and have a positive impact on quality-of-life. Hawkins (1991) considers the effect on quality of life in terms of establishing the habilitative vale of the goals of an intervention. This means defining whether the goal of the intervention provides both short-term quality-of-life benefits (whether things in the individual's environemnt reinforce behaviour) and long-term benefits. Long-term benefits here are concerned with whether the intervention directly exerts control over the individual's behaviour over time, as opposed to factors in the individuals culture maintaining behaviour.

On this, cultural sensitivity of goals should also be addressed prior to commencing treatment. Studies which aim to test the efficacy of certain treatments in new cultural settings, such as study four, should consider that treatments may not be effective or socially valid if they are applied to groups which differ from those that the treatment is already established in.

In order to establish social significance of goals, researchers must be precise about the goals at three levels: broad social goals, categories of behaviour related to the broad social goal, and responses that comprise the goal of interest. In the context of treating restrictive EDs, improving symptoms in any way which moves a person away from malnutrition is arguably a socially significant goal due to the nature of behaviour. However, specific relevant goals may include improving nutritional intake, improving disordered eating behaviours which functionally limit nutritional intake, and improving overall nutritional intake. This aligns with Hawkins (1991) view as the habilitative validity criterion would be met as people who see an improvement in restrictive ED symptoms would have greater access to social reinforcement (inclusion in mealtimes, for example) and a reduction in punishing experiences.

On this, social appropriateness of procedures pertains to the acceptability of methodology used in treatment and research, ensuring that the design is respectful to the rights of the participant. In this case, ensuring participant autonomy throughout treatment and encouraging collaboration and engagement with the participant works towards this. This might mean having regular discussions with the client about their experiences of treatment, or providing an environment away from the primary researcher or practitioner for the client to express concerns or thoughts about this. Moreover, procedures should be accessible and adaptable. Adaptability can be informed by ongoing measures which signal to the practitioner whether treatment is progressing in a therapeutic way, and allows for the intervention to be adapted before it runs its course. In the context of review, in order to assess appropriateness of procedures, the researcher must be clear about the existing or planned treatment, the treatment components, and/or the specific procedures that comprise the treatment. In a majority of cases, studies in this sample which used CBT or CBT derived interventions were not clear about what exactly treatment entailed. Although CBT is largely individualised, it simply cannot be stated in a published article how treatment progressed and adapted for each induvial in a large group design. This again strengthens the premise that SCEDs are appropriate in this space and provide a good basis for EDP, as they can give clear explanations of procedures to both readers and participants. It also strengthens the claim that goals and procedures should address the functional definition used throughout this thesis, and that behaviours which significantly restrict nutritional intake should be addressed effectively.

Considering the impact on quality of life and enhancement of social functioning as an effect of treatment is also relevant to this dimension of social validity. As discussed, mealtimes serve important social functions for people (Hamburg et al., 2014; Papies et al. (2022) and being excluded from these, coupled with the physiological effects of malnutrition, bodes poorly for overall quality of life for those with restrictive EDs. Identified in this sample, the CIA is a measure which aims to comprehensively understand the multifactorial impacts of EDs and measure symptom change across these dimensions. While it still suffers from the shortcomings of closed-question surveys, it does at least provide an impression of the ways in which a specific treatment has improved quality of life for the participant. Further, the CSQ (Client Satisfaction Questionnaire) (Larsen et al, 1979) was identified in the sample, which also works towards establishing client satisfaction with treatment. However, by applying a behavioural lens which addresses the shortcomings of these types of measures, this thesis has argued that a move away from self-report tools like these might better address social validity to improve treatment adherence. It has offered Hawkins (1991) criteria for establishing habilitative value as an alternate, ABA informed way for researchers in this area to consider the validity of treatment goals, procedures and outcomes. It has been well established that in an area in which people have high rates of relapse following treatment (Croce et al., 2024) and associated mortality rates (Fichter & Quadflieg, 2016, Smirk et al., 2012, Steinhausen, 2002), treatment adherence past EOT should be a defining goal of any intervention in this space.

Broadly, in eating disorder treatment, effective procedures which work alongside phycological and behavioural interventions may include psychoeducation about healthy eating habits and nutritional counselling to promote balanced eating patterns. This would ensure that participants do not functionally regress or relapse due to a lack of education, and works to ensure that treatment leads to meaningful behavioural change.

Finally, social importance of effects is concerned with establishing that outcomes produced by treatment have meaningful and lasting effects on the participant, and work towards positive social functioning or quality of life. Effects can be considered at three levels: proximal effect, intermediate effects and distal. Proximal effects may look like an increased variety of foods being consumed routinely, while distal effects may look like a participant engaging more in social meal settings.

In considering what might be done to prevent relapse after EOT, the importance of establishing generalisation and maintenance and assessing these using follow up measures cannot be overstated. As Kimball et al., (2023) explain, preventing relapse is a highly individualised process, and the authors outline a number of strategies which might be employed in specific interventions to mitigate relapse by promoting generalisation and maintenance both during the intervention and at EOT. Based on this work, generalisation here refers to the range of contexts in which behavioural change extends to both during treatment and at EOT, and maintenance refers to this change enduring past EOT.

In this sample, most interventions were conducted in an in-patient or out-patient clinical setting, not the social or physical environment in which the participants would normally engage in eating behaviours in. As CBT was overrepresented as an intervention in this review, it should be noted that CBT does work to promote generalisation by teaching patients specific skills and techniques to reframe their thinking in a range of environments, as well as teaching participants to recognise warning signs of relapse (Beck, 2020). This may be effective, but it is not possible for this thesis to draw conclusions about this due to the fact that group studies using CBT in this sample did not conduct follow ups in naturalistic settings. This is unfortunate, as this means that no firm conclusions about generalisation outside of the primary intervention setting can be made in these cases. To understand if the effects of a treatment are meaningful and lasting, thus meeting the criteria for social validity to be established, it is important that further research in this area demonstrates clear generalisation effects outside of the treatment setting. Furthermore, some studies conducted no follow up measures at all, while others saw a counter therapeutic change at follow up. In this case, any follow up is better than none, but it is not acceptable for a study to observe a regression toward baseline but claim a positive treatment effect due to the life-threatening and enduring nature of the disorders. As it is clear that relapse is commonly observed in AN, BN and ARFID populations (Croce et al., 2024; Strober, Freeman and Morrell, 1997), further research should determine what the timelines for relapse are following each form of treatment for each diagnosis so that follow ups can be conducted at relevant intervals, and treatment can recommence if necessary.

Strengths and Limitations

It should be noted that this study was not without limitations. Systematic reviews are inherently limited by their respective search parameters, and the wealth of research on ED interventions meant that only a narrow review was feasible here. For this reason, the study parameters excluded between-subject designs, which should be examined in subsequent reviews.

Further, the quality assessment tool (the SMRS) informed the limitation of participants to only those who received a DMS-5 diagnosis of one of the three restrictive EDs of interest. Although this method of screening was helpful in that it ensured that only studies on those who met the clinical threshold for ED symptomology were included, it limits the research to only those who’s consummatory behaviour is already harmful enough to present with criteria for a DSM-5 diagnosis, all of which, for these disorders, stipulates significantly reduced body weight. As such, studies which aim to correct maladaptive eating behaviours before they become significantly harmful to participants were excluded. These should be considered in further research with an aim to synthesise the work in this area to inform practice. This this review was strengthened by applying the functional definition of restrictive EDs and applying a behavioural lens to understanding and evaluating the data. This underlying theory allowed for the evaluation of treatments in the restrictive ED space more broadly, and provides a distinct framework for understanding EDs in this way. The systematic review (PRISMA) methodology allowed for the empirical evaluation of literature in this space to be conducted effectively, and for conclusions about the scientific rigor and social validity of the research to be made.

Conclusions

Taken together, this review has illustrated that the research designs and measurement systems in this area need to be adapted to sufficiently address dimensions of social validity and ensure treatment efficacy. I have made recommendations for how this may be done in this context throughout, predominantly stating that goals of treatment and procedures should be individualised and socially significant, and that interventions should establish maintenance and generalisation effects to prevent relapse after EOT.

More broadly, I would like to suggest that further research should aim to not only effectively treat, but effectively prevent participants from reaching a point where the function of their restrictive behaviours qualify them for an ED diagnosis. Advancements should be made in identifying observable symptoms and implementing effective behavioural change strategies before behaviour is so restrictive that it becomes harmful.

References

A.G. Mamhidir, G. Ljunggren, M. Kihlgren, A. Kihlgren, A. Wimo Advances in Child Development and Behavior, 13 (1979), pp. 257-295

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anstine, D., & Grinenko, D. (2000). Rapid screening for disordered eating in college-aged females in the primary care setting. Journal of Adolescent Health, 26, 338–342.

Atwood, M. E., & Friedman, A. (2020). A systematic review of enhanced cognitive behavioral therapy (CBT-E) for eating disorders. *International Journal of Eating Disorders*, *53*(3), 311-330.

Autism and Developmental Disorders, *46*(2), 720–731. https://doi.org/10.1007/s10803-015-2591-6

Axelson, M. L. (1986). The impact of culture on food-related behavior. *Annual review of nutrition*, *6*(1), 345-363.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1, 91-97.

Basile, C., Gigliotti, F., Colaiori, M., Di Santo, F., Terrinoni, A., Ardizzone, I., & Sabatello, U. (2021). Comparison of neuropsychological profiles in children and adolescent with anorexia nervosa and avoidant/restrictive food intake disorder (ARFID). *European Psychiatry*, *64*(S1), S351-S352.

Baum, W. M. (2017). *Understanding behaviorism: behavior, culture, and evolution* (Third Edition.). Wiley.

Beck, J. S. (2020). *Cognitive behavior therapy: Basics and beyond*. Guilford Publications.

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *The International Journal of Eating Disorders*, *45*(3), 428–438. https://doi.org/10.1002/eat.20931

Birch, L. L., & Marlin, D. W. (1982). I don't like it; I never tried it: Effects of exposure on two-year-old children's food preferences. *Appetite, 3*(4), 353–360. https://doi-org.ezproxy.waikato.ac.nz/10.1016/S0195-6663(82)80053-6

Birmingham, C. L., Su, J., Hlynsky, J. A., Goldner, E. M., & Gao, M. (2005). The mortality rate from anorexia nervosa. *International journal of eating disorders*, *38*(2), 143-146.

Boden, G., Homko, C., Barrero, C. A., Stein, T. P., Chen, X., Cheung, P., Fecchio, C., Koller, S., & Merali, S. (2015). Excessive caloric intake acutely causes oxidative stress, GLUT4 carbonylation, and insulin resistance in healthy men. *Science Translational Medicine*, *7*(304), 304re7-304re7. https://doi.org/10.1126/scitranslmed.aac4765

Bohn, K., Doll, H. A., Cooper, Z., O'Connor, M., Palmer, R. L., & Fairburn, C. G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour research and therapy*, *46*(10), 1105-1110.

Bohn, K., Doll, H. A., Cooper, Z., O’Connor, M., Palmer, R. L., & Fairburn, C. G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour Research and Therapy*, *46*(10), 1105–1110. https://doi.org/10.1016/j.brat.2008.06.012

Carpita, B., Muti, D., Cremone, I. M., Fagiolini, A., & Dell’Osso, L. (2022). Eating disorders and autism spectrum: Links and risks. *CNS spectrums*, *27*(3), 272-280.

Carter, J. C., Kelly, A. C., & Norwood, S. J. (2012). Interpersonal problems in anorexia nervosa: Social inhibition as defining and detrimental. *Personality and Individual Differences*, *53*(3), 169–174. https://doi.org/10.1016/j.paid.2012.02.020

Chin, W. W., Johnson, N., & Schwarz, A. (2008). A Fast Form Approach to Measuring Technology Acceptance and Other Constructs. *MIS Quarterly*, *32*(4), 687–703. https://doi.org/10.2307/25148867

Chou, S. H., & Mantzoros, C. (2014). Role of leptin in human reproductive disorders. *J. Endocrinol*, *223*(1), T49-T62.

Cooper, J., Heron, T., & Heward, W. (2020). *Applied Behavior Analysis, Global Edition* (3rd edition.). Pearson Education, Limited.

Cotton, M.A., Ball, C., & Robinson, P. (2003). Four simple questions can help screen for eating disorders. Journal of General Internal Medicine, 18, 53–56.

Croce, S. R., Malcolm, A. C., Ralph-Nearman, C., & Phillipou, A. (2024). The role of identity in anorexia nervosa: A narrative review. *New Ideas in Psychology*, *72*, 101060-. https://doi.org/10.1016/j.newideapsych.2023.101060

Cui, López, M., & Rahmouni, K. (2017). The cellular and molecular bases of leptin and ghrelin resistance in obesity. *Nature Reviews. Endocrinology*, *13*(6), 338–351. https://doi.org/10.1038/nrendo.2016.222

Demmler, J. C., Brophy, S. T., Marchant, A., John, A., & Tan, J. O. A. (2020). Shining the light on eating disorders, incidence, prognosis and profiling of patients in primary and secondary care: national data linkage study. *British Journal of Psychiatry*, *216*(2), 105–112. https://doi.org/10.1192/bjp.2019.153

Dinkler, Yasumitsu-Lovell, K., Eitoku, M., Fujieda, M., Suganuma, N., Hatakenaka, Y., Hadjikhani, N., Bryant-Waugh, R., Råstam, M., & Gillberg, C. (2022). Development of a parent-reported screening tool for avoidant/restrictive food intake disorder (ARFID): Initial validation and prevalence in 4-7-year-old Japanese children. *Appetite*, *168*, 105735–105735. https://doi.org/10.1016/j.appet.2021.105735

Dunlap, G., Newton, J. S., Fox, L., Benito, N., & Vaughn, B. (2001). Family involvement in functional assessment and positive behavior support. *Focus on autism and other developmental disabilities*, *16*(4), 215-221.

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

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.

Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1972). Achievement place: the reliability of self-reporting and peer-reporting and their effects on behavior 1, 2. *Journal of Applied Behavior Analysis*, *5*(1), 19-30.

Fostinelli, S., De Amicis, R., Leone, A., Giustizieri, V., Binetti, G., Bertoli, S., ... & Cappa, S. F. (2020). Eating behavior in aging and dementia: the need for a comprehensive assessment. *Frontiers in nutrition*, *7*, 604488.

Franke, G. H., Jaeger, S., Glaesmer, H., Barkmann, C., Petrowski, K., & Braehler, E. (2017). Psychometric analysis of the brief symptom inventory 18 (BSI-18) in a representative German sample. *BMC Medical Research Methodology*, *17*(1), 14–14. https://doi.org/10.1186/s12874-016-0283-3

Gallon, Ferreira, C. F., Henz, A., Oderich, C. L., Conzatti, M., Ritondale Sodré de Castro, J., Parmegiani Jahn, M., da Silva, K., & Wender, M. C. O. (2022). Leptin, ghrelin, & insulin levels and food intake in premenstrual syndrome: A case-control study. *Appetite*, *168*, 105750–105750. https://doi.org/10.1016/j.appet.2021.105750

Giarelli, E., Wiggins, L. D., Rice, C. E., Levy, S. E., Kirby, R. S., Pinto-Martin, J., & Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and health journal*, *3*(2), 107-116.

Gilsbach, S., Plana, M. T., Castro-Fornieles, J., Gatta, M., Karlsson, G. P., Flamarique, I., ... & Herpertz-Dahlmann, B. (2022). Increase in admission rates and symptom severity of childhood and adolescent anorexia nervosa in Europe during the COVID-19 pandemic: data from specialized eating disorder units in different European countries. *Child and Adolescent Psychiatry and Mental Health*, *16*(1), 46.

Giordano, Simona, '*Eating Disorders: Anorexia and Bulimia Nervosa*', *Understanding Eating Disorders: Conceptual and Ethical Issues in the Treatment of Anorexia and Bulimia Nervosa*, Issues in Biomedical Ethics (Oxford, 2005; online edn, Oxford Academic, 1 Feb. 2006), https://doi-org.ezproxy.waikato.ac.nz/10.1093/0199269742.003.0002, accessed 5 Apr. 2023.)

Giordano, Simona, '*Eating Disorders: Anorexia and Bulimia Nervosa*', *Understanding Eating Disorders: Conceptual and Ethical Issues in the Treatment of Anorexia and Bulimia Nervosa*, Issues in Biomedical Ethics (Oxford, 2005; online edn, Oxford Academic, 1 Feb. 2006), https://doi-org.ezproxy.waikato.ac.nz/10.1093/0199269742.003.0002, accessed 5 Apr. 2023.)

Giordano, Simona, '*Eating Disorders: Anorexia and Bulimia Nervosa*', *Understanding Eating Disorders: Conceptual and Ethical Issues in the Treatment of Anorexia and Bulimia Nervosa*, Issues in Biomedical Ethics (Oxford, 2005; online edn, Oxford Academic, 1 Feb. 2006), https://doi-org.ezproxy.waikato.ac.nz/10.1093/0199269742.003.0002,

Guy, W. (1976). ECDEU assessment manual for psychopharmacology. Washington, DC : United States Department of HEW Publications.

Hagan, K. E., & Walsh, B. T. (2021). State of the art: the therapeutic approaches to bulimia nervosa. *Clinical Therapeutics*, *43*(1), 40-49.

Hamburg, M. E., Finkenauer, C., & Schuengel, C. (2014). Food for love: The role of food offering in empathic emotion regulation. Frontiers in Psychology, 5. https://doi.org/ 10.3389/fpsyg.2014.00032

Hansen, S. J., Stephan, A., & Menkes, D. B. (2021). The impact of COVID-19 on eating disorder referrals and admissions in Waikato, New Zealand. *Journal of eating disorders*, *9*(1), 10.

Hawkins, R. P. (1991). Is social validity what we are interested in? Argument for a functional approach. *Journal of Applied Behavior Analysis*, *24*(2), 205–213. https://doi.org/10.1901/jaba.1991.24-205

Heffner, M., Sperry, J., Eifert, G. H., & Detweiler, M. (2002). Acceptance and commitment therapy in the treatment of an adolescent female with anorexia nervosa: A case example. *Cognitive and behavioral practice*, *9*(3), 232-236.

Houghton, S., Marais, I., Kyron, M., Lawrence, D., Page, A. C., Gunasekera, S., Glasgow, K., & Macqueen, L. (2022). Screening for depressive symptoms in adolescence: A Rasch analysis of the short-form childhood depression inventory-2 (CDI 2:SR[S]). *Journal of Affective Disorders*, *311*, 189–197. https://doi.org/10.1016/j.jad.2022.05.088

House, J., Eisler, I., Simic, M., & Micali, N. (2008). Diagnosing eating disorders in adolescents: A comparison of the eating disorder examination and the development and well-being assessment. *International Journal of Eating Disorders*, *41*(6), 535-541. http://doi.org.ezproxy.waikato.ac.nz/10.1017/S0007114509990286

Huke, V., Turk, J., Saeidi, S., Kent, A., & Morgan, J. F. (2013). Autism spectrum disorders in eating disorder populations: a systematic review. *European Eating Disorders Review*, *21*(5), 345-351.

Hulkoti, V. S., Acharya, S., Shukla, S., Partapa, S. K., & Gupte, Y. (2020). In search of an ideal obesity assessment tool: Is body mass index reliable enough. *J. Evol. Med. Dent. Sci*, *9*, 2556-2560.

Inoue, T., Otani, R., Iguchi, T., Ishii, R., Uchida, S., Okada, A., Kitayama, S., Koyanagi, K., Suzuki, Y., Suzuki, Y., Sumi, Y., Takamiya, S., Tsurumaru, Y., Nagamitsu, S., Fukai, Y., Fujii, C., Matsuoka, M., Iwanami, J., Wakabayashi, A., & Sakuta, R. (2021). Prevalence of autism spectrum disorder and autistic traits in children with anorexia nervosa and avoidant/restrictive food intake disorder. *BioPsychoSocial Medicine*, *15*(1), 9–9. https://doi.org/10.1186/s13030-021-00212-3

J. Magistretti, P., & Allaman, I. (2022). Brain energy and metabolism. In *Neuroscience in the 21st century: from basic to clinical* (pp. 2197-2227). Cham: Springer International Publishing.

James L. Hargrove, History of the Calorie in Nutrition, *The Journal of Nutrition*, Volume 136, Issue 12, December 2006, Pages 2957-2961, https://doiorg.ezproxy.waikato.ac.nz/10.1093/jn/136.12.2957

Johnston, O., Fornai, G., Cabrini, S., & Kendrick, T. (2007). Feasibility and acceptability of screening for eating disorders in primary care. Family Practice, 24, 511–517.

Juarascio, A., Kerrigan, S., Goldstein, S. P., Shaw, J., Forman, E. M., Butryn, M., & Herbert, J. D. (2013). Baseline eating disorder severity predicts response to an acceptance and commitment therapy-based group treatment. *Journal of Contextual Behavioral Science*, *2*(3-4), 74-78.

Juarascio, A., Shaw, J., Forman, E., Timko, C. A., Herbert, J., Butryn, M., Bunnell, D., Matteucci, A., & Lowe, M. (2013). Acceptance and Commitment Therapy as a Novel Treatment for Eating Disorders: An Initial Test of Efficacy and Mediation. Behavior Modification, 37(4), 459-489. https://doi-org.ezproxy.waikato.ac.nz/10.1177/0145445513478633

Kalat, & Rozin, P. (1973). “Learned safety” as a mechanism in long-delay taste-aversion learning in rats. *Journal of Comparative and Physiological Psychology*, *83*(2), 198–207. https://doi.org/10.1037/h0034424

Kambanis, P. E., Breithaupt, L., Zayas, L., Eddy, K. T., & Thomas, J. J. (2023). Cognitive-Behavioral Therapy for Avoidant/Restrictive Food Intake Disorder (CBT-AR). In *The Massachusetts General Hospital Handbook of Cognitive Behavioral Therapy* (pp. 247-258). Cham: Springer International Publishing.

Kesari, K. K., Handa, R., & Prasad, R. (2010). Effect of undernutrition on cognitive development of children. *International Journal of Food, Nutrition and Public Health*, *3*(2), 133-148.

Kimball, R. T., Greer, B. D., Fuhrman, A. M., & Lambert, J. M. (2023). Relapse and its mitigation: Toward behavioral inoculation. *Journal of applied behavior analysis*, *56*(2), 282-301.

Klok, Jakobsdottir, S., & Drent, M. L. (2007). The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obesity Reviews*, *8*(1), 21–34. https://doi.org/10.1111/j.1467-789X.2006.00270.x

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, *2*(3), 197–207. https://doi.org/10.1016/0149-7189(79)90094-6

Leaf, J. B., Leaf, R., McEachin, J., Taubman, M., Ala'i-Rosales, S., Ross, R. K., Smith, T., & Weiss, M. J. (2016). Applied Behavior Analysis is a Science and, Therefore, Progressive. *Journal of*

Luce, K. H., & Crowther, J. H. (1999). The reliability of the eating disorder examination—Self-report questionnaire version (EDE-Q). *International Journal of Eating Disorders*, *25*(3), 349-351.

Lumeng, J. C., & Hillman, K. H. (2007). Eating in larger groups increases food consumption. *Archives of disease in childhood*, *92*(5), 384-387.

Mamhidir, A. G., Ljunggren, G., Kihlgren, M., Kihlgren, A., & Wimo, A. (2006). Underweight, weight loss and related risk factors among older adults in sheltered housing-a Swedish follow-up study. *Journal Of Nutrition Health And Aging*, *10*(4), 255.

Marty, L., Chambaron, S., Nicklaus, S., & Monnery-Patris, S. (2018). Learned pleasure from eating: An opportunity to promote healthy eating in children?. *Appetite*, *120*, 265-274.

Mikami, Y., Motokawa, K., Shirobe, M., Edahiro, A., Ohara, Y., Iwasaki, M., Hayakawa, M., Watanabe, Y., Inagaki, H., Kim, H., Shinkai, S., Awata, S., & Hirano, H. (2022). Relationship between Eating Alone and Poor Appetite Using the Simplified Nutritional Appetite Questionnaire. *Nutrients*, *14*(2), 337-. https://doi.org/10.3390/nu14020337

Miller, L., Rozin, P., & Fiske, A. P. (1998). Food sharing and feeding another person suggest intimacy; two studies of American college students. European Journal of Social Psychology, 28(3), 423–436. https://doi.org/10.1002/(SICI)1099-0992 (199805/06)28:33.0.CO;2-V

Mohd Taib, N., Abdul Khaiyom, J. H., & Mohd Mahudin, (2022). Systematic Review on Eating Disorder Examination-Questionnaire 6.0 (EDE-Q 6.0): An Agenda for a Test Adaptation in Malaysia. *Jurnal Sains Kesihatan Malaysia*, *20*(2), 41–50. https://doi.org/10.17576/JSKM-2022-2002-05

Mond, J. M. Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour research and therapy*, *42*(5), 551-567

Mond, J.M., Myers, T.C., Crosby, R.D., Hay, P.J., Rodgers, B., Morgan, J.F., et al. (2008). Screening for eating disorders in primary care: EDE-Q versus SCOFF. Behaviour Research and Therapy, 46, 612–622.

Morris, M. J., Na, E. S., & Johnson, A. K. (2008). Salt craving: the psychobiology of pathogenic sodium intake. *Physiology & behavior*, *94*(5), 709-721.

Morrison, M. (2002). Using diaries in research. *Research methods in educational leadership and management*, 213-232.

Mulchandani, M., Shetty, N., Conrad, A., Muir, P., & Mah, B. (2021). Treatment of eating disorders in older people: a systematic review. *Systematic Reviews*, *10*, 1-20.

Munekata, P. E. S., Pérez-Álvarez, J. Á., Pateiro, M., Viuda-Matos, M., Fernández-López, J., & Lorenzo, J. M. (2021). Satiety from healthier and functional foods. *Trends in Food Science & Technology*, *113*, 397–410. https://doi.org/10.1016/j.tifs.2021.05.025

National Autism Center. (2015). Findings and conclusions: National standards project, phase 2.

Ogden. (2007). *Health psychology : a textbook* (4th ed.). Open University Press/McGraw-Hill Education.

Oliveira Júnior, M. L., Almeida, M., Santos, C. G., Brown, T. A., & Carvalho, P. H. B. (2023). Psychometric properties of the Eating Disorder Examination Questionnaire among Brazilian cisgender gay and bisexual adult men. *The International Journal of Eating Disorders*, *56*(4), 736–746. https://doi.org/10.1002/eat.23853

Orlandi, E., Mannucci, E., & Cuzzolaro, M. (2005). Bulimic Investigatory Test, Edinburgh (BITE). A validation study of the Italian version. *Eating and Weight Disorders*, *10*(1), e14–e20. https://doi.org/10.1007/BF03354662

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuiness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Medicine*, *18*(3), e1003583–e1003583.

Papies, E, K., van Stekelenburg, A., Smeets, M. A. M., Zandstra, E. H., & Dijksterhuis, G. B. (2022). Situating desire: Situational cues affect desire for food through eating simulations. *Appetite*, *168*, 105679–105679. https://doi.org/10.1016/j.appet.2021.105679

Pinhas, L., Morris, A., Crosby, R. D., & Katzman, D. K. (2011). Incidence and age-specific presentation of restrictive eating disorders in children: a Canadian Paediatric Surveillance Program study. *Archives of Pediatrics & Adolescent Medicine*, *165*(10), 895-899.

Pliner, P., & Hobden, K. (1992). Development of a scale to measure the trait of food neophobia in humans. *Appetite*, *19*(2), 105-120.

Potterton, R., Richards, K., Allen, K., & Schmidt, U. (2020). Eating disorders during emerging adulthood: A systematic scoping review. *Frontiers in Psychology*, *10*, 3062.

Rikani, A. A., Choudhry, Z., Choudhry, A. M., Ikram, H., Asghar, M. W., Kajal, D., ... & Mobassarah, N. J. (2013). A critique of the literature on etiology of eating disorders. *Annals of neurosciences*, *20*(4), 157.

Robinson, & Higgs, S. (2012). Liking Food Less: The Impact of Social Influence on Food Liking Evaluations in Female Students. *PloS One*, *7*(11), e48858–e48858. https://doi.org/10.1371/journal.pone.0048858

Rothman, K. J. (2008). BMI-related errors in the measurement of obesity. *International journal of obesity*, *32*(3), S56-S59.

Rozin, P. (1996). The socio-cultural context of eating and food choice. In *Food choice, acceptance and consumption* (pp. 83-104). Boston, MA: Springer US.

Schienle, Gremsl, A., & Zorjan, S. (2022). Social reward from giving food to others affects food craving and brain potentials: An imagery-based event-related potential study. *Appetite*, *168*, 105722–105722. https://doi.org/10.1016/j.appet.2021.105722

Schwartz, C., Issanchou, S., & Nicklaus, S. (2009). Developmental changes in the acceptance of the five basic tastes in the first year of life. *British Journal of Nutrition*, *102*(9), 1375–1385. https://doi.org/10.1017/S0007114509990286

Skinner, B. F. (1938). The Behavior of Organisms: An Experimental Analysis. New York, NY: Appleton-Century.

Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The Evidence-Based Practice of Applied Behavior Analysis. *Perspectives on Behavior Science*, *37*(1), 41–56. https://doi.org/10.1007/s40614-014-0005-2

Smink, FRE, van Hoeken, D, Hoek, HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr Psychiatry Rep 2012; 14: 406–14.CrossRefGoogle ScholarPubMed

Solmi, M., Wade, T. D., Byrne, S., Del Giovane, C., Fairburn, C. G., Ostinelli, E. G., De Crescenzo, F., Johnson, C., Schmidt, U., Treasure, J., Favaro, A., Zipfel, S., & Cipriani, A. (2021).

Comparative efficacy and acceptability of psychological interventions for the treatment of adult outpatients with anorexia nervosa: a systematic review and network meta-analysis. *The Lancet. Psychiatry*, *8*(3), 215–224. https://doi.org/10.1016/S2215-0366(20)30566-6

Speciani, M., Barak, Y., Damanhuri, H., De Ronchi, D., Panariello, F., & Atti, A. R. (2021). A Perspective on Chronic and Long-Lasting Anorexia Nervosa. *Frontiers in Psychiatry*, *12*, 756669–756669. https://doi.org/10.3389/fpsyt.2021.756669

Spielberger, C. D., Edwards, C. D., Montouri, J., & Lushene, R. (1973). State-trait anxiety inventory for children.

Steinhausen, H. C. (2002). The outcome of anorexia nervosa in the 20th century. *American journal of Psychiatry*, *159*(8), 1284-1293.

Stice, E. , Marti, C. N. & Rohde, P. (2013). Prevalence, Incidence, Impairment, and Course of the Proposed DSM-5 Eating Disorder Diagnoses in an 8-Year Prospective Community Study of Young Women. *Journal of Abnormal Psychology, 122* (2), 445-457. doi: 10.1037/a0030679.

Strober M, Freeman R, Morrell W. The long-term course of severe anorexia nervosa in adolescents: survival analysis of recovery, relapse, and outcome predictors over 10-15 years in a prospective study. Int J Eat Disord. (1997) 22:339–60. doi: 10.1002/(SICI)1098-108X(199712)22:4<339::AID-EAT1>3.0.CO;2-N

Strube-Lahmann, S., Müller-Werdan, U., Norman, K., Skarabis, H., & Lahmann, N. A. (2021). Underweight in Nursing Homes: Differences between Men and Women. *Gerontology (Basel)*, *67*(2), 211–219. https://doi.org/10.1159/000512459

Sulmont-Rossé, C. (2020). Eating in the elderly. *Handbook of eating and drinking: Interdisciplinary perspectives*, 433-457.

Suzuki, K., Jayasena, C. N., & Bloom, S. R. (2012). Obesity and appetite control. *Journal of Diabetes Research*, *2012*.The British Journal of Nutrition, 102 (9) (2009), pp. 1375-1385

Sysko, R., Glasofer, D. R., Hildebrandt, T., Klimek, P., Mitchell, J. E., Berg, K. C., Peterson, C. B., Wonderlich, S. A., & Walsh, B. T. (2015). The eating disorder assessment for DSM-5 (EDA-5): Development and validation of a structured interview for feeding and eating disorders. *The International Journal of Eating Disorders*, *48*(5), 452–463. https://doi.org/10.1002/eat.22388

Törmä, J., Winblad, U., Cederholm, T., & Saletti, A. (2013). Does undernutrition still prevail among nursing home residents? *Clinical Nutrition (Edinburgh, Scotland)*, *32*(4), 562–568. https://doi.org/10.1016/j.clnu.2012.10.007

US National Library of Medicine: https://medlineplus.gov/ency/article/000404.htm, accessed June 7, 2017; Medline Plus Trusted Health

Van Eeden, A. E., Van Hoeken, D., & Hoek, H. W. (2021). Incidence, prevalence and mortality of anorexia nervosa and bulimia nervosa. *Current opinion in psychiatry*, *34*(6), 515-524.

Vannucci, A., Kass, A. E., Sinton, M. M., Aspen, V., Weisman, H., Bailey, J. O., Wilfley, D. E., & Taylor, C. B. (2012). An examination of the Clinical Impairment Assessment among women at high risk for eating disorder onset. *Behaviour Research and Therapy*, *50*(6), 407–414. https://doi.org/10.1016/j.brat.2012.02.009

Vartanian, L. R., Spanos, S., Herman, C. P., & Polivy, J. (2015). Modeling of food intake: A meta-analytic review. Social Influence, 10(3), 119–136. https://doi.org/10.1080/15534510.2015.1008037

Velázquez-Alva, M. C., Irigoyen-Camacho, M. E., Cabrer-Rosales, M. F., Lazarevich, I., Arrieta-Cruz, I., Gutiérrez-Juárez, R., & Zepeda-Zepeda, M. A. (2020). Prevalence of Malnutrition and Depression in Older Adults Living in Nursing Homes in Mexico City. *Nutrients*, *12*(8), 2429-. https://doi.org/10.3390/nu12082429

Vigil Colet, Navarro González, & Morales Vives, 2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*.

What Works Clearinghouse. (2020). What Works Clearinghouse TM Procedures Handbook (version 4.1). https://ies.ed.gov/ncee/wwc/Handbooks

Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart 1. *Journal of applied behavior analysis*, *11*(2), 203-214.

Woolley, K., & Fishbach, A. (2017). A recipe for friendship: Similar food consumption promotes trust and cooperation. Journal of Consumer Psychology, 27(1), 1–10. https://doi.org/10.1016/j.jcps.2016.06.003

Appendix

Appendix A

Formula for combining dimensions on the SMRS:

Research Design (.30) + Dependent Variable (.25) + Participant Ascertainment (.20) + Procedural Integrity (.15) + Generalization and Maintenance (.10).

Appendix B

Table 5.

Title Screening IOA

Title	Me	Observer
Therapeutic alliance in two treatments for adults with severe and enduring anorexia nervosa	Yes	Yes
Dialogues about Emotional Events between Mothers with Anxiety, Depression, Anorexia Nervosa, and No Diagnosis and Their Children	No	No
Anorexia Nervosa in Juvenile Systemic Lupus Erythematosus (SLE): A Causality Dilemma	No	No
Psychological characteristics and childhood adversity of adolescents with atypical anorexia nervosa versus anorexia nervosa	Yes	No
An exploratory investigation of predictors of outcome in face-to-face and online cognitive-behavioural therapy for bulimia nervosa	Yes	No
Physiologic Transdermal Estradiol Replacement Mimics Effects of Endogenous Estrogen on Bone Outcomes in Hypoestrogenic Women with Anorexia Nervosa	No	No
Is the Activity-Based Anorexia Model a Reliable Method of Presenting Peripheral Clinical Features of Anorexia Nervosa?	Yes	No
Mindfulness in persons with anorexia nervosa and the relationships between eating disorder symptomology, anxiety and pain	Yes	No

Rules of Engagement: Qualitative Experiences of Therapeutic Alliance When Receiving In-Patient Treatment for Anorexia Nervosa	Yes	No
Assessment of Insight in Patients with a History of Anorexia Nervosa	Yes	No
Understanding the working alliance with clients diagnosed with anorexia nervosa	No	No
A pilot randomized controlled trial of a cognitive-behavioral therapy guided self-help mobile app for the post-acute treatment of anorexia nervosa: A registered report	Yes	Yes
Using a Teleconsultation-Enhanced Treatment for Avoidant/Restrictive Food Intake Disorder in an Adolescent Male	Yes	Yes
Locating the mechanisms of therapeutic agency in family-based treatment for adolescent anorexia nervosa: A pilot study of clinician/researcher perspectives	Yes	No
Childhood Avoidant/Restrictive Food Intake Disorder: Review of Treatments and a Novel Parent-Based Approach	No	No
Cost-Effectiveness of Internet-Based Cognitive-Behavioral Treatment for Bulimia Nervosa: Results of a Randomized Controlled Trial	No	No
The gut microbiome in anorexia nervosa: relevance for nutritional rehabilitation	No	No
Client attachment and therapist feelings in the treatment of bulimia nervosa	Yes	No
Creating an Opportunity to Reflect: Ear Acupuncture in Anorexia Nervosa – Inpatients' Experiences	No	No
Nutrient Intake and Dietary Inflammatory Potential in Current and Recovered Anorexia Nervosa	Yes	No
Locating the mechanisms of therapeutic agency in family-based treatment for adolescent anorexia nervosa: A pilot study of clinician/researcher perspectives	Yes	No
Change in expressed emotion and treatment outcome in adolescent anorexia nervosa	Yes	Yes

Instilling hope for a brighter future: A mixed-method mentoring support programme for individuals with and recovered from anorexia nervosa	Yes	Yes
Changes in specific and nonspecific psychopathology network structure after intensive cognitive behaviour therapy in patients with anorexia nervosa	Yes	Yes
The minority or the misunderstood? A young man's journey with anorexia nervosa	Yes	Yes
A randomized controlled trial of psychoanalytic psychotherapy or cognitive-behavioral therapy for bulimia nervosa	Yes	Yes
Liver Damage Is Related to the Degree of Being Underweight in Anorexia Nervosa and Improves Rapidly with Weight Gain	No	No
Focused vs. Broad enhanced cognitive behavioral therapy for bulimia nervosa with comorbid borderline personality: A randomized controlled trial	Yes	Yes
Couple-based interventions in the treatment of adult anorexia nervosa: A brief case example of UCAN	Yes	Yes
Predictors of outcome among young adult patients with anorexia nervosa in a randomised controlled trial	Yes	No
Findings from a couple-based open trial for adult anorexia nervosa	Yes	Yes
Therapeutic Alliance and Anorexia Nervosa Treatment Outcomes: Experiences of Young People and Their Families	Yes	Yes
Living as a couple with anorexia nervosa: A dyadic interpretative phenomenological analysis	Yes	Yes
Sensor Technology and Intelligent Systems in Anorexia Nervosa: Providing Smarter Healthcare Delivery Systems	No	No
Patients With Anorexia Nervosa Who Self-Injure: A Phenomenological Study	Yes	No
Prenotification but not envelope teaser increased response rates in a bulimia nervosa mental health literacy survey: A randomized controlled trial	No	No

How do we target the factors that maintain anorexia nervosa? A behaviour change taxonomical analysis	Yes	Yes
Multi-family therapy in anorexia nervosa—A qualitative study of parental experiences	No	No
The experience of specialist inpatient treatment for anorexia nervosa: A qualitative study from adult patients' perspectives	Yes	Yes
Predictors and moderators of treatment outcome in a randomized clinical trial for adults with symptoms of bulimia nervosa	Yes	Yes
Multidisciplinary Inflammatory Bowel Disease Care to Manage Medication Nonadherence and History of Anorexia Nervosa: A Case Report	No	No

Table 6.

Abstract Screening IOA

Title	Me	Observer
Autistic Traits Mediate Reductions in Social Attention in Adults with Anorexia Nervosa	No	No
A 10-Year-Old Girl's Dysfunctional 'Self-Help' in ADHD: Suppression of Hyperkinetic Symptoms via Self-Induced Weight Loss in the Context of Anorexia Nervosa—A Case Report	No	No
Change in expressed emotion and treatment outcome in adolescent anorexia nervosa	No	No
Cognitive-behavioral therapy for avoidant/restrictive food intake disorder: Feasibility, acceptability, and proof-of-concept for children and adolescents	Yes	Yes
Changes in meal-related anxiety predict treatment outcomes in an intensive family-based treatment program for adolescents with anorexia nervosa	No	No
Predictors of Weight Restoration in a Day-Treatment Program that Supports Family-Based Treatment for Adolescents with Anorexia Nervosa	No	No

The experience of specialist inpatient treatment for anorexia nervosa: A qualitative study from adult patients' perspectives	No	No
Self-esteem as a catalyst for change in adolescent inpatients with anorexia nervosa: a pilot randomised controlled trial	No	No
Cognitive behavioral therapy for bulimia nervosa	No	No
A 1-year follow-up study of the longitudinal interplay between emotion dysregulation and childhood trauma in the treatment of anorexia nervosa	No	Yes
Assessment of Insight in Patients with a History of Anorexia Nervosa	No	No
'Learning to live your life again': An interpretative phenomenological analysis of weblogs documenting the inside experience of recovering from anorexia nervosa	No	No
Predictors of outcome among young adult patients with anorexia nervosa in a randomised controlled trial	No	No
Restoring normal eating behaviour in adolescents with anorexia nervosa: A video analysis of nursing interventions	No	No
Enhanced cognitive behavior therapy for severe and extreme anorexia nervosa: An outpatient case series	Yes	Yes
Effectiveness of intensive cognitive behavioral therapy in adolescents and adults with anorexia nervosa	Yes	Yes
Assessment and management of anorexia nervosa during COVID-19	No	No
Internet-delivered cognitive-behavioural therapy v. conventional guided self-help for bulimia nervosa: long-term evaluation of a randomised controlled trial	No	No
Using a Teleconsultation-Enhanced Treatment for Avoidant/Restrictive Food Intake Disorder in an Adolescent Male	Yes	Yes
Hurry Up and "Weight": Innovative Inter-Professional Outpatient Care for a Service User With Severe Anorexia Nervosa Awaiting Admission to (Specialized Elective) Inpatient Care in Canada	No	No

Subjective evaluation of outpatient treatment for adolescent patients with anorexia nervosa	No	No
The relationship of body image with symptoms of depression and anxiety in patients with anorexia nervosa during outpatient psychotherapy: Results of the ANTOP study	No	No
Psychological characteristics and childhood adversity of adolescents with atypical anorexia nervosa versus anorexia nervosa	No	No

Table 7.

Full Text Screening IOA

Title	Me	Observer
Enhanced cognitive behavioral therapy for adolescents with anorexia nervosa: Outcomes and predictors of change in a real-world setting	Yes	Yes
A single-arm pilot study of guided self-help treatment based cognitive behavioral therapy for bulimia nervosa in Japanese clinical settings	Yes	Yes
Inpatient cognitive behaviour therapy for anorexia nervosa: a randomized controlled trial	No	No
A new cognitive behavior therapy for adolescents with avoidant/restrictive food intake disorder in a day treatment setting: A clinical case series	Yes	Yes