

The effectiveness of a domain specific self-esteem
group intervention: a pilot study

Ciping Goh

D.Clin.Psy. thesis (Volume 1), 2018
University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:



Name: Ciping Goh

Date: 22 June 2018

Overview

Volume one of this thesis is presented in three parts. Part one presents a systematic review of domain specific self-esteem measures for adults. A total of 13 papers evaluating 10 domain specific self-esteem measures were reviewed. Results indicated a general lack of psychometric robustness of measures in literature. Future research should focus on the continued validation of these measures.

Part two is an empirical study which investigated the effectiveness of a domain specific self-esteem group intervention developed by Hollingdale (2015). The empirical study was conducted jointly with Emily Dixon. Results indicated an improvement in domain specific self-esteem (i.e. perceived competence) in valued domains. The discrepancy between perceived competence and importance placed in valued domains decreased after the intervention. Attributional styles towards negative events showed a shift towards more external, unstable and specific styles post-intervention and continued moving towards this direction at the one-month follow up. A significant relationship was found between domain specific self-esteem and attributional styles towards negative events. Overall, the domain specific self-esteem group is a promising intervention for self-esteem that requires further study.

Part three provides a critical appraisal of the systematic review and empirical study. It begins with a discussion about definitional and measurement issues faced when conducting the systematic review. The commentary about the empirical study includes reflections on the theoretical issues, measurement issues, group experience and challenges faced. It concludes with a reflection on implementing the domain specific self-esteem group within the National Health Service (NHS).

Impact Statement

Low self-esteem has been found to be an aetiological factor in a variety of mental health diagnoses, such as depression, anxiety, psychosis, eating disorders and obsessive compulsive disorder (Waite, McManus & Shafran, 2012). Difficulties with self-esteem and the impact it has on individuals' mental health and wellbeing are commonly seen in clinical practice (Fennell, 1997). It is therefore critical to develop and evaluate effective treatments for improving self-esteem.

There is extensive evidence in research supporting the domain-specific nature of self-esteem (Byrne, 1996). Domain specific self-esteem refers to an individual's self-appraisals within circumscribed domains, for example, intellect and athleticism (Harter, 2012). Individuals therefore may hold different levels of self-esteem in various domains. Hollingdale (2015) developed the unsatisfactory self-esteem model that considers domain specific self-esteem to be on a spectrum that at times can become "unsatisfactory" for an individual's needs. This depends on their preferred level of functioning, within a specific domain, situation or period in their life.

The present research consisted of two aims. The first was to conduct a systematic review of existing domain specific self-esteem measures for adults in literature. Although some identified measures showed promise in terms of their psychometric properties, notable weaknesses in study methodology and psychometric properties were also found. Future research should therefore focus on the continued validation of these measures. In mental health services, the use of domain specific self-esteem measures would allow clients and therapists to identify domains of focus for therapeutic interventions that are clinically meaningful to work on. Moreover, the developmental stages (e.g. adolescent, adult, older adult) that

some domain specific self-esteem measures account for might be relevant for specific populations in clinical settings. The further development of domain specific self-esteem measures would have real practical implications for how self-esteem is treated in the clinical setting.

The second aim of the present research was to develop and pilot the first known domain specific self-esteem group intervention. The research was conducted as an uncontrolled trial. The four-session Cognitive Behavioural Therapy (CBT) group intervention was based on the unsatisfactory self-esteem model developed by Hollingdale (2015) with the aim of improving domain specific self-esteem. The intervention consisted of the following components: charting individual domain specific self-esteem profiles; identifying valued domains; and employing various CBT techniques such as thought diaries and behavioural experiments.

The brief intervention provided preliminary evidence of clinical benefits such as improvements in domain specific self-esteem and attributional styles towards negative events. Moreover, the treatment gains appeared to be durable, at least for a month after the intervention ended. Clinically, the current group intervention allows clinicians to identify idiosyncratic life domains in which it would be meaningful to intervene in. This is likely to improve clients' engagement in the intervention and increase motivation for change. Moreover, a transdiagnostic intervention, such as the present group intervention might be beneficial for use across multiple mental health diagnoses. Future research should focus on evaluating the present group intervention in a controlled trial and also investigate how the intervention might complement evidenced based therapies that are utilised in local mental health services.

Table of Contents

Acknowledgements	9
Part 1: Literature Review	10
Abstract	11
Introduction	12
Method	18
Results	25
Discussion.....	58
References	68
Part 2: Empirical Paper	85
Abstract	86
Introduction	88
Method	100
Results	109
Discussion.....	118
References	132
Part 3: Critical Appraisal	143
Introduction	144
Reflections on the Literature Review.....	144

Reflections on the Research Paper.....	146
References	159
Appendices.....	162

Tables and Figures

Part 1: Literature Review

Figure 1	20
Table 1	23
Table 2	24
Table 3	26
Table 4	28
Table 5	55
Table 6	57

Part 2: Empirical Paper

Figure 1	89
Figure 2	96
Figure 3	97
Figure 4	110
Table 1	111
Table 2	112

Acknowledgements

It gives me great pleasure to acknowledge the guidance and support given to me by my supervisors Henry Clements and Sue Watson, who continually supported and encouraged me throughout the research process. They were a great help to hurdle the obstacles in the completion of this thesis. I would also like to thank to Emily Dixon for helping to make this joint project such a positive experience. It was a pleasure working with her throughout the entire research process.

I would like to acknowledge Jack Hollingdale, who conceptualised the self-esteem model and group intervention used in the empirical paper. It was a privilege to work with him and take his ideas forward.

I would also like to express my gratitude to Joshua Stott for his guidance with my systematic review and research data analysis. I am thankful for his willingness to help despite his tight schedule. I appreciate the time and expertise offered by Rob Saunders, Ravi Das, and John King for the data analysis portion of the empirical study. I also appreciate Sunjeev Kamboj for his help with the study design.

My gratitude goes out to the people who volunteered their time to participate in this research, for which this study would not have been possible.

Finally, my appreciation goes out to my fiancée and family who have journeyed with me over this period and showered me with numerous encouragements along the way. Above all, this thesis would not have been possible without my Lord and Saviour Jesus Christ, who has given me grace and strength to accomplish this. All glory goes to Him.

Part 1: Literature Review

A Systematic Review of Domain Specific Self-Esteem Measures for Adults

Abstract

Aim. Self-esteem is one of the most studied topics in research and clinical literature. Therefore, the adequate measurement of the construct is crucial. The objective of the present literature review was to summarise and critically appraise the quality of the measurement properties of instruments that measure domain specific self-esteem in adults.

Methods. A range of databases were searched, and articles were selected if their primary purpose was the development or assessment of measurement properties of instruments measuring domain specific self-esteem in adults. Methodological quality was assessed using the COnsensus based Standards for the selection of health Measurement INstruments checklist (COSMIN).

Results. A total of 13 papers evaluating 10 domain specific self-esteem measures met the inclusion criteria. Overall, the Self-Image Profile for Adults (SIP-AD) had more evidence for its psychometric properties compared to other instruments. However, all the measures examined were found to have notable weaknesses. The analysis found a general lack of psychometric robustness in the measures used in current literature.

Conclusion. Future research should focus on the continued validation of these measures, while bearing in mind the complexities around measuring the construct of domain specific self-esteem.

Introduction

Self-esteem is one of the most studied topics in the social sciences and psychology (Mruk, 2006). Despite the amount of research in this area, there is still considerable debate on the definition of self-esteem. Researchers have defined self-esteem as a need (Maslow, 1954), an attitude (Coopersmith, 1967), a belief in one's competence (James, 1980) and being good enough as a person (Rosenberg, 1989). Irrespective of the definitional debate, the construct of self-esteem has been extensively examined in the literature, with many studies indicating links between the construct and various outcomes. High self-esteem is linked to coping with life stresses and achieving more in life (Coopersmith, 1967; Harter, 1990). It is also associated with greater autonomy, sense of mastery, positive relations with others and self-acceptance (Paradise & Kernis, 2002). In contrast, low self-esteem has been identified as an aetiological factor in a variety of mental health conditions such as depression (Brown, Bifulco, & Andrews, 1990), anxiety (O'Brien, Bartoletti, & Leitzel, 2006; Watson, Suls, & Haig, 2002), psychosis (Hall & Tarrier, 2003) and eating disorders (Gual, Perez-Gaspar, Martinez-Gonzalez, Lahortiga, & Cervera-Enguix, 2002). Taking into account the continued interest in self-esteem in research and clinical literature, the adequate measurement of the construct is critical.

A Note on Terminology: Self-Esteem and Self-Concept

One complexity in defining self-esteem is its conflation with self-concept. Most researchers seem to agree that while self-concept implies a broader definition of the construct that includes cognitive, affective, and behavioural aspects, self-esteem is thought to be a more evaluative component of the broader self-concept term (Blascovich & Tomaka, 1991; Wells & Marwell, 1976). However, Shavelson,

Hubner and Stanton (1976) argued that self-concept is both descriptive and evaluative. Therefore, self-concept measures that include statements such as “I am good at mathematics” and “I can run a long way without stopping” have both evaluative and descriptive components (Marsh & Mara, 2008). Typical self-esteem and self-concept instruments consist of items that elicit both descriptive and evaluative components of the self, therefore making it almost impossible to separate these two constructs (Byrne, 1996). Accordingly, it is common for most researchers to use the two terms interchangeably (Hughes, 1984; Shavelson et al., 1976). The complexity of teasing these two constructs apart arguably warrants a separate discussion. Therefore, consistent with other reviews of self-esteem measures (e.g. Byrne, 1996), this review will use both ‘self-esteem’ and ‘self-concept’ terms in the search and evaluation of instruments.

Conceptualising Self-Esteem: Global or Domain Specific?

Harter (1990, p. 292) wisely reflected on the importance of not putting the “methodological cart before the conceptual horse”. Already, past reviews of self-esteem research have found the lack of theoretical basis and the poor quality of measurement instruments in many studies (e.g. Burns, 1979; Shavelson et al., 1976; Wells & Marwell, 1976; Wylie, 1974, 1979). Byrne (1996) noted that variations in definitions of self-esteem at the conceptual level have led to methodological differences at the measurement level. One important issue in self-esteem research is whether self-esteem is conceptualised as a global or a multidimensional concept (Hertherton & Wyland, 2003). Most self-esteem research has traditionally considered the construct as a global concept, that is, an individual’s global evaluation of oneself

that is stable across time and situations (Rosenberg, 1989). Most research examining self-esteem has utilised self-report scales that measure global evaluations of the self.

However, since the 1980s, there is now a wealth of evidence supporting the multidimensional nature of self-esteem (Byrne, 1996). Byrne (1984, p. 427) conducted an extensive review of construct validation research of self-esteem and concluded that self-esteem is indeed “a multidimensional construct, having one general construct and several specific facets”. Furthermore, Marsh and Craven (2006, p. 191) argued that “If the role of self-concept research is to better understand the complexity of self in different contexts, to predict a wide variety of behaviours, to provide outcome measures for diverse interventions, and to relate self-concept to other constructs, then the specific domains of self-concept are more useful than a general domain”.

Significant support for the multidimensional aspect of self-esteem has been evidenced in literature (e.g. Byrne, 1984; Harter, 1985; Marsh 1986; Marsh & Shavelson, 1985). This perspective has emphasised that self-evaluation can happen with respect to specific domains, such as physical appearance and morality (e.g. Harter 1985; Marsh 1986; Marsh & Shavelson 1985). Such evaluations are often referred to as domain specific self-esteem (Donnellan, Trzesniewski, & Robins, 2015). Individuals therefore may hold different levels of self-esteem in various domains (Mruk, 2006). Marsh (1986) identified that domain-specific and global self-esteem shared associations of .06 to .60, suggesting that these constructs were related but not interchangeable. Marsh (1993) also found that academic outcomes were substantially related to academic self-esteem but unrelated to global self-esteem.

Similarly, Marsh and Peart (1988) found that a physical fitness intervention was related to physical self-concept but uncorrelated with nonphysical self-concepts.

Measurement of Domain Specific Self-Esteem

A search by Sheff and Fearon (2004) suggested a figure of 200 measures for self-esteem in the current literature that are conventionally assessed with self-report scales. However, Wylie (1974) suggested that most measures tended to be short lived and of debatable quality. Moreover, most peer reviewed publications used a relatively small set of commonly used self-esteem measures (Donnellan et al., 2015). With the vast amount of self-esteem measures available in the research literature, helpful reviews have been conducted to examine the psychometric properties of some of these measures. Notably, Byrne (1996) conducted an extensive review of 24 self-esteem measures that were categorised into various age groups across the lifespan that included child, adolescent and adult measures. Blascovich and Tomaka (1991) reviewed 11 measures of self-esteem in the literature and Heatherton and Wyland (2003) reviewed three measures. Donnellan et al. (2015) updated and extended previous findings on self-esteem measures by Blascovich and Tomaka (1991) and Heatherton and Wyland (2003); five self-esteem measures were examined in the review.

Most of the instruments reviewed were global self-esteem measures such as the Rosenberg Self-Esteem Scale, (RSES; Rosenberg, 1965) the Coopersmith Self-Esteem Scale (Coopersmith, 1967), the Janis-Field Feelings of Inadequacy Scale (Eagly, 1967; Fleming & Courtney, 1984; Janis & Field, 1969) and the Texas Social Behaviour Inventory (Helmreich, Stapp & Ervin, 1974). None of the reviews conducted have previously focused solely on evaluating domain specific self-esteem

measures. Harter (2012) argued that while most theorists recognise that the self-concept is multidimensional, most measures do not adequately capture this complexity. Only a minority of domain specific self-esteem measures were evaluated in some of the reviews, such as the Tennessee Self-Concept Scale (Fitts, 1965; Roid & Fitts, 1988), the series of Self-Perception Profiles (e.g. Harter, 1985; Harter, 1988) and the series of Self-Description Questionnaires (e.g., Marsh, 1989; Marsh, 1992a; Marsh, 1992b). Despite the growing consensus for the multidimensional perspective of self-esteem, there is a dearth of reviews of instruments which claim to measure it. Therefore, the present review will focus on the evaluation of domain-specific self-esteem measures.

Measurement of Domain Specific Self-Esteem in Adults

While most reviews in the past have focused on self-esteem measures in general, some have focused on self-esteem measures in the context of lifespan development. A review by Butler and Gasson (2005) examined self-esteem and self-concept scales for children and adolescents and evaluated the 14 most frequently cited instruments in the literature. Hughes (1984) did a similar review with 19 most frequently used scales for self-esteem in children aged 3-12 years. Finally, Davis-Kean and Sandler (2001) conducted a meta-analysis of measures for self-esteem for young children in preschool and elementary school.

Adulthood brings developmental changes. Individuals in this age range are typically no longer dependent upon their parents and take on more enduring responsibilities that include career development, financial independence and, for some, marriage and parenting (Harter, 1992). Erikson (1959) identified in his work on psychosocial life stages that this period proceeds from an individual's identity

formation and begins with the exploration of relationships and life opportunities that lead the individual to develop a sense of a meaningful life lived. Harter (1992) added that the period of adulthood also centres on the experimentation in various vocational and occupational opportunities, the development of new friendships, intimate relationships, and the renegotiation of the parent-child bond and of belief systems such as religious, political and moral identifications.

Previous research in self-esteem changes across the lifespan have generally found small, gradual increases in self-esteem across adulthood and a decrease in older adulthood (e.g. Orth, Trzesniewski & Robins, 2010; Galambos, Barker, & Krahn, 2006). Orth et al. (2010) conducted a cohort-sequential longitudinal study exploring self-esteem changes and found that self-esteem followed a quadratic trajectory across the adult life span. It was noted that self-esteem increased during young and middle adulthood, reached a peak at about age 60 years, and then declined in old age (Orth et al., 2010). Moreover, there is evidence that points to self-concept dimensions becoming more differentiated from mid-adolescence (Marsh & Shavelson, 1985). Therefore, global self-esteem measures alone would be limited in capturing the complexity of the adult self.

With the complexity of interactions between adulthood and self-esteem, the need for psychometrically sound instruments that measure and adequately reflect the adult self is crucial. A systematic review using a level of evidence approach has not been previously conducted for domain specific self-esteem measures for adults. Such a review involves the systematic ranking of studies based on the rigour of their methods and ensures that recommendations are made based on studies that are

methodologically sound (Park, Reilly-Spong & Gross, 2013), thereby improving the reliability and validity of findings.

Therefore, the aim of the present review was to summarise and critically appraise the quality of instruments that measure domain specific self-esteem in adults.

Method

Search Strategy

Studies were identified from the following electronic databases: Ovid Medline®, PsycINFO®, Health and Psychological Instruments® and Embase®. The search was: ((domain* or multidimension* or multi-dimension* or hierarch* or facet* or multifacet* or multi-facet*) adj3 (self-esteem or self-concept or self-image or self-perception)). The search terms for self-esteem were adopted from a previous review of self-esteem measures by Byrne (1996). The limits for the search were (i) human; (ii) English language, (iii) Tests and measures.

Selection Criteria

Articles were selected if their primary purpose was to develop or evaluate the measurement properties of a domain specific self-esteem instrument. Only studies that had a measure of domain specific self-esteem administered to adults with a sample mean of 18 years old and above, were included. Only articles that were in the English language were included. Articles were excluded if they measured only one domain of self-esteem, for example a physical self-esteem measure which measures self-esteem solely around physical attributes, or if they measured self-esteem only in a specific population, such as adults with autism, or teachers. Articles were also excluded if the main aim was to test the efficacy of an intervention for the treatment

of self-esteem. The rationale for excluding efficacy studies was detailed by De Vet, Terwee, Mokkink and Knol (2011), who concluded that these studies normally provide indirect evidence of the measurement properties of an instrument.

In addition, Byrne (1996) had already extensively reviewed the psychometric properties for the following domain specific self-esteem measures: Tennessee Self-Concept Scale (TSCS; Fitts, 1965; Roid & Fitts, 1988), the Self Perception Profile for Adults (SPP-A, Messer & Harter, 2012), the Self Perception Profile for College Students (SPP-CS; Neemann & Harter, 2012) and the Self Description Questionnaire III (SDQ III; Marsh, 1989). The studies by Messer and Harter (2012) and Neemann and Harter (2012) were revisions of the original instrument manuals in 1986. Therefore, only articles from 1996 to the present that examined the psychometric properties of these measures were included in the present review.

The titles and abstracts retrieved in the search were initially screened to select the included articles. The full text of articles was assessed for inclusion. The steps involved in identifying and selecting the studies are illustrated in Figure 1. The results from the four databases were combined, identifying a total of 1079 papers. After removing duplicates, 700 articles were identified. The titles and abstracts were screened and articles were excluded based on the relevance of the titles. Following this, 54 full text papers were retrieved. At this point, Byrne (1996) who conducted a review on measures of self-esteem additionally searched the databases for each individual instrument that the papers identified to identify additional papers. For the current search, domain specific self-esteem measures were identified from the papers and searched in the databases. An additional 17 papers were included based on their titles and abstracts. In total, 71 papers were read in full and compared against the

inclusion and exclusion criteria. From these, 13 papers met the inclusion criteria and formed the set of papers for the current review.

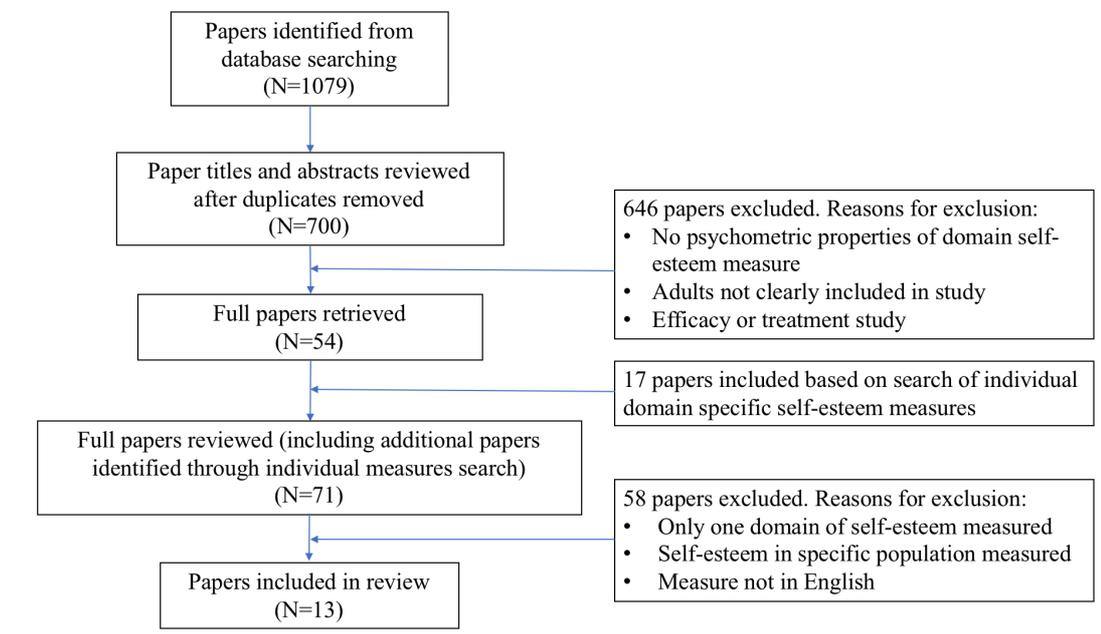


Figure 1. Flow chart of search and selection process

Measurement Properties

The COSMIN taxonomy identifies three domains to assess psychometric properties: reliability, validity, and responsiveness (Mokkink et al., 2010). Reliability refers to the degree to which the instrument is free from measurement error and comprises of three sub-classifications: internal consistency, measurement error, and reliability. Internal consistency signifies the degree of interrelatedness among the items. The measurement error is the systematic and random error that is not attributed to true changes in the construct. Reliability refers to the proportion of the total variance due to true differences among persons.

Validity refers to the extent that the instrument measures the construct(s) it claims to measure. This is further categorised into content validity, construct validity

and criterion validity. Content validity is the extent to which the measure sufficiently reflects every single element of a construct and comprises face validity. Construct validity comprises structural validity, hypothesis testing and cross-cultural validity. Structural validity is the proof that supports the dimensionality of the measure. Hypothesis testing refers to the extent that the measure relates with other constructs which coincide with expectations. Cross-cultural validity refers to the extent to which translated or adapted versions of the measure perform in accordance with the original measure. Cross-cultural validity was not evaluated in the current review as the review did not include papers that examined adapted or translated versions of measures. Criterion validity is the degree that a measure correlates with an accepted ‘gold standard’. Finally, responsiveness is the ability of the measure to detect clinically significant changes in the construct over time. Evaluating responsiveness was beyond the scope of the current review as domain specific self-esteem would need to be measured with two measures at two time points, using one as the ‘true measure’ and evaluate responsiveness against this.

The COSMIN checklist and assessment of instrument quality

The included papers were evaluated for study quality utilising the 4-point scale COSMIN checklist as a guide (Mokkink et al., 2010). Study quality of each measurement property was rated as: *excellent*, *good*, *fair* or *poor*. The “worst score counts” algorithm was used, meaning that the final quality rating for a property is the lowest rating of any item relating to that property. Instruments were evaluated separately for articles that included more than one instrument.

Study results were also assessed and given a *positive*, *negative* or *indeterminate* rating for each measurement property. Criteria for these ratings (Table 1) are outlined by Terwee et al. (2007) and Park, Reilly-Spong and Gross (2013).

Best evidence synthesis

Finally, Table 2 presents the criteria used when combining the results from the study methodological quality and the study result measurement property ratings.

Table 1

Quality criteria for assessment of measurement properties adapted from Terwee et al. (2007) and Park et al. (2013)

Property	Rating	Quality Criteria
Internal Consistency	+	Factor analyses performed on adequate sample size (7* #items and ≥ 100) AND Cronbach's alpha(s) ≥ 0.70
	?	No factor analysis OR doubtful design or method
	-	No information on internal consistency
Reliability	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's $r \geq 0.80$
	?	Neither ICC/weighted Kappa, nor Pearson's r determined
	-	ICC/weighted Kappa < 0.70 OR Pearson's $r < 0.80$
Content Validity	+	A clear description of measurement aims, constructs to be measured and item selection. The target population AND investigators/experts involved in item selection.
	?	A clear description of the measure is lacking. No target population involvement.
	-	The target population considers items on the questionnaire to be incomplete/no information found on target population.
Structural Validity	+	Factors explain at least 50% of the variance OR good or adequate fit (see goodness-of-fit criteria for CFA or EFA ^a)
	?	Explained variance not mentioned OR equivocal fit by goodness-of-fit criteria for CFA or EFA ^a
	-	Factors explain $< 50\%$ of the variance OR poor fit by goodness-of-fit criteria for a CFA or EFA ^a
Hypothesis testing	+	Correlation with an instrument measuring the same construct $\geq 50\%$ but 75% of the results in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs.
	?	Solely correlations determined with unrelated constructs OR $\geq 50\%$ but $< 75\%$ of the results are in accordance with the hypotheses
	-	Correlation with an instrument measuring the same construct < 0.50 OR $< 50\%$ of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs.

^a Good or adequate fit: comparative fit index (CFI) ≥ 0.90 , root mean square of approximation (RMSEA) ≤ 0.08 , standardized root means square residual (SRMR) < 0.10 . Inadequate fit: CFI ≤ 0.85 , RMSEA ≥ 0.10 , SRMR ≥ 0.10 ; Indeterminate fit: the values of the fit indexes ranged in between the adequate criteria and inadequate criteria.

Table 2

Levels of evidence for the overall quality of the measurement properties

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of <i>good</i> methodological quality OR in one study of <i>excellent</i> methodological quality.
Moderate	++ or --	Consistent findings in multiple studies of <i>fair</i> methodological quality OR in one study of <i>good</i> methodological quality.
Limited	+ or -	One study of <i>fair</i> methodological quality.
Conflicting	±	Conflicting findings from studies of comparable quality
Indeterminate	?	Findings from <i>excellent, good or fair</i> studies were not definitively positive or negative
None	na	Findings from <i>excellent, good or fair</i> studies were not available

Table from Park et al. (2013) was used.

+positive result; -negative result; ±both positive and negative findings have been reported by studies of adequate quality; ? findings from studies of adequate quality were not definitively positive or negative; na findings from studies of adequate quality were not available.

Results

A total of 10 domain specific self-esteem measures were evaluated from across the 13 papers. Table 3 describes the characteristics of the studies included in the review. Table 4 details each instrument and provides examples of the items in each instrument. Methodological quality ratings for each study are presented in Table 5; each measurement property is given a rating of either *excellent*, *good*, *fair*, or *poor* based on the COSMIN quality ratings.

Table 6 presents the overall level of evidence synthesis for the instrument's measurement properties; this combines the ratings of methodological quality and study result measurement property ratings. Detailed findings for each measure reviewed are provided in this section.

Table 3

Characteristics of included studies

Study Authors	Population	Sample Size	Age, mean (SD)	Female %	Country
Addeo, Greene and Geisser (1994)	University Students	307	20.55 (SD not reported)	67.8	United States
Bagozzi and Heatherton (1994)	University Students	Sample 1: 102	22.0 (5.2)	50.7	United States
		Sample 2: 428	20.3 (4.3)	66.4	
Bishop, Walling and Walker (1997)	Medical and Nursing Faculty Members	201	Not reported	100	United States
Butler and Gasson (2006)	Primarily Caucasian	1462	17-65	68.2	United Kingdom
Goñi, Madariaga, Axpe, and Goñi (2011)	University Students	1135	30.17 (14.81)	60.1	Spain
Harter and Kreinik (2014)	Primarily Caucasian (i.e., European-American)	203	74.5 (5.69)	59	United States
Heatherton and Polivy (1991)	University Students	Study 1: 428	20.3 (4.3)	66.4	Canada
		Study 2: 102	22.0 (5.2)	70.2	
		Study 3: 128	Not reported	77.3	
		Study 4: 79	Not reported	100	

	Obese Women	Study 5: 18	41.3 (11.2)	100	
McCain, Jonason, Foster and Campbell (2015)	Predominantly European American Students	544	20.25 (4.70)	69	
Rinn and Cunningham (2008)	High Ability & Average University Students	100 High Ability & 196 Average Ability	18.97 (1.00)	70	Australia
			19.26 (1.42)	77	
Robson (1989)	Patients with anxiety or psychotherapy referrals and Controls (adults)	70	Controls: 31 (9.0) Anxiety: 35 (10.6) Psychotherapy: 33 (9.9)	49 76 52	United Kingdom
Stake (1994)	University Students	1665	Majority 18-21	Not reported	United States
Waugh (2001)	University Students	400	Not reported	Not reported	Australia
Yanico and Lu (2000)	University Students	185	24.4 (5.2)	Not reported	United States

Table 4

Characteristics and item examples from the included instruments

Instrument	Construct assessed	Recall period	Dimensions (number of items)	Number of subscales	Response options (range)	Ease of scoring and administration (range of scores)	Sample items
Self-Perception Profile for Older Adults (SPP-OA; Harter & Kreinik, 2014)	Domain Self-Esteem and Global Self-Esteem	None	Relationships with friends (6) Family relationships (6) Nurturance (6) Adequacy as a provider (6) Job competence (6) Cognitive abilities (6) Household management (6) Leisure activities (6) Health status (6) Physical Appearance (6) Morality (6) Global Self-Esteem (6) Life Satisfaction (6) Reminiscence (6)	11 Domain Self-Esteem Subscales and 3 Global Dimension Subscales	Structured Alternative Format	Moderate	“Some adults are very satisfied with the friendships they have formed BUT Other adults are somewhat disappointed that they have not formed friendships that are more rewarding” “Some adults are somewhat disappointed about their job performance BUT Other adults are very satisfied with how they have performed at their job(s)”

							“Some adults are generally happy being the way they are BUT Other adults would like to be different”
Self Image Profile for Adults (SIP-AD; Butler & Gasson, 2006)	Domain Self-Image, Domain Self-Esteem and Domain Self-Certainty	None	SIP - Consideration SIP - Social SIP - Moral SIP - Competence SIP - Physical SIP - Outlook Self-Image Self-Esteem Self-Satisfaction Self-certainty -ve Self-certainty +ve (30 Items in total)	6 Domain and 5 General scales		Easy	Not provided
Personal Self-Concept Questionnaire (PSCQ; Goñi & Fernández, 2007)	Domains of Self-Concept	None	Self-fulfilment (6) Honesty (3) Autonomy (4) Emotional self-concept (5)	4 Domain subscales of self-concept	5 point scale (1 = totally disagree, 5 = totally agree).	Easy	“So far, I have achieved every important goal I have set myself.” “I have yet to achieve anything I consider to be important in my life”
Six- Factor Self-Concept	Domains of Self-Concept	None	Power (7) Task Accomplishment (6)	6 Subscales	Scale from 1 (never or	Easy	Not Provided

Scale (SFSCS; Stake, 1994)			Giftedness (5) Vulnerability (6) Likeability (6) Morality (6)		almost never true of you) to 7(always or almost always true of you).		
State Self Esteem Scale (SSES; Heatherton & Polivy, 1991)	Domains of State Self-Esteem	What you are thinking at the moment	Performance (7) Social (7) Appearance (6)		5 point scale (1 = not at all to 5 = extremely)	Easy	“I feel satisfied with the way my body looks right now.” “I feel as smart as others.” “I am worried about looking foolish.”
Tennessee Self-Concept Scale (TSCS; Fitts, 1965; Roid & Fitts, 1988)			Physical Moral Personal Family Social Identity Satisfaction Behaviour (90 items in total)	5 external aspects of self-concept, 3 internal aspects, and 15 “facets”	5 point scale (1 = completely false, 5 = completely true).		Not Provided
Self-Perception Profile for College Students (SPP-CS; Scholastic Competence	Domain Self-Esteem and Global Self-Esteem	None	Creativity (4) Intellectual Ability (4) Scholastic Competence (4) Job Competence (4) Athletic Competence (4) Appearance (4)	12 Domain Self-Esteem Subscales and 1 Global	Structured Alternative Format	Moderate	“Some students don’t feel that they are very athletic BUT Other students do feel they are athletic” “Some students have

Subscale; Neemann & Harter; 2012			Romantic Relationships (4) Social Acceptance (4) Close Friendships (4) Parent Relationships (4) Finding Humour in One's Life (4) Morality (4) Global Self-Esteem (5)	Dimension Subscale		the ability to develop romantic relationships BUT Other students do not find it easy to develop romantic relationships"
Self- Description Questionnaire III (SDQ III; General Academic Subscale; Marsh & O'Neill, 1984)	Domain Self- Concept and General Self- Concept	None	Math Self-Concept Verbal Self-Concept General Academic Self- Concept Problem Solving Self-Concept Physical Ability Physical Appearance Relations with the Same Sex Relations with the Opposite Sex Relations with Parents Spiritual Values/Religion Honesty/Trustworthiness Emotional Stability General Self-Concept (All scales have 10 or 12 items)	12 Domain Scales and 1 Global Scale	8 point scale, the response options varying from "1-Definitely False" to "8- Definitely True	"I often tell small lies to avoid embarrassing situations" "I have a physically attractive body" "I wish I had more imagination and originality"
Robson Self- Esteem Questionnaire			Significance (5) Worthiness (5) Appearance/ social	7 domains with a total Self-	8 point scale from 0=strongly	Easy "I am not embarrassed to let people know my opinions."

(RSEQ; Robson 1989)			acceptability (5) Resilience and determination (5) Competence (4) Control over personal destiny (4) Value of existence (2)	Esteem Score	disagree to 7=strongly agree	“There are lots of things I'd change about myself if I could.” “I look awful these days.”
Self-Concept Questionnaire (SCQ; Waugh, 2001)	Domains of Self-Concept	None	Capability (10) Perceptions of Achievement (10) Confidence in Academic Life (10) Relationships with Peers of same sex (10) Relationships with Peers of opposite sex (10) Relationships with Family (10) Personal Confidence (10) Physical Self-Concept (10) Honesty/Trustworthiness (10)	9 Domains of Self-Concept	Rate how you would like to be and how you believe that you actually are on a scale of 3 (All the time or nearly all the time) to 0 (None of the time or almost none of the time)	“Having persons of my age and sex enjoy my company” “Treated fairly by my family”

1. The Self Perception Profile for Older Adults (SPP-OA; Harter & Kreinik, 2014)

Description of the measure

The SPP-OA is an 84-item domain specific self-esteem measure for older adults that was validated in a sample of primarily Caucasian older adults with varying educational levels and previous occupational statuses. The measure builds on the theoretical assumption that perceptions of the self reflect multidimensional, specific domains of one's life, as well as a separate domain of global self-worth (Harter, 1992). The SPP-OA was developed because of the lack of self-esteem measures available for older adults. The measure consists of 11 domains of self-esteem and three global indices. These are: Relationships with friends, Family relationships, Nurturance, Adequacy as a provider, Job competence, Cognitive abilities, Household management, Leisure activities, Health status, Physical appearance, Morality, Global Self-Esteem, Life satisfaction and Reminiscence. The study sampled 203 older adults (aged 65-89) that were recruited from senior centres and community centres serving older adults, as well as from newspaper advertisements. The SPP-OA quality ratings were reported for internal consistency, content validity, structural validity and hypothesis testing.

Internal consistency was rated as *intermediate* and the methodology as *fair*. Based on a sample of 203 older adults, Cronbach's alpha was reported for the 11 subscales that ranged from 0.75 to 0.86 (Harter & Kreinik, 2014). However, the sample size was inadequate in the analysis, therefore only meeting the *fair* methodology quality criteria.

In terms of content validity, ratings for the SPP-OA quality were *intermediate* and the methodology was *poor*. Items were generated by the study authors that aimed to be developmentally sensitive to the period of older adulthood. However, an assessment of whether all items were relevant to the construct measured and whether they were relevant to the study population was not conducted. The process of item selection and reduction was also not reported in the study.

Structural validity was rated as *intermediate* in a study of *poor* methodological quality. An Exploratory Factor Analysis (EFA) was performed using an oblique rotation, with the justification provided that the domains were likely to be inter-correlated (Harter & Kreinik, 2014). Findings from the analysis yielded an 11-factor structure that reflected the structure of the initially hypothesised self-esteem domains. However, the methodological quality was rated as *poor* due to the small sample size as compared to the number of variables. For factor analyses, rules of thumb vary between a subject-to-variables ratio of 4:1 to 10:1, with a minimum of 100 subjects (Kline, 1993).

Hypothesis testing was rated as *positive* in a study of *fair* methodological quality. The cognitive competence domain was predicted and found to correlate with educational attainment. A strong relationship between perceived physical appearance and global self-esteem was also predicted and found. Predictions around gender differences with respect to domains such as health status and physical appearance were found.

Levels of evidence conclusions

Overall, the SPP-OA had *limited* evidence for hypothesis testing validity. There was only *intermediate* evidence for its internal consistency due to the small

sample size. Evidence for content validity and structural validity were of *poor* quality and therefore these findings were given no weight in this final synthesis. There is paucity in the psychometric findings reported that reveals the critical need for Confirmatory Factor Analysis (CFA) to confirm the factor structure of the measure. Nonetheless, the SPP-OA fills an important gap in the literature of current measures designed to tap domain specific self-esteem in older adults.

2. Self-Image Profile for Adults (SIP-AD; Butler & Gasson, 2006)

Description of the measure

The SIP-AD is a 30-item instrument designed to measure self-image and self-esteem. While self-esteem relates to an evaluative aspect of self, self-image refers to descriptive characteristics available to an individual in defining the self (Butler & Gasson, 2005). The SIP-AD taps into six aspects of the self that comprise: Consideration, Social, Moral, Competence, Physical and Outlook.

The SIP-AD was designed to address methodological issues such as the lack of a distinct theoretical stance in some measures (Butler & Gasson, 2006). The instrument is grounded in personal construct theory, where items were selected based on commonly used self-descriptions (Bannister & Fransella, 1986; Butler & Green, 1998; Kelly, 1955), and the developmental and organizational model of the self proposed by Harter (1999). Butler and Gasson (2006) also argued that previous studies included geographically limited samples, with little correspondence with a national census that therefore created problems with generalisation. The current study sampled 1462 British adults from a variety of backgrounds across the age groups of 17-65 years. The SIP-AD quality ratings have been reported for the

following measurement properties: internal consistency, content validity, structural validity and hypothesis testing.

Internal consistency itself was rated *intermediate* and the methodological quality was rated *poor*, as there were no subscale internal consistencies reported. Cronbach's alpha for total self-image score was 0.90.

Content validity was rated *positive* and the methods used were *excellent*. A pool of initially developed items was subjected to examination by a sample of 1303 adults to ensure familiarity and meaningfulness of the items (Butler & Gasson, 2006). Item selection and reduction was conducted. An additional sub-sample of males and females were recruited to ensure validity of items.

Structural validity was rated *positive* while the methodology used was rated *fair*. Principal Component Analysis (PCA) was used to determine the number of factors following a varimax rotation, which revealed six clear factors.

Acknowledging the problem of inflated loadings, as well as other extensively documented limitations associated with principal components analyses (e.g., Gorsuch, 1990; Hubbard & Allen, 1987; Snook & Gorsuch, 1989), an EFA would have been more appropriate for the factor analysis. CFA was conducted and found support for the six-factor structure, albeit one item that failed to load on the 'competence' factor. However, the CFA was conducted on the same sample as the PCA, which is not recommended when examining structural validity.

Hypothesis testing was rated as *positive* and the methods used were rated *good*. The hypothesis testing validity was tested through the examination of the associations of the measure with other well-known self-esteem measures. Self-image and self-esteem scores were correlated with both the Rosenberg Self-Esteem Scale

(RSES; Rosenberg, 1965) and the Tennessee Self-Concept Scale-2 (TSCS-2; Fitts & Warren, 1996).

Levels of evidence conclusions

The SIP-AD is a soundly constructed measure of domain specific self-esteem that had *strong* evidence for its content validity as it is linked with a sound theoretical framework. Evidence for hypothesis testing validity was *moderate*. However, the SID-AD had *limited* evidence for its structural validity and could benefit from the validation of its factorial structure by means of using a new sample to conduct CFA. Evidence for internal consistency was of *poor* quality and therefore the findings were given no weight in this final synthesis.

3. Personal Self-Concept Questionnaire (PSCQ; Goñi & Fernández, 2007)

Description of the measure

The PSCQ is a 22-item measure that taps four domains of self-concept: Self-fulfilment, Autonomy, Honesty and Emotional self-concept. The PSCQ aims to measure personal self-concept which refers to the more specific, individual or private aspects of oneself, as opposed to a more external aspect of the self, such as the social self-concept (Goñi, Madariaga, Axpe, & Goñi, 2011). Based on this conceptual model, the researchers charted the development of research that investigated personal self-concept and developed the PSCQ to tap into these domains. Goñi and Fernández (2007) had previously conducted a study to ascertain its internal reliability and factor structure. Another replication study was conducted by Goñi (2009) on a broader sample group to determine its psychometric properties. Unfortunately, these papers were in the Spanish language and were not included in the review. The current study was conducted as a follow-up to determine if the empirical data confirmed the

structure of the PSCQ that was proposed by the earlier studies. A Confirmatory Factor Analysis (CFA) was conducted on a sample of 559 participants between the ages of 15 and 65 years old. The PSCQ quality ratings have been reported for its structural validity.

Structural validity of the PSCQ was rated *positive* and the methodology used was rated *good*. The goodness of fit for three models were tested and the four interrelated factors model had Root Mean Square Error of Approximation (RMSEA) =.071, Confirmatory Fit Index (CFI) =.94, Standardized Root Mean Residual (SRMR) =.06, which all met the criteria for good fit. The sample size included in the analysis was also deemed appropriate.

Levels of evidence conclusions

Overall, the PSCQ had *moderate* evidence for its structural validity, given that there was only one study of *good* methodological quality. Findings from this study provide support for the construct validity of the PSCQ and built on the earlier findings by Goñi (2009) and Goñi and Fernández (2007).

4. Six-Factor Self-Concept Scale (SFSCS; Stake, 1994)

Description of the measure

The SFSCS is a 36-item multidimensional measure of adult self-concept conceptualised to have broad applicability across life settings, roles and activities (Stake, 1994). The measure was created to measure domain specific self-esteem at a mid-level specificity: that meant that it aimed to provide maximum generalisability across situations (e.g. work, relationships etc.) with maximum distinctiveness in categories (Rosch, 1978). The authors sought to adopt these categories from previously evidenced multidimensional self-esteem measures, such as the Social

Self-Esteem Scale (SSES; Stake, 1985) and the Performance Self-Esteem Scale (PSES; Stake, 1979). The SFSCS consisted of six subscales: Likability, Morality, Task Accomplishment, Giftedness, Power and Vulnerability.

Two papers were examined that explored the reliability and validity of the measure (Stake, 1994; Yanico & Lu, 2000). Stake (1994) developed and validated the measure on a predominantly Euro-American sample comprising of 476 undergraduate students and 365 non-university participants. Yanico and Lu (2000) explored the psychometric properties of the measure in a sample of 185 undergraduate ethnic minority women. The SFSCS quality ratings have been reported for the following measurement properties: internal consistency, reliability, content validity, structural validity and hypothesis testing.

In the study by Stake (1994), internal consistency was rated *positive* in the undergraduate sample but *intermediate* in the non-university sample. The Cronbach's alphas for all the subscales in the undergraduate sample were above 0.70. However, in the non-university sample, the Cronbach's alphas of three subscales (Morality, Task Accomplishment, Vulnerability) were below 0.70. In the study by Yanico and Lu (2000), internal consistency was rated *positive*; Cronbach alphas ranged from 0.76 to 0.86. The methodology used for internal consistency was rated *good* in both the papers examined.

Reliability was rated as *intermediate* and the methods used were rated as *fair* in the study by Stake (1994), as not all reliability coefficients met the *positive* rating criteria of $r \geq .80$. The test-retest reliability coefficients in a sample of 57 undergraduates over a period of four weeks ranged from 0.74 to 0.88; test-retest

reliability in 61 undergraduates over a period of six weeks ranged from 0.72 to 0.85 (Stake 1994).

Content validity was rated as *intermediate* and the methods used were rated as *fair* in the study by Stake (1994). An initial pool of 115 items were originally referenced from the Social Self-Esteem Scale (Stake, 1985) and Performance Self-Esteem Scale (Stake, 1979). Additional items were added from other self-concept questionnaires and research studies. Items were then given to undergraduate students to examine face validity. However, the relevance of the items was not assessed in the target general adult population.

In the study by Stake (1994), structural validity was rated *negative* in the undergraduate sample but *positive* in the non-university sample. The methodology was rated as *fair*. EFA in the undergraduate sample revealed a six-factor structure but the factors only accounted for 48% of the total variance and failed to meet the positive rating criteria ($\geq 50\%$) for quality in the review. However, EFA in the non-university sample revealed a six-factor solution and the factors accounted for 54% of the total variance. CFA was also conducted on the non-university sample and the chi-square/degrees of freedom ratio was (1246/579) 2.15, which indicated a good fit with the six-factor model. However, methodologically, using the chi-square as the only indicator of a model's goodness of fit is inappropriate due to its sensitivity to sample size. In fact, to carry out a thorough assessment of a model's fit, it is essential to adopt a holistic approach which includes other indices which currently exist, such as the CFI and RMSEA (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

In the study by Yanico and Lu (2000), structural validity was rated *negative* and the method was rated *fair*. A PCA was conducted in the ethnic minority women

sample (Yanico & Lu, 2000), and the results replicated the six-factor structure of the instrument, but two items (*Strong* and *Law Abiding*) had failed to load adequately on their respective subscales (*Power* and *Morality*). Moreover, the factors only accounted for 47% of the common variance and failed to meet the positive rating criteria for quality in the review. In this instance, a CFA approach instead should have been used to validate the structure of the SFSCS.

Hypothesis testing was rated as *positive* and the methods used were rated as *fair* in both studies (Stake, 1994; Yanico & Lu, 2000). Stake (1994) reported evidence for convergent and discriminant validity of the instrument. Consistent with the predicted relationships, the SFSCS correlated highly with self-esteem (RSES; Rosenberg, 1965). Yanico and Lu (2000) replicated the findings on a ethnic minority female sample and found that the SFSCS also correlated with the RSES (Rosenberg, 1965). Stake (1994) found the measure to correlate with wellbeing (Monge Wellbeing Scale; Monge, 1973). In addition, the SFSCS correlated with memories of childhood behaviour and current life events and behaviours.

Levels of evidence conclusions

Overall, the SFSCS had *moderate* evidence for its hypothesis testing validity. There was *conflicting* evidence for its structural validity. It had *intermediate* evidence for its internal consistency, content validity and reliability. In particular, internal consistency and structural validity studies of the instrument in samples representative of the adult population are needed before the psychometric soundness of its internal consistency and structure can be judged appropriately. Further studies that replicate the factor structure using CFA are required with the provision of adjustment indices such as the CFI and RMSEA.

5. State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991)

Description of the measure

The SSES is a 20-item self-esteem scale that measures short-lived changes in self-esteem in three domains: Performance, Social and Appearance self-esteem. State self-esteem was conceptualised by research indicating that self-esteem showed momentary fluctuations across situations, although the fluctuations did not seem to be large (e.g. Croker & Major, 1989; Gergen, 1971; Markus & Kunda, 1986; Rosenberg, 1986; Wells, 1988). The SSES was developed with the purpose of measuring these momentary changes in domain specific self-esteem. Three papers were examined that explored the reliability and validity of the measure (Bagozzi & Heatherton, 1994; Heatherton & Polivy, 1991; McCain, Jonason, Foster, & Campbell, 2015). Heatherton and Polivy (1991) examined the reliability and validity of the SSES in a sample of 428 undergraduate students. Bagozzi and Heatherton (1994) and McCain et al. (2015) examined only the structural validity of the measure. Bagozzi and Heatherton (1994) examined the measure in two samples of university students (Sample 1, $n = 102$; Sample 2, $n = 428$). McCain et al. (2015) examined the measure in a sample of 544 university students. The SSES quality ratings have been reported for the following measurement properties: internal consistency, content validity, structural validity and hypothesis testing.

Internal consistency was rated as *intermediate* and the methods used were rated as *poor* in the study by Heatherton and Polivy (1991). While the overall scale

had high internal consistency of Cronbach alpha 0.92 in a sample of 428 undergraduates, internal consistency was not calculated for each subscale.

Content validity was rated as *intermediate* and the methods used were rated as *poor* in the study by Heatherton and Polivy (1991). There was a clear description of the aim of the measure and the authors utilised items from another validated measure of self-esteem, the Janis-Field Feelings of Inadequacy Scale (JFS; Pliner, Chaiken, & Flett, 1990; Fleming & Courtney, 1984) to develop the SSES. However, there was no study population involved in the item selection process.

Structural validity was rated as *positive* for the studies (Bagozzi & Hetherington, 1991; Heatherton & Polivy, 1991). However, structural validity was rated as *negative* for the study by McCain et al. (2015). The methodology for all three papers was rated as *good*. Heatherton and Polivy (1991) conducted an EFA and the factors in the final model of a three-factor solution accounted for 50.4% of the overall variance. Bagozzi and Heatherton (1994) conducted a CFA, and the results supported the findings of Heatherton and Polivy (1991) for a three-factor solution. The Relative Non-Centrality Index (RNI) scores for both models, which are comparable to the CFI, was 0.90. However, McCain et al. (2015) conducted a CFA and found support instead for a bi-factor model that fit the data tolerably better with CFI = 0.87, RMSEA = 0.08 compared to the three-factor solution with CFI = 0.73, RMSEA = 0.11.

Hypothesis testing was rated as *positive* and the methods used were rated as *fair*. The measure was correlated with the JFS and Rosenberg Self-Esteem Scale (RSES) as predicted. The SSES was also predicted and found to be highly related to anxiety (State Anxiety subscale of the State-Trait Anxiety Inventory; STAI;

Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983) and depression (Beck Depression Inventory; BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Evidence for its discriminant validity was also found (Heatherton & Polivy, 1991).

Levels of evidence conclusions

Overall, the SSES showed *conflicting* evidence for its structural validity and *limited* evidence for its hypothesis testing validity. Further CFA is required to ascertain whether a bifactor or three-factor solution is more adequate. Evidence for internal consistency and content validity were of *poor* quality and therefore the findings were given no weight in this final synthesis.

Research is essential to determine the reason for the inconsistencies to better understand the factor structure of the SSES. Moreover, internal consistency of the subscales requires further examination as they were not identified in either of the papers evaluated. Finally, support for content validity of the measure was lacking.

6. Tennessee Self-Concept Scale (TSCS; Fitts, 1965; Roid & Fitts, 1988)

Description of the measure

The TSCS (Fitts, 1965; Roid & Fitts, 1988) is a widely used measure for self-concept in both research and clinical settings (Donnellan et al., 2015). The TSCS is theoretically grounded in a taxonomic model of self-esteem where there are three self-concept facets and each with two or more levels. The first facet includes five levels of the external frame-of-reference facet. These include physical, moral, personal, family, and social self-concepts. Each of these traits have three internal frames of reference (the second facet): identity (e.g., “what I am”), satisfaction (e.g., “how I feel about myself”), and behaviour (e.g., “what I do, or how I act”) (Fitts, 1965; Roid & Fitts, 1988). Identity refers to the private internal self; satisfaction

represents the gap between the actual and ideal self; and behaviour reflects the external observable self (Fitts, 1965; Roid & Fitts, 1988, 1996). Finally, the third facet involves the wording of items which serves the purpose of controlling response bias (Fitts, 1965; Roid & Fitts, 1988). The measure consists of 90 items and is divided into eight subscales: Physical, Moral, Personal, Family, Social, Identity, Satisfaction and Behaviour self-concept. Bishop, Walling and Walker (1997) aimed to validate the factor structure of the TSCS after a failure by Tzeng, Maxey, Fortier, and Landis (1985) to do so. A convenience sample of 111 female medical and nursing faculty members from a university was used. The TSCS quality ratings have been reported for the following measurement properties: internal consistency and structural validity.

Internal consistency was rated as *intermediate* and the methods used were rated as *poor*. The overall scale had high internal consistency, Cronbach alpha was 0.92. However, alphas were not calculated for each subscale.

Structural validity was rated as *intermediate* and the methods used were rated as *poor*. The study conducted an PCA and was unable to replicate the factor structure originally proposed by Roid and Fitts (1988). The three level self-concept structure was not replicated in the study and neither was there evidence for the sublevels (Bishop et al., 1997). However, the use of CFA that allows for testing of a priori factor structures to validate the TSCS would have been more appropriate. The sample size did not meet the required subject-to-variables ratio and was comprised of females only. Therefore, the methodology was rated as *poor*.

Levels of evidence conclusions

Evidence for internal consistency and structural validity of the TSCS in the study examined were of *poor* quality and therefore the findings were given no weight in this final synthesis. While the current study points at a failure to replicate results of previous validation studies of the TSCS, methodological robustness of the examined study was lacking.

7. SPP-College Students (SPP-CS; Scholastic Competence Subscale; Neemann & Harter; 2012)

Description of the measure

The SPP-CS (Neemann & Harter; 2012) is a 54-item questionnaire that evaluates self-esteem in the following 12 domains and one measure of global self-worth in university students: creativity, intellectual ability, scholastic competence, job competence, athletic competence, appearance, romantic relationships, social acceptance, close friendships, parent relationships, humour and morality. Similar to the SPP-OA, the SPP-CS is founded on the theoretical basis that perceptions of the self reflect multidimensional, specific domains of one's life, as well as a separate domain of global self-worth (Harter, 1992). In the current study, Rinn and Cunningham (2008) examined the appropriateness of using the Scholastic Competence subscale of the SPP-CS in a sample of 100 high-ability and 196 average-ability university students. The SPP-CS quality ratings have been reported for the following measurement properties: internal consistency and hypothesis testing.

Internal consistency was rated as *good* and the methods used were rated as *good*. Cronbach's alpha was 0.76 among high-ability students and 0.77 among average-ability students. The methodology was rated as *good* although no factor

analysis was performed as the paper referred to other studies where factor analyses were performed (e.g. Neemann & Harter, 2012).

Hypothesis testing was rated as *intermediate* and the methods used were rated as *fair*. With average-ability students, the study found moderate correlations between scholastic competence and American College Testing (ACT) scores (a test around English, mathematics, reading, and science reasoning skills) and the students' Grade Point Average (GPA). With high-ability students, it was unexpected that scholastic competence did not correlate with students' GPA but did so with ACT scores. This was possibly attributed to the lack of variability in GPA scores, whereas ACT scores provided more variability. No correlations were found with students' aspirations or year in college. However, the study methodology was rated as *fair* as prior hypotheses were not formulated clearly at the beginning, but rather assumed.

Levels of evidence conclusions

Overall, evidence for the internal consistency was *moderate* and hypothesis testing is *intermediate* for the Scholastic Competence domain of the SPP-CS in this study.

8. Self-Description Questionnaire III (SDQ III; General Academic Subscale; Marsh & O'Neill, 1984)

Description of the measure

The SDQ III (Marsh & O'Neill, 1984) is a measure of self-concept that contains 136 items measuring 12 domains of self-concept and one general self-concept score. The theoretical basis postulates that general self-concept is a higher order factor that comprises multiple, domain-specific self-concepts, which, although correlated, can be interpreted as separate constructs (Marsh & O'Neill, 1984). This

theoretical model is based on the Shavelson model (Shavelson et al., 1976) which has undergone extensive construct validation (Byrne, 1996). Shavelson et al. (1976) proposed that multidimensional self-concept is composed of four first-order facets, each with additional second-order facets: physical self-concept (physical ability, physical appearance); social self-concept (relations with the same sex, relations with the opposite sex, relations with parents); emotional self-concept (spiritual values/religion, honesty/trustworthiness, emotional stability); and academic self-concept (math, verbal, general academic and problem-solving).

In the current study, Rinn and Cunningham (2008) examined the appropriateness of using the General Academic subscale of the SDQ III in a sample of high-ability and average-ability college students. The SDQ III quality ratings have been reported for the following measurement properties: internal consistency and hypothesis testing.

Internal consistency was rated as *good* and the methods used were rated as *good*. Cronbach's alpha was 0.84 among high-ability students and 0.88 among average-ability students. The methodology was rated as *good* although no factor analysis was performed as the paper referred to other studies where factor analyses were performed (e.g. Marsh, 1989; Marsh & O'Neill, 1984).

Hypothesis testing was rated as *intermediate* and the methods used were rated as *fair*. The findings were similar to those for the SPP-CS that was used in the same study. With average-ability students, the study found moderate correlations between scholastic competence and ACT scores and the students' GPA. With high-ability students, it was unexpected that scholastic competence did not correlate with students' GPA but did so with ACT scores. This again was attributed to the lack of

variability in GPA scores, as compared to ACT scores which provided more variability. No correlations were found with students' aspirations or year in college. However, the study methodology was rated as *fair* as prior hypotheses were not formulated clearly at the beginning, but rather assumed.

Levels of evidence conclusions

Overall, evidence for the internal consistency was *moderate* and hypothesis testing is *intermediate* for the General Academic subscale of the SDQ III in this study.

9. Robson Self-Esteem Questionnaire (RSEQ; Robson 1989)

Description of the measure

The RSEQ (Robson, 1989) is a 30-item self-esteem measure based on a multidimensional model of self-esteem. Based on the definition that self-esteem is “the sense of contentment and self acceptance that results from a person's appraisal of his own worth, significance, attractiveness, competence, and ability to satisfy his aspirations” (Robson, 1989, p. 514), seven components of self-esteem were defined: the subjective sense of significance; worthiness; appearance and social acceptability; competence; resilience and determination; control over personal destiny; and the value of existence.

Two papers explored the reliability and validity of the measure (Addeo, Greene, & Geisser, 1994; Robson, 1989). Robson (1989) developed and validated the RSEQ in three samples that included 51 outpatients with Generalised Anxiety Disorder (GAD), 47 patients undergoing psychotherapy and a control group of 70 adults with no evidence of psychological disorder. Addeo et al. (1994) examined the structural validity hypotheses testing of the measure in a sample of 307

undergraduate students. The RSEQ quality ratings have been reported for the following measurement properties: internal consistency, reliability, structural validity, content validity and hypothesis testing.

In the study by Robson (1989), internal consistency was rated as *intermediate* and the methods used were rated as *poor* in the study by Robson (1989). The overall Cronbach's Alpha was 0.89, but no factor analysis was conducted in the study. In the study by Addeo et al., (1994), internal consistency was rated as *intermediate* and the methods used were rated as *good*. Addeo et al. (1994) performed a factor analysis and found three factors, Self-Depreciation, Attractiveness and Self-Respect. While Cronbach's Alpha was 0.90 for the overall scale, the factor reliability of the subscales were 0.85, 0.21 and 0.68 respectively. Therefore, only internal consistency for one of the subscales (Self-Depreciation) was good.

Reliability was rated as *good* but the methods used were rated as *poor* in the study by Robson (1989). Correlation of overall scores was 0.87 across the a 4-week interval. However, a small sample size of only 21 university students was examined, which only met the *poor* methodological rating.

Content validity was rated as *positive* and the methods used were rated as *fair* in the study by Robson (1989). Although measurement aims were clearly described and constructs to be measured were adequately elaborated upon, there was no elaboration of the sample characteristics or the sample size used to norm the measure.

Structural validity was rated as *negative* and the methods used were rated as *good* in the study by Addeo et al. (1994). It should be noted that Robson (1989) mentioned that a factor analysis was conducted but the analysis and results were not

reported in the paper. Addeo et al (1994) later reported the results that Robson (personal communication, 1991) conducted a maximum likelihood factor analysis with oblique and equamax rotation and found five factors. Addeo et al. (1994) conducted a CFA using an oblique rotation on a sample of 307 undergraduate students and found that a three-factor model had a better fit than the five-factor model. However, the three-factor model only accounted for 33.1% of the variance and other the fit indices were not presented.

Hypothesis testing was rated as *positive* and the methods used were rated as *fair* for the study by Robson (1989), due to the inadequate sample size used in some analyses. The RSEQ correlated strongly with the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) and self-evaluations of self-esteem as expected. The measure also negatively correlated with depression scores on the Beck Depression Inventory (BDI; Beck et al., 1961), patients with anxiety, and patients referred for psychotherapy as predicted. Hypothesis testing was rated as *positive* and the methods used were rated as *good* for the study by Addeo et al., (1994). Replication by Addeo et al. (1994) also found that the RSEQ correlated with the RSES. In addition, the RSEQ was positively correlated with global self-efficacy, social self-efficacy, trait curiosity, and negatively correlated with anxiety and depression as predicted.

Levels of evidence conclusions

Overall, the RSEQ had *moderate* evidence for the hypothesis testing validity. It had *limited* support for its content validity. However, it had *conflicting* evidence for its structural validity and *intermediate* evidence for its internal consistency. The evidence for reliability was of *poor* quality and therefore those findings were given

no weight in this final synthesis. Further examination for its factor structure and its internal consistency in an adult population is particularly needed given the conclusions.

10. Self-Concept Questionnaire (SCQ; Waugh, 2001)

Description of the measure

The SCQ (Waugh, 2001) is a 90-item questionnaire that is theoretically based on a similar model to the Shavelson et al. (1976) model. Domain-specific self-concept consisted of three first order facets, each with three second-order facets: academic self-concept (capability, achievement and confidence), social self-concept (same-sex peer, opposite-sex peer, and family) and self-concept presentation of self (personal confidence, physical and honest/trustworthy). The items were developed based on evidence provided by Bracken (1996), Hattie (1992), Marsh (1992a), Marsh (1992b) and Marsh and Hattie (1996) for the Shavelson et al (1976) model. The questionnaire consists of 45 items involving a ‘how I would like to be’ self-concept and 45 items corresponding to ‘how I actually am’ self-concept. A convenience sample of 400 university students was used for the study which applied the Item Response Theory (IRT) in the development and evaluation of the measure.

In the COSMIN analysis for IRT, methodology was rated similarly to the Classical Test Theory counterparts (*excellent, good, fair or poor*) but quality ratings were not provided due to the lack of quality ratings for IRT study findings. However, a discussion of the levels of evidence conclusion is made below. The general methodological requirements met by the IRT model was rated as *good*. The extended logistic model of Rasch was used with the computer programme Rasch Unidimensional Measurement Models (RUMM) to analyse the data. The Rasch

model estimates a common discrimination parameter for all items, and it is advantageous due to its parsimony (Edelen & Reeve, 2007). The IRT model and software of was adequately described in the paper. The sample size of 400 was deemed adequate as Edelen and Reeve (2007) recommended a sample size of 100 subjects for Rasch models.

However, the study only partly checked the assumptions for estimating parameters of the IRT model. One important assumption of unidimensional parametric IRT models is that the construct being measured is in fact unidimensional (Edelen & Reeve, 2007). This is usually done through an item factor analysis in IRT methodology, but this was not conducted in the paper examined.

Internal consistency methodology was rated as *fair*, as the unidimensionality of the measure was not checked. The person separation index is used instead of reliability indices (e.g. Cronbach alphas) in Rasch models. This referred to the proportion of observed variance considered to be true. The Index of Person Separability values for the 45-item and 66-item scales were .945 and .946, respectively, meaning that the proportion of observed variance considered to be true is 94% in each scale.

Structural validity methodology was also rated as *fair*, as the unidimensionality of the measure was not checked. From the results, 24 items from the 90-item scale did not fit the model. These items were removed and the 66 items scale (consisting of 45 'how I actually am' plus the remaining 21 'how I would like to be' items) and the 45-item scale (consisting of only 45 'how I actually am' items) and fit the model well. The responses were consistent and logical with the order response format used. The item-trait interaction was significant at 445 and 307 for

the 66-item and 45-item scale respectively, indicating good agreement for all items across participants with differing self-concept. There was good consistency of response patterns from the item and person fit statistic. Finally, the power of the tests of fit was found to be “Excellent”. The analysis found that items were not as well targeted to students with higher self-concept. Therefore, items targeted specifically for this could be added to the measure.

Levels of evidence conclusions

Overall, the SCQ had *limited* evidence for its internal consistency and structural validity, due to methodological limitations. Nonetheless, the analysis supported the conceptual framework of self-concept that is based on a multifaceted, hierarchical model with first order and second order facets. The measure had sound theoretical foundations that gives evidence for its content validity. The Rasch analysis provided support for the fit of the model for the 45-item and 66-item scales as a valid measurement tool for multidimensional self-concept, supporting its structural validity. The value added through the IRT analysis was the detailed item-level information that differentiated individuals with lower and higher self-concept.

However, a major limitation of the study is that no unidimensional analysis was conducted, which is an important assumption in IRT models. Choosing the Rasch model also has other limitations; although the model increases unidimensionality of a scale, there might be a decrease in validity. This is because it uses the requirements of measurement to model the data instead of choosing a model that fits with the data (Andrich, 1989). The Rasch approach also rejects items that do not fit the measurement criteria, which might also result in a loss of validity. Studies examining these assumptions with respect to the measure are required.

Table 5

Methodological quality of each study per measurement property and instrument

Measure	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Criterion Validity	Hypothesis testing
Self-Perception Profile for Older Adults (SPP-OA)							
Harter and Kreinik, (2014)	Fair			Poor	Poor		Fair
Self-Image Profile for Adults (SIP-AD)							
Butler and Gasson, (2006)	Poor			Excellent	Fair		Good
Personal Self-Concept (PSC) Questionnaire							
Goñi et al. (2011)					Good		
Six Factor Self-Concept Scale (SFSCS)							
Stake (1994)	Good	Fair		Fair	Fair		Fair
Yanico and Lu (2000)	Good				Fair		Fair
State Self-Esteem Scale (SSES)							
Heatherton and Polivy (1991)	Poor			Poor	Good		Fair
Bagozzi and Heatherton (1994)					Good		

Measure	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Criterion Validity	Hypothesis testing
McCain et al (2015) Tennessee Self-Concept Questionnaire (TSCS)					Good		
Bishop et al. (1997) SPP- College Students (Scholastic Competence Subscale)	Poor				Poor		
Rinn and Cunningham (2008) SDQ III (General Academic Subscale)	Good						Fair
Rinn and Cunningham (2008) Robson Self-Esteem Questionnaire (RSEQ)	Good						Fair
Robson (1989)	Poor	Poor		Fair			Fair
Addeo et al. (1994) Self-Concept Questionnaire (SCQ)	Good				Good		Good
Waugh (2001)	Fair				Fair		

Table 6

Levels of evidence synthesis: Quality of measurement properties per instrument

Instrument	Internal consistency	Reliability	Content Validity	Structural Validity	Hypothesis testing
SPP-OA	?	na	na	na	+
SID-AD	na	na	+++	+	++
PSCQ	na	na	na	++	na
SFSCS	?	?	?	±	++
SSES	na	na	na	±	+
TSCS	na	na	na	na	na
SPP-CS (Scholastic Competence Subscale)	++	na	na	na	?
SDQ III (General Academic Subscale)	++	na	na	na	?
RSEQ	?	na	+	±	++
SCQ	+	na	na	+	na

See Table 2 for levels of evidence descriptors

Discussion

The aim of this paper was to systematically review the psychometric properties of instruments that measure domain specific self-esteem in adults. The COSMIN checklist was used to evaluate the measurement properties and the methodological quality of 10 instruments. Internal consistency, reliability, content validity, structural validity, criterion validity and hypothesis testing were assessed across the included studies.

Summary of findings

Overall, it was encouraging that the findings using the COSMIN checklist suggest that most studies had at least *fair* methodological quality. However, the measures reviewed were not without flaws. In terms of psychometric properties alone, it was difficult to conclude which measure was the most suitable as all the measures examined showed strengths and weaknesses. Moreover, there were relatively few replication studies conducted on the measures examined and therefore *moderate* evidence was the highest level most measures could reach if they were evaluated in only one *good* quality study.

The SIP-AD had relatively more evidence than its counterparts for its psychometric properties. It had *strong* evidence for content validity and *moderate* evidence for hypothesis testing validity. The PSCQ had *moderate* evidence for its structural validity and this provided support for the earlier studies to confirm the measure's factor structure. The SPP-CS (Scholastic Competence Subscale) and SDQ III (General Academic Subscale) had *moderate* evidence for internal consistency. The SFSCS and RSEQ had *moderate* evidence for hypothesis testing validity.

The remaining instrument properties provided *limited, intermediate* or *conflicting* evidence.

Some reasons are hypothesised for the lack of evidence for these instruments. The methodological quality of many of the studies in this review was compromised by the problem of inadequate reporting of methodology, such as the lack of reporting Cronbach alpha for subscales (e.g. Bishop et al., 1997; Butler & Gasson, 2006; Heatherton & Polivy, 1991) or reporting appropriate goodness of fit indices in EFA or CFA (e.g. Butler & Gasson, 2006; Harter & Kreinik, 2014; Stake, 1994; Yanico & Lu, 2000 etc.).

In addition, some studies reviewed here had content validity methodology that was poor or lacking (e.g. Robson, 1994; Stake, 1994). This was because they had not included the general population of adults in the process of item testing and selection, although they had suggested that the measure should tap self-esteem in this population.

Most of the studies reviewed here used convenience sampling methods with university students. Therefore, there appears to be a genuine need to validate these measures on adult samples that are representative of the general adult population.

Generally, there was a lack of evidence for reliability analysis in most of the papers reviewed here. Test-retest reliability to examine whether scores changed under repeated measurements was absent or the methodology was rated as poor (e.g. Robson, 1994).

Test development and norming

While most of the studies reviewed were published in the United States with American norms and a few published in Australia, only the study by Butler and Gasson (2006) examined and normed their measure, the Self-Image Profile for

Adults (SIP-AD), in a British population. This review found some evidence for its psychometric properties as a measure of domain specific self-esteem in the general British adult population. Moreover, it was encouraging that the instrument had been tested on a large British adult sample, which contributes to the measure's generalisability.

While most of the measures had used Classical Test Theory (CTT) in the development and validation of measures, Waugh (2001) employed IRT methods for development of the SCQ. IRT methodology might serve to provide rich item level information as captured by the study by Waugh (2001). While it is beyond the scope of the review to examine the differences between CTT and IRT in detail, the value of IRT analysis seems to be gaining presence in psychological test development (e.g. Zanon, Hutz, Yoo, & Hambleton, 2016) due to its advantages (Embretson, 1996; Hambleton, Robin & Xing, 2000).

Other Relevant Measures

This review also provides a summary of the psychometric properties of the instruments examined here that Byrne (1996) had already reviewed. They include the Self-Description Questionnaire III (SDQ III; Marsh, 1989), the Self Perception Profile for College Students (SPP-CS) and the Tennessee Self-Concept Scale (TSCS; Fitts, 1965; Roid & Fitts, 1988).

The results of the present review add to Byrne's (1996) review of the SDQ III. Marsh (1989) examined the internal consistency of the Self-Description Questionnaire III (SDQ III) and found that Cronbach's alpha ranged from 0.76 to 0.95 on the subscales. The findings in the present review builds on this by providing support for the internal consistency of the General Academic subscale. Byrne (1996) also indicated that both EFA and CFA had been conducted on the instrument with

results revealing strong factor structures, with each of the 13 subscales being clear. Marsh and Richards (1988) had also found evidence of its convergent validity with the TSCS (Fitts, 1965; Roid & Fitts, 1988). Finally, strong evidence was found for the test-retest reliability of the SDQ III (Marsh, Richards, & Barnes, 1986).

This review examined the Scholastic Competence subscale of the SPP-CS and also adds to the findings made by Byrne (1996). The internal consistencies for the SPP-CS 12 subscales ranged from 0.76 to 0.92. The findings in the present review builds on this by providing support for the internal consistency of the Scholastic Competence subscale. In terms of its structural validity, Neemann and Harter (1986) conducted a PCA and the results suggested a 12-factor structure. However, cross loadings were not reported in the study. Moreover, given the widely recognized limitations associated with principal components analyses (e.g., Gorsuch, 1990; Hubbard & Allen, 1987; Snook & Gorsuch, 1989), the findings should be interpreted with caution. Crocker and Ellsworth (1990) examined five subscales of the measure and found support for the factorial structures and internal consistency. Overall, the instrument requires further evidence to establish its psychometric properties.

Byrne (1996) reviewed the TSCS and reported that Roid and Fitts (1994) had established the TSCS as a reliable and valid measure. They had conducted EFA, CFA, as well correlations with theoretical models of self-concept and personality scales (Byrne, 1996). Byrne (1996) therefore concluded that the TSCS had established itself as a sound measure. Although the paper included in the present review suggested a failure to replicate the factor structure of the TSCS, the findings were given no weight due to the *poor* methodological quality. It should be noted that Fitts and Warren (1996) developed a second edition of the TSCS. However, the

manual was not retrievable and no other articles were found which examined its psychometric properties for the purpose of the present review.

Finally, although the Self Perception Profile for Adults (SPP-A; Messer & Harter, 2012) was not included in the present review as no papers examining its psychometric properties were found since 1996, the instrument nonetheless fits the criteria for a domain-specific self-esteem measure for adults. A brief description about the review that Byrne (1996) did is presented. The instrument aligns itself with developmental theory of self-concept (e.g. Harter, 2012; Marsh, 1989; Shavelson et al., 1976). In terms of the instrument's psychometric properties, internal consistency Cronbach's alpha for the subscales ranged from 0.65 to 0.92. Test-retest reliability was not reported. In terms of the validity, Byrne (1996) noted that a bigger sample size was required to test the structural validity of the measure that yielded a more adequate variable to item ratio. An EFA was conducted that yielded a clear 10-factor solution, but the Job Competence domain could not be defined. Overall, there seemed to be a need for further studies that utilised CFA strategies in testing for the validity of the factor structure, as well as evidence for its hypothesis testing validity.

Conceptual Issues

Besides the evaluation of the measures' psychometric properties, the present review also seeks to examine conceptual issues of the instruments that might influence their utility. The review of the 10 domain specific self-esteem measures highlight conceptual issues of interest to researchers and clinicians assessing domain specific self-esteem in adults.

Firstly, it was positive to note that most measures had attempted to link the instruments to a strong body of theory. This is critical because of the previously acknowledged complexity in defining self-esteem. This also allows the measure to be

tested for its construct validity. Even within the scope of domain-specific self-esteem measures, different theoretical perspectives were identified in the reviewed instruments. It is therefore critical that researchers clarify the particular theoretical framework they wish to adopt and then decide the most appropriate domain-specific self-esteem measure to use accordingly. For example, a researcher who decides to manipulate domain-specific self-esteem to examine temporary changes might opt to use the State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991), which is aimed at measuring momentary changes in self-esteem in different contexts.

In addition, depending on the theoretical stance, the measures reviewed had varying levels of specificity or abstractness of the self-esteem domains. For example, the SFSCS has six domains: likability, morality, task accomplishment, giftedness, power and vulnerability. Theoretically, these more abstract categories were hypothesised to be more relevant to a broader range of adult life experiences (Norem-Hebeisen, 1976), and therefore a general applicability across adult roles and situations. This was in contrast to the SCQ with a deeper level of specificity which included first and second order facets: academic self-concept (capability, achievement and confidence), social self-concept (same-sex peer, opposite-sex peer, and family) and self-concept presentation of self (personal confidence, physical and honest/trustworthy). Therefore, selecting one measure over another requires theoretical clarity of what researchers want to examine when selecting amongst different domain-specific self-esteem measures.

Although the measures examined were used in the general adult population, the Self-Perception Profile – Older Adults (SPP-OA; Harter & Kreinik, 2014) was the only instrument that aims to measure domain specific self-esteem in older adults. The older population comprises of individuals with varying lifestyles: some might be

living independently while others might be dependent on others in a care home. Moreover, some older people might be employed while others might be retired. Byrne (1996) reviewed the literature and found at that time that the most common approach to the tapping of self-esteem for older people had been through the use of interview techniques. These methods however lacked evidence of validity. As such, the SPP-OA seems to be a promising measure for domain specific self-esteem in older adults that aims to measure the diversity of domains inherent in different contexts.

The present review identified ten domain specific self-esteem measures for adults 18 years old and above, and academic self-esteem was measured in a number of the instruments (e.g. SDQ III, SCQ, SPP-CS). This domain would solely be applicable for university students and not working adults. Therefore, researchers would have to take the population demographic examined into account when deciding on a domain specific self-esteem measure to use.

Clinical Implications

Most of the instruments reviewed were designed for research purposes. However, some studies alluded to using the measures in clinical settings. For example, the SPP-OA (Harter & Kreinik, 2014), the SIP-AD (Butler & Gasson, 2006), and the TSCS (Fitts, 1965; Roid & Fitts, 1988) highlight the utility of individual self-esteem profiles in the clinical setting. Some of these measures purport that individuals attach meaningful importance to various domains of the self and their perceived competence in those domains (e.g. SPP-OA and SIP-AD). Through this, the SPP-OA and the SIP-AD provide individual profiles that might help identify domains of focus for a particular therapeutic intervention that is clinically meaningful for the client. For example, domains identified as problem areas for

clients can be the focus for therapy. Validation of these measures in a clinical population will be valuable in achieving these goals.

The developmental stages that some domain specific self-esteem questionnaires take into account might be relevant to specific populations in clinical settings. For example, an older adult mental health service might consider using the SPP-OA that taps into domains that might be relevant specifically in an older adult population, such as Reminiscence (enjoyment in looking back on one's life) and Nurturance (nurturance towards children or others) (Harter & Kreinik, 2014).

Finally, judging the utility of domain specific self-esteem questionnaires in a clinical setting requires consideration of their response burden. The number of items in the questionnaires included in this review range from 18 (e.g. PSCQ; Goñi and Fernández, 2007) to over 100 (e.g. SDQ III; Marsh & O'Neill, 1984). This might be a consideration for clinicians especially if they plan to administer the questionnaires pre and post treatment. Clients might find it burdensome to fill in long questionnaires. It might be useful for future studies to report information for clinical application such as the assessment time and the completion rate of questionnaires. Alternatively, clinicians might want to consider using portions of the questionnaire if they have identified particular domains for intervention. This is similar to the study by Rinn and Cunningham (2008) who only utilised the academic subscales of the SDQ III and the SPP-CS for the university population they examined.

Strengths and Limitations

One of the strengths of the current review is the use of the COSMIN rating tool which introduces rigour into the process of evaluating measurement properties and the methodological quality of studies that report on them. This ensures that studies are evaluated systematically against evidenced-based criteria.

Another strength of this review is the attempt to bridge the gap between research examining global self-esteem and domain specific self-esteem. With the increasing evidence supporting the multidimensional nature of self-esteem in literature, the present review is an attempt to build on this by appraising measures that claim to measure this.

One limitation was that only articles written in English were reviewed as there were no resources for translation available. This would have introduced selection bias. For example, the development and validation of Personal Self-Concept Questionnaire (PSCQ; Goñi & Fernández, 2007) was in the Spanish language and therefore was not included in the present review. Indeed, cross-cultural examination of domain specific self-esteem measures warrants exploration. This will be helpful given that research has found differences in how individuals in different cultures appraise self-esteem (e.g. Cai, Brown, Deng, & Oakes, 2007; Cai, Wu, Shi, Gu, & Sedikides, 2016).

Another limitation of the review is that the entire process of search and review was conducted by a single researcher. Because of this, it is possible that a small number of pertinent studies might have been left out. In addition, having only one researcher evaluate study and measurement quality might affect the reliability of ratings. Nonetheless, the present review was undertaken within a rigorous supervision framework. Study methodology and instrument quality were discussed within supervision to ensure that the ratings given were appropriate.

Conclusion

A systematic search of measures of domain specific self-esteem for adults was undertaken. Although some identified measures showed promise in terms of their psychometric properties, notable weaknesses were also found. Future research

should therefore focus on the continued validation of these measures, while bearing in mind the complexities around measuring domain specific self-esteem.

Given the evidence for domain specific self-esteem in literature, the present review provides a foundation for this and acts as a starting point to evaluate measurement quality. Finally, the further development of theory and understanding of domain specific self-esteem would have real practical implications for how we understand self-esteem in the clinical context at both the individual and the societal levels.

References

- Addeo, R. R., Greene, A. F., & Geisser, M. E. (1994). Construct validity of the Robson self-esteem questionnaire in a college sample. *Educational and Psychological Measurement, 54*(2), 439–446. Retrieved from <http://journals.sagepub.com.libproxy.ucl.ac.uk/doi/pdf/10.1177/0013164494054002018>
- Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, & S. H. Lovibond (Eds.), *Mathematical and Theoretical Systems* (pp. 7–16). North-Holland: Elsevier Science Publishers.
- Backman, C., & Wylie, R. C. (1962). *The self concept: A critical survey of pertinent research literature*. (Vol. 27). Lincoln: University of Nebraska Press. <https://doi.org/10.2307/2089838>
- Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(1), 35–67. <https://doi.org/10.1080/10705519409539961>
- Bannister, D., & Fransella, F. (1986). *Inquiring man: The psychology of personal constructs*. London: Croom Helm.
- Beck, A. ., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–71. <https://doi.org/10.1001/archpsyc.1961.01710120031004>

- Bishop, S. L., Walling, D. P., & Walker, B. A. (1997). The emperor's clothes: Assessing the validity of scores on the Tennessee Self-Concept Scale. *Educational and Psychological Measurement, 57*(1), 150–163.
<https://doi.org/10.1177/0013164497057001011>
- Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 115–160). San Diego, CA, US: Academic Press.
https://doi.org/10.1300/J103v11n02_05
- Borgatta, E. F., Kercher, K., & Stull, D. E. (1986). A cautionary note on the use of principal components analysis. *Sociological Methods & Research, 15*(1–2), 160–168. <https://doi.org/10.1177/0049124186015001011>
- Bracken, B. A. (1996). *Handbook of self-concept: Development, social and clinical considerations*. New York: John Wiley.
- Brown, G. W., Bifulco, A., & Andrews, B. (1990). Self-esteem and depression. *Social Psychiatry and Epidemiology, 25*, 244–249.
- Brown, G. W., Dutton, K. A., & Cook, K. E. (2001). From the top down: Self-esteem and self-evaluation. *Cognition and Emotion, 15*(5), 615–631.
- Burns, R. B. (1979). *The self-concept: Theory, measurement, development and behaviour*. London: Longman.
- Butler, R., & Green, D. (1998). *The child within: The exploration of personal construct theory with young people*. Oxford: Butterworth Heinemann.

- Butler, R. J., & Gasson, S. L. (2005). Self esteem/self concept scales for children and adolescents: A review. *Child and Adolescent Mental Health, 10*(4), 190–201.
<https://doi.org/10.1111/j.1475-3588.2005.00368.x>
- Butler, R. J., & Gasson, S. L. (2006). Development of the Self-Image Profile for Adults [SIP-AD]. *European Journal of Psychological Assessment, 22*(1), 52–58. <https://doi.org/10.1027/1015-5759.22.1.52>
- Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research, 54*, 427–456.
- Byrne, B. M. (1996). *Measuring self-concept across the lifespan*. Washington, DC: American Psychological Association.
- Cai, H., Brown, J. D., Deng, C., & Oakes, M. A. (2007). Self-esteem and culture: Differences in cognitive self-evaluations or affective self-regard? *Asian Journal of Social Psychology, 10*(3), 162–170. <https://doi.org/10.1111/j.1467-839X.2007.00222.x>
- Cai, H., Wu, L., Shi, Y., Gu, R., & Sedikides, C. (2016). Self-enhancement among Westerners and Easterners: A cultural neuroscience approach. *Social Cognitive and Affective Neuroscience, 11*(10), 1569–1578.
<https://doi.org/10.1093/scan/nsw072>
- Coopersmith, S. (1967). *The antecedents of self esteem*. San Francisco: W. H. Freeman.

- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96(4), 608–630.
<https://doi.org/10.1037/0033-295X.96.4.608>
- Crocker, P. R. E., & Ellsworth, J. P. (1990). Perceptions of competence in physical education students. *Canadian Journal of Sport Sciences*, 15(4), 262–266.
- Davis-Kean, P. E., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for young children: A framework for future measures. *Child Development*, 72(3), 887–906. <https://doi.org/10.1111/1467-8624.00322>
- De Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine: A practical guide*. Cambridge: Cambridge University Press.
<https://doi.org/http://dx.doi.org.myaccess.library.utoronto.ca/10.1017/CBO9780511996214>
- Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Measures of self-esteem. In *Measures of personality and social psychological constructs* (pp. 131–157). <https://doi.org/10.1016/B978-0-12-386915-9.00006-1>
- Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-Esteem: Enduring issues and controversies. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences*. Wiley-Blackwell.
- Eagly, A. H. (1967). Involvement as a determinant of response to favorable and unfavorable information. *Journal of Personality and Social Psychology*, 7(3), 1–15. <https://doi.org/10.1037/h0025080>

- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*, 5–18. <https://doi.org/10.1007/s11136-007-9198-0>
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341–349. <https://doi.org/10.1037/1040-3590.8.4.341>
- Erikson, E. H. (1959). Identity and the life cycle: Selected papers. *Psychological Issues, 1*, 1–171. <https://doi.org/10.1001/archinte.1966.00290150102025>
- Fitts, W. H. (1965). *Tennessee Self Concept Scale*. Nashville, TN: Counselor Recordings and Tests.
- Fitts, W. H., & Warren, W. L. (1996). *Tennessee Self Concept Scale Manual: Second edition*. Los Angeles, CA: Western Psychological Services.
- Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology, 46*(2), 404–421. <https://doi.org/10.1037/0022-3514.46.2.404>
- Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, self-esteem, and anger in emerging adulthood: Seven-year trajectories. *Developmental Psychology, 42*(2), 350–365. <https://doi.org/10.1037/0012-1649.42.2.350>
- Gentile, B., Dalon-Pascoe, B., Twenge, J., Maitiono, A., Grabe, S., & Wells, B. E. (2009). Gender differences in domain-specific self-esteem. A meta-analysis. *Review of General Psychology, 13*(1), 34–45.

- Gergen, K. J. (1971). *The concept of self*. New York: Holt, Rinehart & Winston.
- Goñi, E. (2009). *El autoconcepto personal: Estructura interna, medida y variabilidad*. Leioa: UPV/EHU.
- Goñi, E., & Fernández, A. (2007). Los dominios social y personal del autoconcepto. *Revista de Psicodidáctica*, *12*, 179–194.
- Goñi, E., Madariaga, J. M., Axpe, I., & Goñi, A. (2011). Structure of the Personal Self-Concept (PSC) questionnaire. *Int J Clin Health Psychology*, *11*(3), 509–522.
- Goodman, N., & Coopersmith, S. (1969). The antecedents of self-esteem. *American Sociological Review*, *34*(1), 116. <https://doi.org/10.2307/2092806>
- Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*(1), 33–39. https://doi.org/10.1207/s15327906mbr2501_3
- Gual, P., Perez-Gaspar, M., Martinez-Gonzallaz, M. A., Lahortiga, J., & Cervera-Enguix, S. (2002). Self-esteem, personality and eating disorders: baseline assessment of a prospective population based cohort. *International Journal of Eating Disorders*, *31*, 261–273.
- Hall, P. L., & Tarrier, N. (2003). The cognitive-behavioural treatment of low self-esteem in psychotic patients: A pilot study. *Behaviour Research and Therapy*, *41*(3), 317–332. [https://doi.org/10.1016/S0005-7967\(02\)00013-X](https://doi.org/10.1016/S0005-7967(02)00013-X)

- Hambleton, R. K., Robin, F., & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 553–581). San Diego, CA, US: Academic Press.
<https://doi.org/10.1016/B978-012691360-6/50020-3>
- Harter, S. (1985). *The self-perception profile for children: Revision of the perceived competence scale for children*. Unpublished manuscript, University of Denver, Colorado, USA.
- Harter, S. (1988). *Self-perception profile for adolescents*. Denver, CO: University of Denver Press.
- Harter, S. (1990). Issues in the assessment of the self-concept of children and adolescents. In A. M. La Greca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 292–325). Needham Heights, MA, US: Allyn & Bacon.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford Press.
- Harter, S. (2012). *The construction of the self: Developmental and sociocultural perspectives*. New York: Guilford Press.
- Harter, S., & Kreinik, P. (2014). The Self-Perception Profile for Older Adults, 1–75.
- Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: Lawrence Erlbaum.

- Heatherton, T. E. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*(6), 895–910. <https://doi.org/10.1037/0022-3514.60.6.895>
- Heatherton, T. F., & Wyland, C. L. (2003). Assessing self-esteem. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures*. (pp. 219–233). Washington, DC, US: American Psychological Association. <https://doi.org/10.1037/10612-014>
- Helmreich, R., Stapp, J., & Ervin, C. (1974). The Texas Social Behavioral Inventory (TSBI): An objective measure of self-esteem or social competence. *JSAS Catalog of Selected Documents in Psychology*, *4*(79).
- Hubbard, R., & Allen, S. J. (1987). A cautionary note on the use of principal components analysis. *Sociological Methods & Research*, *16*(2), 301–308. <https://doi.org/10.1177/0049124187016002005>
- Hughes, H. M. (1984). Measures of self-concept and self-esteem for children ages 3-12 years: A review and recommendations. *Clinical Psychology Review*, *4*(6), 657–692. [https://doi.org/10.1016/0272-7358\(84\)90011-4](https://doi.org/10.1016/0272-7358(84)90011-4)
- James, W. (1890). The principles of psychology. *New York Holt*, *1*, 697. <https://doi.org/10.1037/10538-000>
- Janis, I. ., & P.B, F. (1969). A behavioral assessment of persuasability consistency of individual differences. In C. I. Hovland & I. L. Janis (Eds.), *Personality and Persuasibility* (pp. 55–68). New Haven, CT: Yale Univ. Press.

- Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton.
- Kline, P. (1993). *The handbook of psychological testing (2nd ed.)*. Florence, KY, US: Taylor & Frances/Routledge. [https://doi.org/10.1016/S0191-8869\(96\)90047-1](https://doi.org/10.1016/S0191-8869(96)90047-1)
- Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, *51*(4), 858–866. <https://doi.org/10.1037/0022-3514.51.4.858>
- Marsh, H. W. (1989). *Self-Description Questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept: An interim test manual and a research monograph*. New South Wales, Australia: University of Western Australia.
- Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology*, *51*(6), 1224–1236.
- Marsh, H. W. (1992a). *Self-Description Questionnaire I: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept—A test manual and research monograph*. Macarthur, NSW, Australia: University of Western Sydney.
- Marsh, H. . (1992b). *Self-Description Questionnaire II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept—A test manual and research monograph*. Macarthur, NSW, Australia: University of Western Sydney.

Marsh, H. W. (1993). Academic Self-Concept: Theory, measurement and research.

In J. M. Suls (Ed.), *Psychological Perspectives on the self, Vol. 4. The self in social perspective* (pp. 59–98). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-

concept. In *Handbook of self-concept: Developmental, social and clinical considerations* (pp. 38–90). New York: John Wiley.

Marsh, H. W., & Peart, N. D. (1988). Competitive and cooperative physical fitness

training programs for girls: Effects on physical fitness and multidimensional self-concepts. *Journal of Sport and Exercise Psychology, 10*, 390–407.

Retrieved from <http://psycnet.apa.org/psycinfo/1989-24006-001>

Marsh, H. W., & Richards, G. E. (1988). The Tennessee Self-Concept Scales:

Reliability, Internal Structure, and Construct Validity. *Journal of Personality and Social Psychology, 55*, 612–624.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and

performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2),

133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-

concept, self-esteem, achievement, and attainment over seven adolescent years:

Unidimensional and multidimensional perspectives of self-concept. *Personality*

and Social Psychology Bulletin, 34(4), 542–552.

<https://doi.org/10.1177/0146167207312313>

Marsh, H. W., & O’Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, 21(2), 153–174.
<https://doi.org/10.1111/j.1745-3984.1984.tb00227.x>

Marsh, H. W., Richards, G. E., & Barnes, J. (1986). Multidimensional self-concepts: A long-term follow-up of the effect of participation in an outward bound program. *Personality and Social Psychology Bulletin*, 12(4), 475–492.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20(3), 107–123.
https://doi.org/10.1207/s15326985ep2003_1

Maslow, A. H. (1954). *Motivation and Personality*. Oxford, England: Harpers.

McCain, J. L., Jonason, P. K., Foster, J. D., & Campbell, W. K. (2015). The bifactor structure and the “dark nomological network” of the State Self-Esteem Scale. *Personality and Individual Differences*, 72, 1–6.
<https://doi.org/10.1016/j.paid.2014.08.006>

Messer, B., & Harter, S. (2012). The Self-Perception Profile for Adults: Manual and questionnaires. *University of Denver*.

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., ... de Vet, H. C. (2010). The COSMIN checklist for evaluating the

methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, 10(1), 22.

<https://doi.org/10.1186/1471-2288-10-22>

Monge, R. H. (1973). Developmental trends in factors of adolescent self-concept.

Developmental Psychology, 8, 382–393.

Mruk, C. J. (2006). *Self-esteem research, theory and practice: Toward a positive psychology of self-esteem*. New York: Springer Publishing Company.

Neemann, J., & Harter, S. (2012). The Self-Perception Profile for College Students.

University of Denver.

Norem-Hebeisen, A. . (1976). A multidimensional construct of self-esteem. *Journal of Educational Psychology*, 68, 559–565.

O'Brien, E. J., Bartoletti, M., & Leitzel, J. (2006). Self-esteem, psychopathology and psychotherapy. In *Self-esteem: Issues and answers. A sourcebook of current perspectives* (pp. 306–316). East Sussex: Psychology Press.

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: A cohort-sequential longitudinal study.

Journal of Personality and Social Psychology, 98(4), 645–658.

<https://doi.org/10.1037/a0018769>

Paradise, A. W., & Kernis, M. H. (2002). Self-esteem and psychological well-being:

Implications of fragile self-esteem. *Journal of Social and Clinical Psychology*,

21(4), 345–361. <https://doi.org/10.1521/jscp.21.4.345.22598>

- Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life Research, 22*(10), 2639–2659. <https://doi.org/10.1007/s11136-013-0395-8>
- Pliner, P., Chaiken, S., & Flett, G. L. (1990). Gender differences in concern with body weight and physical appearance over the life span. *Personality and Social Psychology Bulletin, 16*(2), 263–273. <https://doi.org/10.1177/0146167290162007>
- Rinn, A. N., & Cunningham, L. G. (2008). Using self-concept instruments with high-ability college students: Reliability and validity evidence. *Gifted Child Quarterly, 52*, 232–242. <https://doi.org/10.1177/0016986208319458>
- Robson, P. (1989). Development of a new self-report questionnaire to measure self-esteem. *Psychological Medicine, 19*(2), 513–518.
- Roid, G. H., & Fitts, W. H. (1988). *Tennessee Self-Concept Scale, revised manual*. Los Angeles, CA: Western Psychological Services.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.
- Rosenberg, M. (1986). Self-concept from middle childhood through adolescence. In S. & A. G. Greenwald (Ed.), *Psychological perspectives on the self* (Hillsdale, pp. 107–135). Erlbaum.

- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Scheff, T. J., & Fearon, D. S. (2004). Cognition and emotion? The dead end in self-esteem research. *Journal for the Theory of Social Behaviour*, 34, 73–90.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
<https://doi.org/10.1002/0470010940>
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441.
<https://doi.org/10.3102/00346543046003407>
- Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106(1), 148–154.
<https://doi.org/10.1037/0033-2909.106.1.148>
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press. <https://doi.org/10.1002/9780470479216.corpsy0943>
- Stake, J. E. (1979). The ability/performance dimension of self-esteem: Implications for women's achievement behavior. *Psychology of Women Quarterly*, 3, 365–377.

- Stake, J. E. (1985). Predicting reactions to everyday events from measures of self-esteem. *Journal of Personality*, *53*, 530–542.
- Stake, J. E. (1994). Development and validation of the Six-Factor Self-Concept Scale for Adults. *Educational and Psychological Measurement*, *54*(1), 56–72.
Retrieved from
<http://journals.sagepub.com.libproxy.ucl.ac.uk/doi/pdf/10.1177/0013164494054001006>
- Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology*, *54*, 1038–1051.
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., ... de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & De Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657.
<https://doi.org/10.1007/s11136-011-9960-1>
- Tzeng, O. C. ., Maxey, W. A., Fortier, R., & Landis, D. (1985). Construct evaluation of the Tennessee Self Concept Scale. *Educational and Psychological Measurement*, *45*, 63–77.

- Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology*, 83, 185–197.
- Waugh, R. F. (2001). Measuring ideal and real self-concept on the same scale, based on a multifaceted, hierarchical model of self-concept. *Educational and Psychological Measurement*, 61(1), 85–101. Retrieved from <http://journals.sagepub.com.libproxy.ucl.ac.uk/doi/pdf/10.1177/00131640121971086>
- Wells, A. J. (1988). Variations in mothers' self-esteem in daily life. *Journal of Personality and Social Psychology*, 55(4), 661–668.
<https://doi.org/10.1037//0022-3514.55.4.661>
- Wells, L. E., & Marwell, G. (1976). *Self-esteem: Its conceptualization and measurement*. Beverly Hills: Sage.
- Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & et al. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27(4), 552–565. <https://doi.org/10.1037/0012-1649.27.4.552>
- Wylie, R. C. (1979). *The self-concept: Theory and research on selected topics (Vol. 2)*. Lincoln: University of Nebraska Press.
- Wylie, R. C. (1974). *The self-concept: A review of methodological considerations and measuring instruments*. Lincoln: University of Nebraska Press.

Yanico, B. J., & Lu, T. G. C. (2000). A psychometric evaluation of the Six-Factor Self-Concept Scale in a sample of racial/ethnic minority women. *Educational and Psychological Measurement, 60*(1), 86–99.

<https://doi.org/10.1177/00131640021970385>

Zanon, C., Hutz, C. S., Yoo, H. (Henry), & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*(1), 18. <https://doi.org/10.1186/s41155-016-0040-x>

Part 2: Empirical Paper

**The effectiveness of a domain specific self-esteem group intervention:
a pilot study**

Abstract

Aim. The present research evaluated the effectiveness of a domain specific self-esteem group intervention, based on the unsatisfactory self-esteem model developed by Hollingdale (2015). The main aims of the study were to assess changes in domain specific self-esteem (i.e. perceived competence) in valued domains, discrepancy between perceived competence and importance placed in valued domains, and attributional styles towards negative and positive events. In addition, the relationship between domain specific self-esteem and attributional styles was examined.

Method. The present study utilised an uncontrolled design and students from University College London (UCL) were recruited for the study. Domain specific self-esteem in valued domains, discrepancy between perceived competence and importance placed in valued domains, and attributional style were assessed at pre-intervention, post-intervention and one-month follow up. A correlational analysis was also conducted between domain specific self-esteem and attributional styles.

Results. The results indicated that participants showed improvements in domain specific self-esteem (i.e. perceived competence) in their valued domains. The discrepancy between perceived competence and importance placed in valued domains decreased after the intervention. These findings were maintained at one month follow up. Attributional styles towards negative events showed a shift towards more external, unstable and specific styles post-intervention and continued moving in this direction at one-month follow up. No changes were observed in attributional styles towards positive events. A significant relationship was also found between domain specific self-esteem and attributional styles towards negative events but not for positive events.

Conclusion. The domain specific self-esteem group is a promising intervention for self-esteem that requires further study.

Introduction

Self-esteem has been extensively researched in the fields of social sciences and psychology for many years (Fennell, 1997). Historically, it has been defined as a person's subjective evaluation of their self-worth (Donnellan, Trzesniewski & Robins, 2011). Many studies have explored the relationships between self-esteem and other outcomes. High self-esteem is linked to coping with life stresses, achieving more in life and maintaining positive relationships with others (Coopersmith, 1967; Harter, 1990; Paradise & Kernis, 2002). On the other hand, low self-esteem has been identified as an aetiological factor in psychiatric diagnoses including depression (Brown, Bifulco, & Andrews, 1990), anxiety (O'Brien, Bartoletti & Leitzel, 2006; Watson, Suls & Haig, 2002), psychosis (Hall & Tarrier, 2003), obsessive compulsive disorder (Ehnholt, Salkovskis, & Rimes, 1999) and eating disorders (Gual et al., 2002). Moreover, low self-esteem has been found to be related to substance abuse (Akerlind, Hornquist, & Bjurulf, 1988; Brown, Andrews, Harris, Adler, & Bridge, 1986; Button, Sonuga-Barke, Davies, & Thompson, 1996) and chronic pain (Soares & Grossi, 2000).

Difficulties with self-esteem and the impact it has on individuals' mental health and wellbeing are commonly seen in clinical practice (Fennell, 1997). It is therefore critical to develop and evaluate effective treatments for improving self-esteem.

Cognitive Therapy for Low Self-Esteem

Fennell (1997) developed a cognitive model of low self-esteem (Figure 1). Low self-esteem is defined as the negative image of the self, which tends to be global, persistent and enduring (Fennell, 1997). The model is built upon Beck's (1976) cognitive model of emotional disorders that was originally targeted at

depression and anxiety. Fennell's (1997) model for low self-esteem suggests that individuals form global negative judgements, known as core beliefs or the 'bottom line', about themselves, others and the world, which are shaped by early life experiences. Individuals develop dysfunctional assumptions or 'rules for living' to compensate for these negative beliefs and are able to cope providing that they adhere to these assumptions (Fennell, 1997). However, situational events might activate these negative beliefs, triggering automatic negative thoughts which elicit feelings and behaviours that maintain the negative core beliefs (Fennell, 1997).

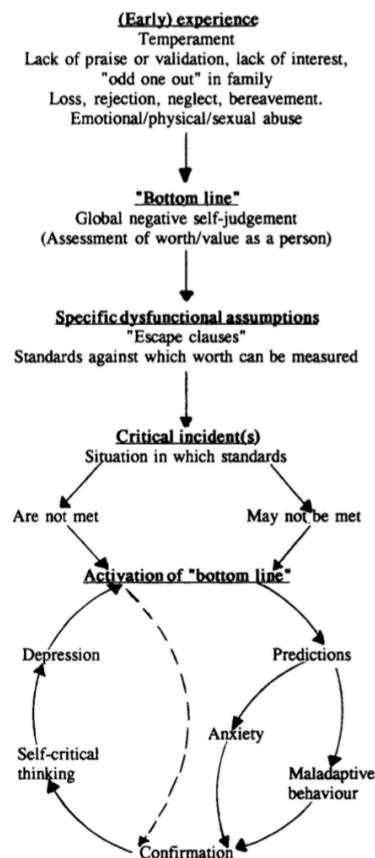


Figure 1. Fennell's (1997) cognitive model of low self-esteem

Stemming from this model, Fennell (1997) developed a Cognitive Behavioural Therapy (CBT) protocol for low self-esteem. This combined standard CBT practices (e.g., Beck, Rush, Shaw, & Emery, 1979) and schema approaches

(Young, Klosko, & Weishaar, 2003). The protocol applied these approaches to low self-esteem and focused on challenging individuals' negative core beliefs to help them develop a more balanced view about themselves (Fennell, 1997). This was aimed at individuals who are prone to biases of identifying perceived failures and ignoring any possible contradictory evidence (Fennell, 1997).

Strengths and Weaknesses of Fennell's (1997) Model

Fennell's model of self-esteem was the first of its kind and is currently extensively used in clinical practice (Waite, McManus & Shafran, 2012). It provides a useful heuristic for clients to make sense of their difficulties, including how they are developed and maintained, and provides a framework for treatment. However, there are few studies which have systematically evaluated the use of Fennell's (1997) model of low self-esteem. To date, the protocol had been evaluated in only one Randomised Control Trial (RCT; Waite et al., 2012). Results indicated a significant improvement in self-esteem scores in the treatment group compared to the waitlist group at the end of treatment and follow-up (Waite et al., 2012). A limitation of this study, noted by its authors, was the small sample size of 11 participants, comprising of mainly highly educated women in each group. Although the initial findings were promising, there has been a lack of replication studies.

Besides the RCT conducted by Waite et al. (2012), there have been single case examples (e.g. Butler, Fennell & Hackmann, 2008; Chatterton, Hall, & Tarrier, 2007; Fennell, 1997; McManus, Waite, & Shafran, 2009) that have utilised Fennell's (1997) model. A few uncontrolled evaluations of adaptations of the model and treatment protocol for group settings (e.g. Rigby & Waite, 2007; Morton, Roach, Reid, Stewart, 2012) have also been conducted. Rigby and Waite (2007) combined Fennell's (1997) model with narrative techniques in a group setting and found

significant improvements in self-esteem. Morton et al. (2012) also conducted a CBT group intervention for women with low self-esteem and found similar results.

However, the way self-esteem is defined in Fennell's (1997) model presents two key limitations. Firstly, the model assumes that low self-esteem is global. Most self-esteem research has traditionally considered the construct as a global concept, that is, an individual's global evaluation of themselves that is stable across time and situations (Rosenberg, 1965). However, since the 1980s, there is now a wealth of evidence also supporting the multidimensional or domain-specific nature of self-esteem (Byrne, 1996). Byrne (1984, p. 427) conducted an extensive review of construct validation research of self-esteem and concluded that self-esteem is indeed "a multidimensional construct, having one general construct and several specific facets". Marsh and Shavelson (1985) also argued that self-esteem cannot be adequately understood if its multidimensionality is ignored. Marsh (1986) found that domain-specific and global self-esteem shared associations of .06 to .60, suggesting that these constructs were related but not interchangeable.

Harter (2012) conceptualised self-esteem as domain-specific, which refers to an individual's self-appraisals within more circumscribed domains, for example, intellectual, athleticism and appearance. Individuals therefore may hold different levels of self-esteem in various domains (Mruk, 2006). In self-esteem literature, several authors have found support for domain specific self-esteem (e.g., Harter, 1985; Marsh 1986; Marsh & Shavelson, 1985; Swann 1987). Marsh and Craven (2006) also appraised a large body of research which indicated that academic outcomes were related to academic self-esteem but unrelated to global self-esteem; this suggested a differentiation between domain specific and global self-esteem. While there has been agreement in literature on the multidimensional nature of self-

esteem, there is currently a lack of self-esteem interventions targeting domain specific self-esteem.

Secondly, while Fennell (1997, p.2) purported that low self-esteem is “enduring over time and across situations”, research has indicated the variability and fluctuations of self-esteem across life situations and contexts (e.g. Galambos, Barker, & Krahn, 2006; Harter & Whitesell, 2003; Orth, Trzesniewski & Robins, 2010; Wigfield, Eccles, Mac Iver, Reuman & Midgley, 1991). Developmental changes and transitions across the lifespan have been found to lead to changes in self-esteem that might be accounted for by changes in role demands, maturational changes, physical functioning and the individual’s socioeconomic status (Orth et al., 2010). For example, a cohort-sequential longitudinal study exploring self-esteem changes found that self-esteem increased during young and middle adulthood, reached a peak at about age 60 years, and then declined in old age (Orth et al., 2010). Moreover, evidence points to self-concept dimensions becoming more differentiated from mid-adolescence (Marsh & Shavelson, 1985). Neemann and Harter (2012) identified that as adolescents age, they accept more responsibility for their own lives and educational goals which lead to differentiated self-esteem in life domains. Therefore, treatments for self-esteem should consider meaningful and developmentally appropriate domains across the lifespan.

The Unsatisfactory Self Esteem Model: Conceptual Issues (Hollingdale, 2015)

To address the limitations of Fennell’s model, an alternative, unpublished CBT model of self-esteem and a related domain specific self-esteem group intervention session plan (Appendix F) were proposed by Hollingdale (2015), a trainee clinical psychologist at University College London (UCL). Based on a multidimensional framework of self-esteem, this model integrates theories including

Fennell's (1997) low self-esteem model and Abramson, Seligman and Teasdale's (1978) attributional styles. The new model serves two purposes: first, it aims to provide a more helpful and meaningful model of self-esteem to support clients' and clinicians' understanding of the concept; secondly, it aims more accurately to identify areas to target with self-esteem interventions.

In Hollingdale's (2015) model, self-esteem is conceptualised as multidimensional, or what Harter (2012) terms as domain specific self-esteem. Self-esteem varies amongst domains (e.g. intellectual, athleticism and appearance) and is a deeply personal and complex experience that cannot be accurately identified or meaningfully interpreted with an arbitrary threshold of "low" or "high". Instead, the model considers domain specific self-esteem to be on a spectrum that at times can become "unsatisfactory" for an individual's needs. It can become "unsatisfactory" for the individual, dependent on their preferred level of functioning, within a specific domain, situation or period in their life. This has clinical implications in understanding a client's difficulties and acknowledging that they might not be experiencing global "low" self-esteem but "unsatisfactory" self-esteem in specific domains in their current life situation.

Furthermore, the importance or value placed in a specific domain is a key concept in Hollingdale's (2015) model. This is based on work by James (1982) who theorised that perceptions of competence in domains deemed important were the best predictors of self-esteem. Similarly, the unsatisfactory self-esteem model posits that the importance placed on a specific domain will influence an individual's self-esteem in that domain. For example, an individual may place no value on being a good athlete and therefore potential threats towards this domain will not violate an individual's self-esteem in this domain. Indeed, the individual may not even

perceive threats in this domain. However, the individual at the same time may place significant value on academic achievements and so perceived threats to their competence in this domain will result in unsatisfactory self-esteem in that domain. This is consistent with Neemann and Harter's (2012) conceptualisation of domain specific self-esteem, where perceived competence and importance placed in various domains are assessed through a questionnaire. A discrepancy score, indicating the difference between one's perceived competence and one's importance ratings can also be calculated for each domain (Neemann & Harter, 2012).

Finally, the proposed model also suggests that domain specific self-esteem will fluctuate over the course of an individual's life. This is consistent with research indicating that self-esteem fluctuates over the lifespan and across contexts (e.g. Galambos et al., 2006; Harter & Whitesell, 2003; Orth et al., 2010; Wigfield et al., 1991). An individual may react very differently to perceived violations of domains across their lifespan. Clinically, it therefore becomes critical to ascertain changes in an individual's self-esteem across time within developmentally appropriate domains in which they are currently experiencing difficulties.

The Unsatisfactory Self Esteem Model (Hollingdale, 2015)

The unsatisfactory self-esteem model (Figure 2) posits that an individual's early life experience contributes to their core beliefs about themselves in different life domains. Familial, social and cultural experiences construct an individual's values and specifically how much value or importance one places in each life domain, such as family, relationships, academic achievement, career, appearance, etc.

An individual's life experiences result in the development of attributional styles that influence how the individual perceives and interprets events in various

domains. Attributional styles refer to a general tendency to make internal (versus external), stable (versus temporary), and global (versus specific) attributions for positive and negative events (Abramson et al., 1978). Past research has indicated that individuals who tend to experience deficits in their self-esteem attribute negative events to more internal, stable and global causes (Abramson et al., 1978; Feather, 1983; Ickes & Layden, 1978; Seligman, Abramson, Semmel, & Baeyer, 1979). However, the association between self-esteem and attributional styles for positive events seems to be less clear. While Feather (1983) and Tennen and Herzberger (1987) found that individuals with high self-esteem tend to attribute successes to internal, stable, and global causes, Tennen, Herzberger and Nelson (1987) found that self-esteem was not associated with attributional styles for positive events.

Different combinations of attributional styles result in the development of core beliefs and assumptions about the self, others and the world. When situations encountered are perceived to violate an individual's self-esteem in a valued domain, negative core beliefs and assumptions are activated, which subsequently trigger feelings and behaviours that perpetuate their core beliefs.

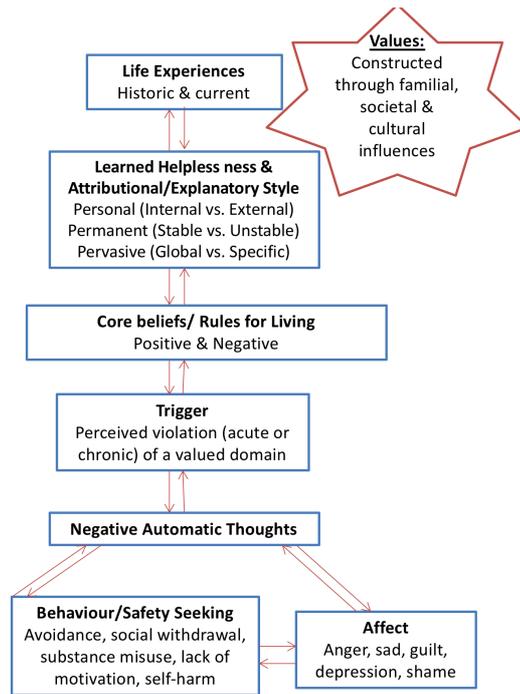


Figure 2. The unsatisfactory self-esteem model (Hollingdale, 2015)

The Unsatisfactory Self-Esteem Model: Group Intervention (Hollingdale, 2015)

Hollingdale (2015) conceptualised a four-session CBT group session protocol based on the unsatisfactory self-esteem model. Group sessions begin with psycho-education about the model and an exploration of each individual’s domain specific self-esteem profile. Through the self-esteem profile, individuals are able to identify their domain specific self-esteem (i.e. perceived competence) and also the importance of these domains. An example is represented in figure 3. The solid line represents the individual’s importance placed in the 12 domains while the dotted line represents the individual’s perceived competence in the 12 domains. A higher score indicates greater perceived competence or importance placed in that domain. Taking the example of the domain of close friendships in figure 3, the individual’s importance placed in the domain is rated as four, while the perceived competence is rated as three.

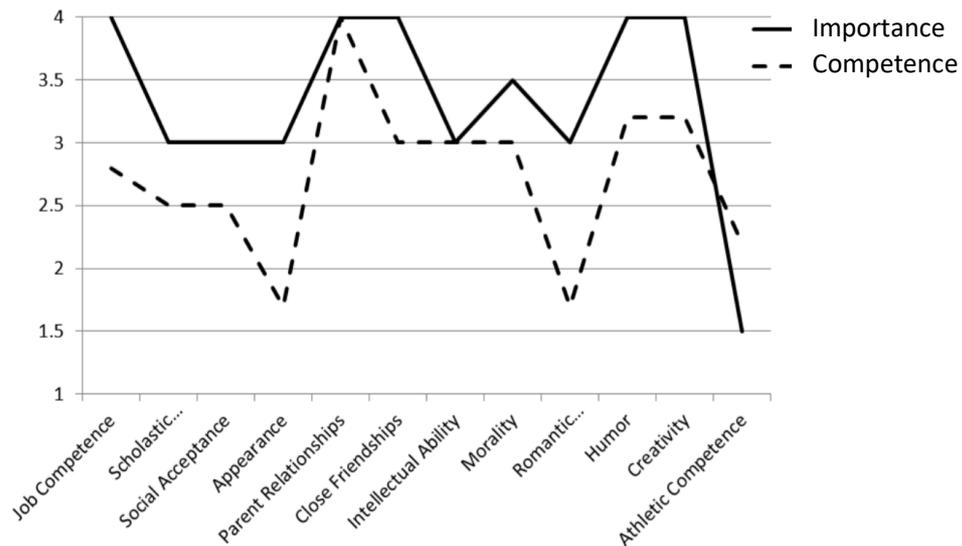


Figure 3. Domain specific self-esteem profile example

A collaborative formulation for each individual is used to identify and explore possible antecedents, triggers and maintaining factors with regard to unsatisfactory self-esteem in valued domains. Positive data logs, identification of automatic negative thoughts and behavioural experiments are utilised in and out of sessions with the aim of increasing domain specific self-esteem (i.e. perceived competence) in valued domains. This thereby leads to a reduction in the discrepancy between perceived competence and importance placed in valued domains following the intervention.

Moreover, it was anticipated that the intervention should also result in attributional styles changes when encountering negative and positive events. Through the use of thought diaries and behavioural experiments, participants might perceive that negative outcomes might not be contingent on acts in their repertoires and instead be due to the external situation which might be less internal, global and stable. This would result in a shift from more internal, stable and global attributions to a more external, unstable and specific attributional style for negative events. With regard to attributions of positive events, the intervention aims to help individuals get

a balanced view of successes resulting in a shift from more external, unstable, and specific attributions to more internal, stable and global attributions for positive events. Finally, the group intervention concludes with a relapse prevention plan.

The present research also sought to explore the relationship between domain specific self-esteem and attributional styles, to build on findings for the existing relationship between self-esteem and attributional styles (e.g. Abramson et al., 1978; Feather, 1983; Ickes & Layden, 1978; Tennen, et al., 1987). The current literature indicates more evidence for the relationship between self-esteem and attributional styles for negative events as compared to positive events. As such, the present study will explore these relationships with regard to domain specific self-esteem.

This project is a joint one with Emily Dixon, who is also a trainee clinical psychologist. Different outcomes and data were examined. This study examined the effectiveness of the domain-specific self-esteem group by assessing changes in domain-specific self-esteem in participants' valued domains, discrepancy scores in valued domains and attributional styles. In addition, this study explored the relationship between attributional styles and domain specific self-esteem. Dixon (2018) explored whether the group intervention would lead to improvements on scores of anxiety, depression, psychological wellbeing, and global self-esteem. In addition, the relationship between domain-specific and global self-esteem was examined. Finally, participants' feedback on their experience of the group was collected and analysed. This was gathered through a questionnaire that included both quantitative and qualitative items. The details of joint working are presented in Appendix A.

Hypotheses

The present study aimed to explore the effectiveness of a brief group intervention using Hollingdale's (2015) model of unsatisfactory self-esteem. The following hypotheses were proposed:

1. Domain specific self-esteem (i.e. perceived competence) in valued domains will increase post-intervention and this change will be maintained at follow-up.
2. The discrepancy between domain specific self-esteem (i.e. perceived competence) and importance placed in valued domains will decrease post-intervention and this change will be maintained at follow-up.
3. Attributional styles for negative events will shift from internal, stable and global styles towards more external, unstable and specific styles post intervention and this change will be maintained at follow-up.
4. Attributional styles for positive events will shift from external, unstable and specific styles towards more internal, stable and global styles post intervention and this change will be maintained at follow-up.
5. Individuals who adopt more internal, stable and global attributional styles towards negative events will be more likely to experience deficits in their domain specific self-esteem in valued domains.
6. Similarly, individuals who adopt more external, unstable and specific attributional styles towards positive events will be more likely to experience deficits in their domain specific self-esteem in valued domains.

Method

Design

The intervention is untested and therefore the study utilised an uncontrolled design to test the effectiveness of a potentially promising intervention for self-esteem. We hoped to find a possible effect of the group on domain-specific self-esteem as an initial test prior to pitting it against a control group or other interventions in future studies.

Responsibility for recruitment, administering measures and conducting group sessions was shared between the two researchers. The entire research process and facilitation of the groups were conducted under a rigorous supervision framework. This ensured treatment fidelity and that issues arisen during the research process were reflected upon and discussed. Participants were required to attend a four-session group programme, held weekly, and a follow-up session one month later at the University College London (UCL) campus. The repeated measures variable 'Time' had three levels (pre-intervention, post-intervention, and one-month follow up) and the dependent variables were measures of domain specific self-esteem (i.e. perceived competence) in valued domains, discrepancy between perceived competence and importance placed in valued domains, and attributional styles.

Participants

Participants were UCL students recruited for the purpose of this study between December 2016 and January 2018. The inclusion criteria for the study were as follows: The participant believed that they had difficulties with their self-esteem, was a student at UCL, a fluent English speaker, over 18 years old, had normal visual acuity and was computer literate with internet access. The only exclusion criterion was if potential participants experienced daily thoughts of suicide and self-harm.

The study was advertised through the following means: UCL Newsletter emails, posters displayed across the UCL campus, the waiting room at Student Psychological Services (SPS) at UCL and word of mouth.

Procedure

Potentially interested participants were directed to an online screening questionnaire on the UCL Qualtrics Survey Platform. They were provided with a Participant Information Sheet (PIS; Appendix D) describing the study and asked to give their consent to participate (Appendix E). The PIS included the researchers' contact details if participants had questions about the study. The online screening questionnaire comprised of the following measures: Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001), General Anxiety Disorder Questionnaire-7 (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006) and the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). After the completion of these measures, participants deemed not appropriate for the study were excluded automatically, based on the exclusion criterion. This was assessed through item 9 ("Thoughts that you would be better off dead, or of hurting yourself in some way") score on the PHQ-9; participants who scored 3 (*nearly every day*) on that item were excluded from the study. Anyone who scored one or more on item 9 on the PHQ-9 was provided with an online information sheet setting out how they could seek help through crisis hotline details and numbers for services that they could contact if required.

Demographic information was subsequently collected through the Qualtrics platform for participants who were eligible for the study. Participants also filled in the possible dates they were available to attend the group. Participants were then

contacted via email, confirming their participation in the groups with the dates and location provided.

Group intervention

Broadly, groups began with defining domain specific self-esteem and having participants chart their own domain specific self-esteem profile based on the completed questionnaires during the initial session. Based on their valued domains, participants employed various CBT techniques, including thought diaries and behavioural experiments, with the goal of increasing domain specific self-esteem in their valued domains. The specific details of the intervention at each session are presented in Appendix F, and the presentation slides for the group sessions are presented in Appendix G.

The general overview for each session was as follows:

Session 1: Collect pre-group self-report measures; explore the definition of domain specific self-esteem; introduce the generic CBT model.

Session 2: Participants given their domain specific self-esteem profile and attributional styles profile. Introduce the unsatisfactory self-esteem model and vicious cycles of unsatisfactory self-esteem; identify self-critical thoughts and/or unhelpful assumptions and/or core beliefs through using a thought diary.

Session 3: Introduction to behavioural experiments to test validity of thoughts and develop alternative thoughts.

Session 4: Design and develop more behavioural experiments; complete domain specific self-esteem group therapy blueprint; administer post-intervention questionnaires.

Follow up: Identify problems and solutions to difficulties that may have arisen since the completion of the group. Providing a refresher of session content that

participants may be struggling with or would like to expand on. Administer follow up questionnaires.

Ethical Approval

Ethical approval for the study was sought from UCL research Ethics Committee (Appendix B). Participants were given a Participant Information Sheet (PIS; Appendix D) detailing what the study involved and were asked to provide informed consent (Appendix E). They were reminded that they have the right to withdraw consent at any point in time. Participants were also given a participant code to ensure paperwork and data collection remained anonymised and confidential.

As mentioned above, individuals who scored 3 on item 9 (risk question) on the PHQ-9 during the online screening questionnaire were informed that they were not suitable for the study. They were provided with information on how to seek further help and guidance (e.g. directed to their GP, A&E department, UCL Student Psychological Services, or Samaritans helpline). Participants who scored 1 or 2 on item 9 were eligible for the study but were also given guidance about suicide and self-harm.

At the end of the group intervention, individuals were given a list of psychological support services if they wanted to seek further help (e.g. UCL Student Psychological Services, IAPT services).

As this was an untested self-esteem group programme, we did not suggest to participants that this group would increase their self-esteem when advertising the study. Moreover, participants who felt that the group was unsuitable or unhelpful had the option to discontinue participation at any time. All participants were provided with a list of psychological support providers (e.g. IAPT services) for them to seek psychological support outside of the group, should they wish.

Measures

Demographic details

Participants were asked to provide information about their gender, age, ethnicity, the course they were currently undertaking at UCL, email address and contact telephone number.

Self-Perception Profile for College Students (SPP-CS; Neemann & Harter, 2012)

The SPP-CS (Appendix H) is a validated 54-item domain specific self-esteem measure which comprises 12 specific domains of self-esteem and a measure of global self-worth. The SPP-CS is founded on the theoretical basis that perceptions of the self are reflected in specific domains of one's life (Harter, 1992). Respondents rated themselves on their perceived competence in each domain (e.g. relationships, physical appearance etc.). Each self-esteem domain has four items. The self-esteem items consist of two contrasting statements (e.g., "Some students like the kind of person they are" but "Other students wish that they were different"). Respondents were asked first to decide which statement pertains to them and then to indicate whether the choice is 'really true' or 'sort of true'. Each item was then scored on a 4-point scale (1 = *really negative*, 2 = *sort of negative*, 3 = *sort of positive*, 4 = *really positive*). Scores were aggregated and averaged for each domain. There were good internal consistencies across subscales, with Cronbach's alphas ranging from .76 to .92 (Neemann & Harter, 2012).

The questionnaire also includes importance ratings to assess the importance of each domain to the respondent. The items consist of two contrasting statements (e.g., "Some students feel it's important to be good at athletics" but "Other students do not feel athletics is all that important"). Respondents were asked first to decide which statement pertains to them and then to indicate whether the choice is 'really

true' or 'sort of true'. Each item was then scored on a 4-point scale (1 = *not very important*, 2 = *only sort of important*, 3 = *pretty important*, 4 = *very important*).

Domains that were rated as *very important* to an individual at pre-intervention were considered valued domains; the manual conceptualised that competence scores only affected one's self-worth if the domain is considered *very important* to an individual (Neemann & Harter, 2012). This was determined from a university student population which the measure was normed with (Neemann & Harter, 2012).

Therefore, perceived competence scores were used only for domains that were rated as *very important* to participants. These scores were then averaged for each participant to get a single score for domain specific self-esteem in valued domains. Internal consistencies for importance ratings of domains had Cronbach's alphas that ranged from .53 to .94 (Neemann & Harter, 2012).

Discrepancy scores were determined by calculating the difference between an individual's domain specific self-esteem (i.e. perceived competence) and the importance in domains that are rated as *very important* (Neemann & Harter, 2012). Consistent with the information above, the manual only used domains with an importance of 4 to calculate discrepancy scores as it is conceptualised that competence scores only affected one's self-worth if the domain is considered *very important* to an individual (Neemann & Harter, 2012).

The SPP-CS was chosen because of the similar theoretical basis to the unsatisfactory self-esteem model, which reflects the multidimensionality of self-esteem, the importance placed in domains and the attention to relevant life stages. *Attributional Style Questionnaire (ASQ)*; (Peterson et al., 1982)

The ASQ (Appendix I) is a validated 48-item questionnaire that measures an individual's explanatory style for positive and negative events. The questionnaire is

made up of 12 hypothetical events (6 positive and 6 negative). Each event is followed by four questions: (1) a free-response question about the cause of the hypothetical event, (2) a question about whether the event has an internal or external cause (i.e. the extent respondents believe they themselves are responsible for the event) (3) a question about whether the event has a stable or unstable cause (i.e. the extent respondents believe that the cause of the event is present over time), (4) a question about whether the event has a global or specific cause (i.e. the extent respondents believe the cause of the event occurs across different conditions; Paterson et al., 1982). Scores were tabulated into two categories: Composite Negative Attributional Style (CoNeg) and Composite Positive Attributional Style (CoPos). CoNeg refers to the attributional style towards negative events; the higher the CoNeg score indicates a more internal, stable and global style of attribution towards negative events. CoPos refers to the attributional style towards positive events; the higher the CoPos score indicates a more internal, stable and global style of attribution towards positive events. The CoNeg and CoPos scores were aggregated from the six items in the negative and positive events respectively and subsequently divided by six, with scores ranging from three to 21 (Paterson et al., 1982). CoNeg and CoPos scores were used as they are the most valid and reliable as compared to the individual dimension (i.e. internal, stable, global) scores (Peterson et al., 1982). Good internal consistencies across CoPos and CoNeg scores were reported, with Cronbach's alphas of .75 for positive events and .72 for negative events (Peterson et al., 1982). The ASQ is widely used due to critiques about other unvalidated attributional style measures (e.g. Alloy, 1982; Raps, Peterson, Reinhard, Abramson & Seligman, 1982). The ASQ had been found to have satisfactory criterion (Eaves &

Rush, 1984), convergent (Blaney, Behar, & Head, 1980), and discriminant validity (Raps et al., 1982).

Power calculation

As the study utilised a novel self-esteem group intervention, it was not possible to anticipate exactly what the effect sizes would be. No previous published research used the SPP-CS (Neemann & Harter, 2012) as part of an intervention. A power analysis was therefore informed by considering the results reported for similar CBT self-esteem group interventions. Morton et al. (2012) conducted a CBT group intervention for self-esteem using Fennell's (1997) model and treatment protocol and found a large effect size. However, we were using a new, untested model and intervention (Hollingdale, 2015). Therefore, we anticipated a more conservative effect size of Cohen's $d = 0.5$. A power calculation was carried out using G Power, giving an estimated sample size of 34 participants to provide 80% power with an alpha level of 0.05 for a dependent means design, to detect a medium effect size.

Analysis

Statistical analysis was conducted using the Statistical Package for the Social Sciences Version 24 (SPSS). Domain-specific self-esteem, discrepancy and attributional style scores were tested to see if they met parametric assumptions and transformations were attempted if the variables were not normally distributed.

To address the study's hypotheses, data were analysed in the following steps:

Four separate mixed-model analyses, using Howell's (2015) method (see Appendix K for SPSS syntax) were conducted to assess changes in measures of domain specific self-esteem (i.e. perceived competence) in valued domains, discrepancy scores in valued domains, attributional style for negative events and attributional style for positive events. These changes were assessed between time

points (pre-intervention, post-intervention and one-month follow up), with time being the within-subjects factor. Based on Neemann and Harter's (2012) conceptualisation, valued domains were operationalised as domains that were rated as *very important* (i.e. importance rating = 4) to participants at pre-intervention.

Mixed model analysis was chosen over the General Linear Model as it prevented exclusion of cases where any post intervention or follow up data were missing. Compared to an ANOVA, the mixed model analysis does not remove the other scores from the participant when there is missing data. This therefore allows all the data to be included in the analysis (Howell, 2015). In addition, the mixed model analysis does not require the assumption that the data was missing at random or assume sphericity (Howell, 2015).

In the mixed model analysis, the Akaike Information Criterion (AIC) is a measure of model adequacy. A lower AIC statistic indicates a better fitting model (Howell, 2015). Therefore, in the present study, the model with the lowest AIC statistic was selected as the most appropriate model (i.e. compound symmetry or autoregressive).

Post hoc comparisons were conducted when a statistically significant effect was found, and the Bonferroni correction was used where multiple testing could result in a Type I error inflation.

Effect sizes were calculated for the mixed model analysis and post hoc comparisons. As a standard measure of effect size for this type of model has yet to be established, Cohen's d_z for dependent pairs was calculated where .2 is a small effect, .5 medium and .8 large (Cohen, 1992). The common language effect size indicator (CL) was also computed where significant differences were found (McGraw & Wong, 1992).

Finally, two sets of correlational analyses were conducted using data at pre-intervention to explore the relationship between: 1. domain specific self-esteem in valued domains and attributional style for negative events; 2. domain specific self-esteem in valued domains and attributional style for positive events.

Results

Recruitment and attrition

118 participants completed the online questionnaire and met the eligibility criteria for the study. None of the participants who completed the online questionnaires were excluded based on the exclusion criteria. Depending on the availability of participants to make the scheduled group dates, a total of 89 participants indicated availability on the proposed group dates. A total of five domain specific self-esteem groups were conducted, each consisting of around eight to 12 participants.

A total of 51 participants were assessed at the pre-group, 39 participants were assessed at post-group and 24 participants were assessed at the one-month follow-up. The mean attendance was $M = 3.33$, $SD = 1.39$. The main reasons for missing sessions included: UCL term break, other appointments or activities and being unwell. The CONSORT diagram for the recruitment process is shown in Figure 4.

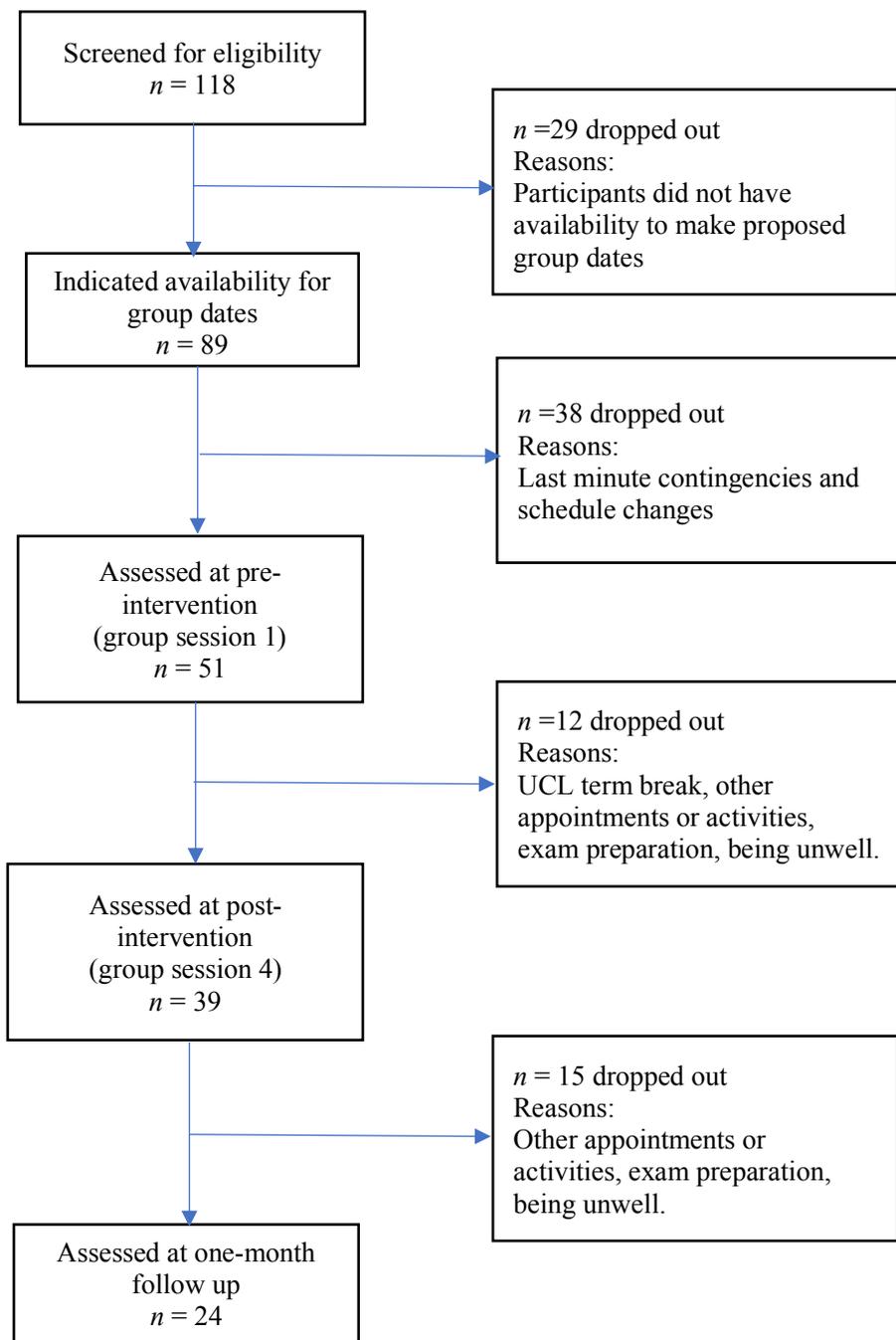


Figure 4. COSORT diagram of the recruitment process

Participant demographics

Table 1 summarises the demographic data of participants in the study. In the overall sample of 51 participants, there were 43 females (84.3%) and 8 males (15.7%). Participant age ranged from 17 to 52 years old ($M = 23.96$, $SD = 7.32$), 32 (62.7%) were undergraduate and 19 (37.3%) were postgraduate students at UCL.

Table 1

Sample Characteristics

	<i>Sample</i>
<i>N</i>	51
Age (M, SD) in years	23.96 (7.32)
Gender (number, %)	Female 43 (84.3%), Male 8 (15.7%)
Student Status (number, %)	
Undergraduate	32 (62.7%)
Postgraduate	19 (37.3%)

Intervention Outcomes

Table 2 shows the descriptive statistics relating to the primary outcome hypotheses.

Table 2

State measure means and standard deviations for participants at pre-, post-intervention and one-month follow up

Measure	Pre <i>M (SD)</i>	Post <i>M (SD)</i>	Follow up <i>M (SD)</i>	Pre-post <i>p</i>	Effect size pre-post (<i>d</i> ₁)	Effect size post-follow up (<i>d</i> ₂)
Domain specific self-esteem (i.e. perceived competence)	2.39 ^a (.49) .18 ^b (.15)	2.58 ^a (.52) .13 ^b (.16)	2.65 ^a (.56) .10 ^b (.17)	.01*	.49	.00
Discrepancy score	1.61 (.55)	1.06 (.60)	.98 (.68)	< .001*	1.14	.00
Attributional Style for negative events (CoNeg)	14.85 (2.38)	13.75 (2.55)	12.88 (2.86)	.001*	.56	.48
Attributional Style for positive events (CoPos)	13.14 (2.43)	13.17 (2.58)	13.09 (3.01)	0.98	.00	.00

^a= Non transformed means and standard deviations. ^b= Transformed means and standard deviations. * $p < .05$

Domain Specific Self-Esteem in Valued Domains

A mixed model analysis was conducted to compare the effect of time on domain specific self-esteem (i.e. perceived competence) scores in valued domains. This was done to investigate the first hypothesis which predicted that domain specific self-esteem (i.e. competence evaluations) in valued domains would increase post-intervention and that this change would be maintained at follow-up. Domain specific self-esteem scores were inversely transformed to address non-normality of the variable. In the mixed model analysis, AIC for compound symmetry was -136.24 and for autoregressive was -135.22, therefore the compound symmetry model was chosen due to the smaller AIC.

The effect of time on domain specific self-esteem scores was significant $F(2, 60.59) = 9.38, p < .001$. A Bonferroni post hoc comparison revealed that there was a significant difference in domain specific self-esteem at pre-intervention ($M = .18, SD = .15$) and post-intervention ($M = .13, SD = .16$), ($p = .01$) 95% CI [.01, .1], Cohen's $d_z = .49$ and follow-up ($M = .10, SD = .17$), ($p < .001$) 95% CI [.03, .14], Cohen's $d_z = .73$. Domain specific self-esteem scores at the one-month follow-up did not differ from post intervention ($p = .43$), indicating that the initial decrease in scores were sustained over that period. Given that this data was inversely transformed to address non-normality of the variable, the mean perceived competence scores following the intervention was shown as lower compared to baseline. This implied that domain specific self-esteem in valued domains increased at post intervention and was maintained at follow-up. The CL effect sizes indicate that the chance of a randomly selected participant rating their domain specific self-esteem score higher post-intervention compared to pre-intervention was 69%. Similarly, the probability that a

randomly selected participant scored higher domain specific self-esteem at follow-up compared to baseline was 77%.

Discrepancy Scores in Valued Domains

A mixed model analysis was conducted to compare the effect of time on discrepancy scores in valued domains. This was done to investigate the second hypothesis which predicted that the discrepancy between domain specific self-esteem (i.e. perceived competence) and importance placed in valued domains would decrease post-intervention and this change would be maintained at follow-up. Discrepancy scores were calculated through the difference between the domain specific self-esteem (i.e. perceived competence) scores and importance scores. In the mixed model analysis, AIC for compound symmetry was 146.60 and for autoregressive was 150.11, therefore the compound symmetry model was chosen due to the smaller AIC.

The analysis found that the effect of time on discrepancy scores was significant $F(2, 61.3) = 39.2, p < .001$. A Bonferroni post hoc comparison revealed that there was a significant difference in discrepancy scores at pre-intervention ($M = 1.61, SD = .55$) and post-intervention ($M = 1.06, SD = .60$), ($p < .001$) 95% CI [.37, .72], Cohen's $d_z = 1.14$ and follow-up ($M = .98, SD = .68$), ($p < .001$) 95% CI [.41, .83], Cohen's $d_z = 1.23$. The mean discrepancy score at the one-month follow-up did not differ from post intervention ($p = 1.00$), indicating that the initial decrease in discrepancy scores was sustained over that period. The CL effect sizes indicate that the chance that a randomly selected participant had a smaller discrepancy score at post-intervention compared to baseline was 87%. Similarly, the probability that a randomly selected participant had a smaller discrepancy score at follow-up compared to baseline was 89%.

A sensitivity analysis was also conducted on importance scores to analyse whether the scores had changed at post-intervention and follow up. This was because one potential hypothesis was that the reason for change in discrepancy scores might also be due to the importance scores changing. As the study only examined domains that were *very important* to participants, the distribution of importance scores was heavily skewed and non-normal. Transformation of the data was not possible and there was no nonparametric equivalent for mixed model analysis. A nonparametric equivalent of a repeated measures ANOVA (Friedman's test) was therefore used to conduct a sensitivity analysis on the importance scores. Data were excluded listwise if there were incomplete data. A non-parametric Friedman's test was conducted with $n = 23$, which rendered a Chi-square value of 28.32 which was significant ($p < .001$). A Bonferroni post hoc comparison using the Wilcoxon Signed-Ranks test indicated that post-intervention importance scores, $Mdn = 3.75$ were significantly lower than pre-intervention importance scores $Mdn = 4$, $Z = -4.86$, $p < .001$. The follow up importance scores, $Mdn = 3.67$ were significantly lower than pre-intervention importance scores $Mdn = 4$, $Z = -3.93$, $p < .001$. There was no significant difference between the post-intervention importance scores and follow up importance scores ($p = 2.04$). A caveat about the sensitivity analysis was that removing so much data might have affected the results, so conclusions related to this analysis would be necessarily tentative.

Attributional Style for Negative Events

A mixed model analysis was conducted to compare the effect of time on CoNeg scores. This was conducted to investigate the third hypothesis which predicted that attributional styles for negative events would shift from internal, stable and global styles towards more external, unstable and specific styles post

intervention and this change would be maintained at follow-up. In the mixed model analysis, AIC for compound symmetry was 451.78 and for autoregressive was 457.47, therefore the compound symmetry model was chosen due to the smaller AIC.

The mixed models analysis found that the effect of time on CoNeg scores was significant $F(2, 62.49) = 17.8, p < .001$. A Bonferroni post hoc comparison revealed that there was a significant difference in CoNeg scores at pre-intervention ($M = 14.85, SD = 2.38$) and post-intervention ($M = 13.75, SD = 2.55$), ($p = .001$) 95% CI [.39, 1.81], Cohen's $d_z = .56$ and follow-up ($M = 12.88, SD = 2.86$), ($p < .001$) 95% CI [1.13, 2.81], Cohen's $d_z = .99$. There was also a significant difference in CoNeg scores at post-intervention ($M = 13.75, SD = 2.55$) and follow-up ($M = 12.88, SD = 2.86$), $p = .04$, 95% CI [.02, 1.73], Cohen's $d_z = .48$, indicating that CoNeg scores continued to decrease post-intervention to follow-up. The CL effect sizes indicate that the chance of a randomly selected participant rating their attributional style for negative events as more external, unstable and specific post-intervention compared to pre-intervention was 71%. Similarly, the probability that a randomly selected participant rating their attributional style for negative events as more external, unstable and specific at follow-up compared to baseline was 83%. Finally, the probability that a randomly selected participant rating their attributional style for negative events as more external, unstable and specific at follow-up compared to post-intervention was 68%.

Attributional Style for Positive Events

A mixed model analysis was conducted to compare the effect of time on CoPos scores. This was done to investigate the fourth hypothesis which predicted that attributional styles for positive events would shift from external, unstable and

specific styles towards more internal, stable and global styles post intervention and this change would be maintained at follow-up. In the mixed model analysis, AIC for compound symmetry was 458.01 and for autoregressive was 456.29, therefore the autoregressive model was chosen due to the smaller AIC.

The mixed models analysis found that the effect of time on CoPos scores was non-significant $F(2, 64.51) = 0.02, p = .98$. Bonferroni corrected post-hoc comparisons also confirmed that the CoPos scores did not differ from each other across pre-intervention ($M = 13.14, SD = 2.43$), post-intervention ($M = 13.17, SD = 2.58$) and follow-up ($M = 13.09, SD = 3.01$).

Relationship between Attributional Styles and Domain Specific Self-esteem

To test the fifth hypothesis, which predicted that individuals who adopt more internal, stable and global attributional styles towards negative events would be more likely to experience deficits in their domain specific self-esteem in valued domains, a Pearson's correlational analysis was conducted between domain specific self-esteem in valued domains and CoNeg scores for the sample ($n = 48$). Domain specific self-esteem scores in valued domains were significantly negatively correlated with CoNeg scores ($r = -.42, p = .003$).

To test the sixth hypothesis, which predicted that individuals who adopt more external, unstable and specific attributional styles towards positive events would be more likely to experience deficits in their domain specific self-esteem in valued domains, a Pearson's correlational analysis was conducted between domain specific self-esteem scores in valued domains and CoPos scores for the sample ($n = 48$). Domain specific self-esteem scores in valued domains were not significantly correlated to CoPos scores ($r = .20, p = .18$).

Discussion

Summary of findings

This, to our knowledge, is the first study that evaluated a CBT group intervention targeted at domain specific self-esteem. This intervention builds on the idea of the multidimensional nature of self-esteem that has been evidenced in literature (Byrne, 1996). While past research on CBT group interventions for self-esteem targeted global self-esteem (e.g. Morton et al., 2011; Rigby & Waite, 2006), the current domain specific self-esteem group focussed on the multidimensional aspect of the construct by intervening in valued domains in which participants had unsatisfactory self-esteem in.

The main aims of the study were, following from the group intervention, to assess changes in domain specific self-esteem in valued domains, discrepancy between domain specific self-esteem (i.e. perceived competence) and importance placed in valued domains, and attributional styles towards negative and positive events. In addition, the relationship between domain specific self-esteem and attributional styles was examined. The results of this preliminary study provide encouraging evidence for the four-session domain specific self-esteem group intervention.

The overall results indicate that participants who attended the domain specific self-esteem group showed improvement in domain specific self-esteem (i.e. perceived competence) in their valued domains. In addition, the discrepancy between perceived competence and importance placed in valued domains decreased after the intervention. These findings were observed at post-intervention and maintained at one month follow up. These were consistent with the study's first and second hypotheses.

In accordance with the third hypothesis of the study, attributional styles towards negative events showed a shift towards more external, unstable and specific styles after the intervention and continued moving in this direction at the one-month follow up. Contrary to what was predicted in the fourth hypothesis of the study, no changes were observed in attributional styles towards positive events.

Finally, a significant relationship was found between domain specific self-esteem and attributional styles towards negative events, which was consistent with the study's fifth hypothesis. However, contrary to the study's sixth hypothesis, no relationship was found between domain specific self-esteem and attributional styles towards positive events.

Changes in Domain Specific Self-Esteem, Discrepancy Scores and Attributional Styles

Reasons for improvements observed

Although the specific mechanism explaining the improvements observed in domain specific self-esteem, discrepancy scores and attributional styles had not been examined in this study, a number of hypotheses can be made. Firstly, participants' awareness of their personal domain specific self-esteem profile set the foundation for goal directed behaviour change. Cognitive behaviour therapy is goal oriented and problem focused (Beck, 1976). Participants were able to identify valued domains in which they had unsatisfactory self-esteem, which provided a problem focus that was meaningful for goal setting and intervention.

The process of collaborative empiricism (Beck, 1976) formed the basis of the work with participants in promoting change. A collaborative formulation for each participant was used to identify and explore possible antecedents, triggers and current distress in regard to domain specific self-esteem and its maintaining factors.

The group facilitators and participants collaboratively identified maladaptive cognitions and behaviours to test. The facilitators had also met with each participant individually within the group sessions to plan behavioural experiments that were relevant and meaningful to them. This collaborative process might have resulted in an increase in participants' motivation towards change (Beck et al., 1979).

In the general cognitive behavioural framework, participants were introduced to CBT techniques that underpinned the intervention, such as identifying negative automatic thoughts in a process that culminated in 'challenging' them through behavioural experiments (Beck, 1976). The use of these techniques was centred on the valued domains in which participants had unsatisfactory self-esteem in. This provided a focus for each participant to identify negative automatic thoughts and plan behavioural experiments, a powerful method that can bring about change in cognitive therapy. Research conducted by Bennett–Levy (2003) found that behavioural experiments were rated as having promoted greater cognitive, affective, and behavioural change compared to purely verbal cognitive techniques that lacked an experiential component.

Along with the domain specific self-esteem profile, participants were also provided with their attributional styles profile, which gave an indication of participants' general tendencies when attributing the causes for positive and negative events. This might have been useful in unpicking participants' unhelpful thinking styles when examining their automatic thoughts. Moreover, through the experience of conducting behavioural experiments, participants may have learnt that negative outcomes might not be contingent on acts in their repertoires. Instead, they might be due to the external situation, which reflect less internal, global and stable causes, resulting in attributional style changes.

Finally, participants in the domain specific self-esteem group were also encouraged continually to learn through reflection and practice and to continually devise behavioural experiments after the group ended. The one-month follow up session provided a space for participants to reflect on challenges faced during the implementation of the skills acquired. This might have further consolidated learning and ensured that intervention gains were maintained thereafter.

Attributional styles towards positive events

Contrary to what was predicted, no changes were observed in attributional styles towards positive events at post-intervention and one-month follow up. As the group was tailored towards intervening in self-esteem domains deemed as “unsatisfactory” to participants, most participants identified negative events and the related thoughts, feelings and behaviours associated with them. For example, some negative events included doing poorly in an exam, receiving poor feedback from supervisors or being rejected by friends. It was rare for participants to mention positive events associated with these domains.

It might be hypothesised that individuals with unsatisfactory self-esteem in valued domains pay more attention to negative events but are less concerned with positive events in the prescribed domains. Therefore, there might have been a general lack of examination of participants’ attribution toward positive events during the group intervention. This might have resulted in no changes being observed for attributional styles towards positive events at post-intervention and one-month follow up.

Changes in Importance Scores

The sensitivity analysis conducted on the importance scores utilising the nonparametric Friedman’s test indicated a significant decrease in importance scores

in valued domains at post-intervention and maintained at one-month follow up. This was an interesting finding that warrants further investigation. This implied that the importance placed in participants' initially valued domains had become less important following the group intervention. Some participants had mentioned in their qualitative feedback that they valued some of these domains as less important compared to pre-intervention (Dixon, 2018). This might provide an explanation that for some participants, a combination of both an increase in domain specific self-esteem and a decrease in the importance placed in initially valued domains resulted in a lower discrepancy score between the two at post-intervention.

Future studies should examine the relationship between domain specific self-esteem and the importance placed in valued domains. Moreover, studies should examine changes in importance of participants' domains throughout the group intervention, which might shed light on the effect the intervention has on the importance placed on domains. It should be noted nonetheless that these present findings indicating the decrease in importance scores should be interpreted with caution due to the small sample size included in the analysis.

Relationship between Attributional Styles and Domain Specific Self Esteem

Past research established a relationship between global self-esteem and attributional styles (Abramson et al., 1978; Feather, 1983; Ickes & Layden, 1978; Seligman et al., 1979). Specifically, deficits in self-esteem have been associated with attributions of negative events to internal, stable and global causes (Abramson et al., 1978; Feather, 1983; Ickes & Layden, 1978; Seligman et al., 1979). The current study expands on this finding to suggest that individuals with lower domain specific self-esteem scores in their valued domains also tend to attribute negative events to more internal, stable and global causes. For example, an individual who has

unsatisfactory self-esteem in academic performance and experiences failure in that domain would tend to attribute the failure a lack of ability that is recurrent across time and situations. This might be due to a sense of personal helplessness felt in valued domains (Abramson et al., 1978).

This relationship, however, did not seem to hold for attributional styles for positive events and domain specific self-esteem. This was contrary to what the study initially hypothesised. The result seems to suggest no relationship for domain specific self-esteem and attributional styles towards positive events. Similar to the earlier discussions, one possible explanation is that individuals with unsatisfactory self-esteem pay more attention to negative events but are less concerned with positive events in the prescribed domains. Presently, the literature also indicates mixed findings for the relationship between global self-esteem and attributional styles towards positive events (e.g. Feather, 1983; Tennen & Herzberger, 1987; Tennen et al., 1987). Nonetheless, the absence of significant correlations between domain specific self-esteem and attributional styles for positive events does not necessarily mean that they do not exist. Further research is required to better understand the relationship between domain specific self-esteem and attributional styles.

Limitations and Future Directions

Methodological Issues

The present study has a number of limitations which are important to consider. The one-group pretest posttest design, although appropriate as a feasibility study, was limited as it lacked a control group. Without a control group, possible threats to internal and construct validity cannot be discounted (Barker, Pistrang, & Elliott, 2016). For example, spontaneous remission is one possible explanation for

symptom reduction that might account for the improvement in scores with an uncontrolled study design. Another possible threat to internal validity would be expectancy effects, where participants might have benefitted from the group because they expected to, rather than as the result of the group intervention (Barker, et al., 2016). Researcher and participant expectations prior to the group intervention should be formally assessed to reduce bias in the future. Other possible factors for bias such as the use of medication and previous or current psychological treatments should also be assessed in future studies.

Having a randomised control trial will pit the current domain specific self-esteem intervention against treatment as usual (TAU) for self-esteem to determine if the group intervention has an effect over and above the current treatments. For example, various Improving Access to Psychological Therapies (IAPT) services employ group or workshop-based interventions for improving self-esteem. Even though these are targeted at global self-esteem, it will be useful to determine if the domain specific self-esteem group intervention provides additional value beyond these existing interventions. Another limitation of the study is that the groups were facilitated by the two researchers conducting the study. This might have the potential to introduce demand characteristics and experimenter effects in participant responses, which might subsequently bias the data either consciously or unconsciously (Rosnow & Rosenthal, 1997). The present group employed a rigorous supervision framework that allowed for the reflection on possible issues. Future replication studies would benefit from employing other measures such as having independent researchers administer the questionnaires and collate the data. Additionally, having a formal measure of facilitator adherence would improve the intervention's fidelity.

Further research is also necessary to ascertain the durability of the effect that was observed after the one-month follow up. Due to the time constraints of the project, investigations into whether the treatment gains persisted for a longer period were not feasible. It might be that gains at a three-month or six-month follow up would be maintained, but that they could also dissipate. The effects of the group might maintain or grow as participants continue to implement the skills and techniques learnt during the intervention. The incorporation of ‘booster’ sessions might also sustain the clinical improvements observed. Future studies therefore should consider administering measures at various time points following the end of the group to investigate the durability of the intervention effects.

Participants were not assessed for the presence of any comorbidity with other psychiatric conditions. An understanding of comorbidity might provide insight to the effectiveness of a transdiagnostic intervention such as the present domain specific self-esteem group with different mental health diagnoses. Past research had found high levels of comorbidity between low self-esteem and other mental health diagnoses (Waite et al., 2012). There is the evidence in literature suggesting that the relationship between low self-esteem and psychiatric disorders may be circular, suggesting that self-esteem can be both an aetiological factor and a maintaining factor in mental health disorders (Waite et al., 2012). Therefore, further research to examine the effectiveness of the present group intervention in the context of psychiatric comorbidity would be valuable.

The study’s patient flow showed that a considerable number of participants did not complete the group. From the 51 participants that were present at the first session, only 39 attended the final session and 24 participants were assessed at one-month follow up. This indicated an attrition rate of 23.5% from pre-intervention to

post-intervention and 38.5% from post-intervention to one-month follow up.

Participants had cited reasons such as student activities, classes, the school term breaks and upcoming exams as the main reasons they were unable to attend sessions.

While none of the participants cited any reasons pertaining to the group itself that might have resulted in them dropping out, future studies might benefit from having an anonymous drop out survey that might provide a wider range of possible explanations for the attrition rate.

The majority of participants were female (84.3%), and therefore the recruitment of males was lacking. Possible future studies might benefit from examining potential barriers to accessing help through the domain specific self-esteem group and to determine barriers that men might face when accessing a group such as this. In addition, participants were university students so there is an issue of the extent to which the current findings are generalisable to a clinical population, both in terms of the relative clinical profiles and the relative ranges of intellectual abilities of the two populations.

Finally, some hypotheses were suggested in the previous section about the elements of the group that might have contributed to changes in domain specific self-esteem and attributional styles. Further research is required to determine the relative contribution of different components of the group intervention to the various improvements observed. This might be achieved through weekly participant ratings of intervention components of the group sessions. This could be conducted through the collection of both quantitative and qualitative feedback.

Reliable and Clinically Significant Change

Reliable and clinically significant changes are important concepts to consider when interpreting the results of a study (Jacobson & Truax, 1991). Reliable change is

defined as changes in scores observed that is unlikely due to measurement unreliability (Jacobson & Truax, 1991). This is determined through statistically significant changes observed in participants' scores in a measure post-intervention and follow up. This was found in the domain specific self-esteem group intervention for changes observed in domain specific self-esteem scores, discrepancy scores and attributional styles towards negative events.

Clinically significant change is defined as the extent to which an intervention results in scores moving outside the range of a 'dysfunctional' population or within the range of the 'functional' population (Jacobson & Truax, 1991). This is usually determined by a clinical cut-off score on a measure. One limitation of the SPP-CS is the lack of predefined norms for clinical cut-offs in the university population that was examined. Therefore, it is unclear whether the present findings indicating statistically significant changes had meaningful clinical change too. Future studies should therefore examine clinical change cut-offs in relation to the SPP-CS in order to determine clinically significant changes.

Relatedly, the current study is limited in determining whether participants initially fell below the 'clinical' threshold for unsatisfactory self-esteem in valued domains. There might have been some participants who were above the 'clinical' threshold to begin with. Conclusions as to whether the intervention actually moves individuals from the 'clinical' range to the 'functional' range is therefore uncertain. Having a sample of participants who initially score below 'clinical' threshold is important in future studies in order to assess if scores move out of the 'clinical' range following the intervention.

Instrument Limitations

While the SPP-CS had good subscale internal consistencies for competency evaluations in domains (Neemann & Harter, 2012), internal consistencies for some of the importance subscales were not as high. Moreover, the structural validity of the measure seems to require further validation. While Neemann and Harter (2012) deemed the factor structures of subscales to be appropriate after conducting a Principal Components Analysis (PCA), it was difficult to determine the extent each factor was clearly defined as cross loadings were not reported. Moreover, there are widely recognised limitations using PCA (e.g. Gorsuch, 1990; Hubbard & Allen, 1987; Snook & Gorsuch, 1989). While the SPP-CS fills the gap in instruments measuring domain specific self-esteem in a university population, it would benefit from further validation of its factor structure through Confirmatory Factor Analysis (CFA).

Furthermore, while the ASQ has been psychometrically well validated compared to other attributional style measures (Peterson et al., 1982; Tennen & Herzberger, 1986), it does present some conceptual limitations. Firstly, the ASQ employs hypothetical situations in the questionnaire to assess attributional styles, which might impose internal biases as compared to reporting actual causes for real events (Seligman, 1985, cited in Tennen et al., 1987). Moreover, the 16 hypothetical situations in the ASQ are also limited in providing a range of events associated with domain specific self-esteem. While the hypothetical situations cover domains such as romantic relationships, friendships, work and physical appearance, the instrument lacks other domains that might be relevant to the current student sample, for example, academic performance and athleticism.

In summary, more research is needed to examine the potential value of the domain specific self-esteem group. Current limitations should be addressed in future research by: using a control group comparison; examining the effect durability of the group; examining intervention elements contributing to outcomes; having independent researchers administering questionnaires; incorporating an adherence to the group checklist; examining the impact of group in the context of psychiatric comorbidity; further validation of the instruments; and capturing perceptions of the group through a qualitative methodology.

Clinical Implications

The domain specific self-esteem group intervention is the first of its kind that is conceptualised to target domain specific self-esteem. As self-esteem is a deeply personal and complex construct, the reconceptualisation of self-esteem from “low” and “high” to being “satisfactory” or “unsatisfactory” for an individual’s needs can be more meaningful for clients in the clinical setting. Although giving clients a general label of low self-esteem can be helpful sometimes, it can be arbitrary and difficult to interpret. Instead, the unsatisfactory self-esteem model promotes an understanding of the multidimensionality of self-esteem and help clients make sense of the difficulties they face within the various domains (e.g. career, relationships, etc.).

The domain specific self-esteem profile allows clients and clinicians to identify specific domains of self-esteem that might be considered “unsatisfactory” for their needs. This is valuable when planning specific treatment goals with regard to particular life domains, in comparison to the broader self-esteem goals traditional self-esteem approaches employ (e.g. Fennell, 1997). Treatment goals can therefore be easily identified with the client based on the domains with which clients are

struggling. These goals can be subsequently monitored and reviewed throughout the course of treatment.

Moreover, a key concept of the unsatisfactory self-esteem model is the value or importance individuals place in life domains. Clients and clinicians can therefore prioritise domains for change. For example, although a client's perceived competence score might be low in the academic domain of self-esteem, it might not be meaningful to work on academic self-esteem if it is deemed unimportant to the client during that time in their lives. Therefore, identifying a client's valued domains would create a meaningful focus for treatment. This will also likely to improve clients' engagement in the intervention and increase motivation for change.

The group intervention also encourages the collaborative exploration of self-esteem domains in which clients identify and acknowledge as "satisfactory" for their needs. This can provide insight to a client's existing strengths and ways that their self-esteem is maintained in these domains. Therefore, identifying these strengths during treatment might provide ideas and insight into ways clients can improve their self-esteem in domains in which they deemed as "unsatisfactory".

The attributional styles profile for clients might also be valuable for identifying particular cognitive biases clients have. These biases would usually be elicited through the 'homework' exercises (e.g. thought diaries) that the intervention employs. The awareness of one's attributional style might enable clients to identify possible cognitive distortions when attributing causes to negative events. For example, a client might realise that they have a general tendency to attribute negative events in the domain of romantic relationships to internal, stable and global causes. Patterns of these attributions can be identified, and the validity of these assumptions can be subsequently tested through behavioural experiments.

Finally, it is typical that clients present with various kinds of diagnostic comorbidity in routine clinical practice. There is currently little evidence to guide clinicians in deciding how to structure or combine interventions for clients who meet the criteria for multiple psychiatric diagnoses (Harvey, Watkins, Mansell, & Shafran, 2004). Given the evidence that low self-esteem is an aetiological and maintaining factor in various psychiatric diagnoses (Waite et al., 2012), a transdiagnostic intervention such as the present domain specific self-esteem group might be beneficial to be used across multiple diagnoses in addition to existing evidenced based CBT for specific disorders. Further research is necessary to determine the effectiveness of this group as a single pathway intervention for psychiatric comorbidity.

Conclusion

The present study examined a novel CBT group intervention for domain specific self-esteem. As difficulties with self-esteem are closely linked with poor mental health and wellbeing, developing and evaluating effective treatments is critical. The present group intervention appears to provide preliminary evidence of clinical benefits such as improvements in domain specific self-esteem and attributional styles towards negative events. Moreover, the treatment gains appear durable, at least for a month after the intervention ended. Future research should focus on evaluating this intervention in a controlled trial to understand how this intervention might fare against conventional CBT methods.

References

- Abramson, L. Y., Seligman, M. E. P., & Teasdale, I. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*(1), 49–74. <https://doi.org/10.1037//0021-843X.87.1.49>
- Åkerlind, I., Hernquist, J. O., & Bjurulf, P. (1988). Prognosis in alcoholic rehabilitation: The relative significance of social, psychological, and medical factors. *Substance Use and Misuse, 23*(11), 1171–1195. <https://doi.org/10.3109/10826088809056193>
- Alloy, L. B. (1982). The role of perceptions and attributions for response-outcome noncontingency in learned helplessness: A commentary and discussion. *Journal of Personality, 50*(4), 443–479. <https://doi.org/10.1111/j.1467-6494.1982.tb00229.x>
- Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology: An introduction for students and practitioners* (Third ed.).
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders. Cognitive therapy and the emotional disorders* (Vol. 8). New York: International Universities Press. [https://doi.org/10.1016/S0005-7894\(77\)80293-1](https://doi.org/10.1016/S0005-7894(77)80293-1)
- Bennett-Levy, J., Butler, G., Fennell, M., Hackman, A., Mueller, M., & Westbrook, D. (2004). *Oxford guide to behavioural experiments in cognitive therapy*. New

York, NY, US: Oxford University Press.

<https://doi.org/http://dx.doi.org/10.1093/med:psych/9780198529163.001.0001>

Bennett-Levy, J. (2003). Mechanisms of change in cognitive therapy: The case of automatic thought records and behavioural experiments. *Behavioural and Cognitive Psychotherapy*, 31(3), 261–277.

<https://doi.org/10.1017/S1352465803003035>

Blaney, P. H., Behar, V., & Head, R. (1980). Two measures of depressive cognition: Their association with depression and with each other. *Journal of Abnormal Psychology*, 89, 678–682.

Brown, G. W., Andrews, B., Harris, T., Adler, Z., & Bridge, L. (1986). Social support, self-esteem and depression. *Psychological Medicine*, 16(4), 813.

<https://doi.org/10.1017/S0033291700011831>

Brown, G. W., Bifulco, A., & Andrews, B. (1990). Self-esteem and depression - IV. Effect on course and recovery. *Social Psychiatry and Psychiatric Epidemiology*, 25(5), 244–249. <https://doi.org/10.1007/BF00788645>

Butler, G., Fennell, M., & Hackmann, A. (2008). *Cognitive-behavioral therapy for anxiety disorders: Mastering clinical challenges*. New York; London: Guilford Press. Retrieved from

<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2008-10718-000>

Button, E. J., Sonuga-Barke, E. J. S., Davies, J., & Thompson, M. (1996). A prospective study of self-esteem in the prediction of eating problems in

adolescent schoolgirls: Questionnaire findings. *British Journal of Clinical Psychology*, 35(2), 193–203. <https://doi.org/10.1111/j.2044-8260.1996.tb01176.x>

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*, 54, 427–456.

Byrne, B. M. (1996). *Measuring self-concept across the lifespan*. Washington, DC: American Psychological Association.

Chatterton, L., Hall, P. L., & Tarrier, N. (2007). Cognitive therapy for low self-esteem in the treatment of depression in an older adult. *Behavioural and Cognitive Psychotherapy*, 35, 365–369.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>

Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: Freeman.

Dixon, E. (2018). *A feasibility study of a Cognitive Behavioural Therapy group for domain-specific self-esteem and its impact on global self-esteem, depression, anxiety and psychological wellbeing*. Unpublished clinical psychology doctoral thesis, Department of Clinical, Educational, and Health Psychology, University College London.

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-esteem: Enduring issues and controversies. In *The Wiley-Blackwell Handbook of*

Individual Differences (pp. 718–746).

<https://doi.org/10.1002/9781444343120.ch28>

Eaves, G., & Rush, A. J. (1984). Cognitive patterns in symptomatic and remitted unipolar major depression. *Journal of Abnormal Psychology, 93*, 31–40.

Ehnholt, K. a, Salkovskis, P. M., & Rimes, K. a. (1999). Obsessive-compulsive disorder, anxiety disorders, and self-esteem: an exploratory study. *Behaviour Research and Therapy, 37*(8), 771–81. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10452177>

Feather, N. T. (1983). Some correlates of attributional style: Depressive symptoms, self-esteem, and Protestant Ethic values. *Personality & Social Psychology Bulletin, 9*(1), 125–135. <https://doi.org/10.1177/0146167283091018>

Fennell, M. J. V. (2009). Overcoming low self-esteem : a self-help guide using cognitive behavioral techniques. *Overcoming Series*, xi, 366 p.

Fennell, M. J. V. (1997). Low Self-Esteem: A cognitive perspective. *Behavioural and Cognitive Psychotherapy, 25*(1), 1–25. <https://doi.org/10.1017/S1352465800015368>

Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, self-esteem, and anger in emerging adulthood: Seven-year trajectories. *Developmental Psychology, 42*(2), 350–365. <https://doi.org/10.1037/0012-1649.42.2.350>

- Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25(1), 33–39. https://doi.org/10.1207/s15327906mbr2501_3
- Grant, A. M., Franklin, J., & Langford, P. (2002). The Self-Reflection and Insight Scale: A new measure of private self-consciousness. *Social Behavior and Personality: An International Journal*, 30(8), 821–835. <https://doi.org/10.2224/sbp.2002.30.8.821>
- Gual, P., Prez-Gaspar, M., Martnez-Gonzlez, M. A., Lahortiga, F., De Irala-Estvez, J., & Cervera-Enguix, S. (2002). Self-esteem, personality, and eating disorders: Baseline assessment of a prospective population-based cohort. *International Journal of Eating Disorders*, 31(3), 261–273. <https://doi.org/10.1002/eat.10040>
- Hall, P. L., & Tarrier, N. (2003). The cognitive-behavioural treatment of low self-esteem in psychotic patients: A pilot study. *Behaviour Research and Therapy*, 41(3), 317–332. [https://doi.org/10.1016/S0005-7967\(02\)00013-X](https://doi.org/10.1016/S0005-7967(02)00013-X)
- Harter, S. (2012). *The construction of the self: Developmental and sociocultural perspectives*. New York: Guilford Press.
- Harter, S. (1985). *The self-perception profile for children: Revision of the perceived competence scale for children*. Unpublished manuscript, University of Denver, Colorado, USA.
- Harter, S., Whitesell, N., & Kowalski, P. (1986). *The impact of educational transitions upon children's perceived competence and motivational orientation*. Unpublished manuscript, University of Denver.

- Harter, S. (1990). Issues in the assessment of the self-concept of children and adolescents. In A. M. La Greca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 292–325). Needham Heights, MA, US: Allyn & Bacon.
- Harter, S., & Whitesell, N. R. (2003). Beyond the debate: Why some adolescents report stable self-worth over time and situation, whereas others report changes in self-worth. *Journal of Personality*, *71*(6), 1027–1058.
<https://doi.org/10.1111/1467-6494.7106006>
- Harvey, A. G., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford Oxford University Press. Oxford: Oxford University Press. <https://doi.org/10.1002/erv.647>
- Hollingdale, J. (2015). CBT for self-esteem: An Introduction to a new model. Unpublished presentation given at University College London. University College London, London.
- Howell, D. C. (2015). Treatment of missing Data-part 1. Retrieved from https://www.uvm.edu/~dhowell/StatPages/Missing_Data/Missing.html
- Hubbard, R., & Allen, S. J. (1987). A cautionary note on the use of principal components analysis. *Sociological Methods & Research*, *16*(2), 301–308.
<https://doi.org/10.1177/0049124187016002005>

- Ickes, W., & Layden, M. A. (1978). Attributional styles. In H. J. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research (Vol. 2)*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Jacobson, N. S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.
- James, W. (1982). *Psychology: The briefer course*. New York: Holt, Rinehart, & Winston.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology, 51*(6), 1224–1236.
- Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*, 107–125.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>

- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. <https://doi.org/10.1037/0033-2909.111.2.361>
- McManus, F., Waite, P., & Shafran, R. (2009). Cognitive-Behavior Therapy for low self-esteem: A case example. *Cognitive and Behavioral Practice*, *16*(3), 266–275. <https://doi.org/10.1016/j.cbpra.2008.12.007>
- Morton, L., Roach, L., Reid, H., & Stewart, S. H. (2012). An evaluation of a CBT group for women with low self-esteem. *Behavioural and Cognitive Psychotherapy*, *40*(2), 221–225. <https://doi.org/10.1017/S1352465811000294>
- Mruk, C. J. (2006). *Self-esteem research, theory and practice: Toward a positive psychology of self-esteem*. New York: Springer Publishing Company.
- Neemann, J., & Harter, S. (2012). *Self-Perception Profile for College Students: Manual and questionnaires*. Unpublished manuscript, University of Denver, Colorado, USA.
- O'Brien, E. J., Bartoletti, M., & Leitzel, J. D. (2006). Self-Esteem, psychopathology, and psychotherapy. In M. H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 306–315). New York, NY, US: Psychology Press. <https://doi.org/10.4324/9780203759745>
- Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-Esteem development From young adulthood to old Age: A cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, *98*(4), 645–658. <https://doi.org/10.1037/a0018769>

- Paradise, A. W., & Kernis, M. H. (2002). Self-esteem and psychological well-being: Implications of fragile self-esteem. *Journal of Social and Clinical Psychology, 21*(4), 345–361. <https://doi.org/10.1521/jscp.21.4.345.22598>
- Peterson, C., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. P. (1982). The Attributional Style Questionnaire. *Cognitive Therapy and Research, 6*, 287–300. <https://doi.org/10.1007/BF01173577>
- Raps, C. S., Peterson, C., Reinhard, K. E., Abramson, L. Y., & Seligman, M. E. P. (1982). Attributional style among depressed patients. *Journal of Abnormal Psychology, 91*(2), 102–108. Retrieved from 10.1037/0021-843X.91.2.102%5Cn<http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1982-21321-001&site=ehost-live>
- Rigby, L., & Waite, S. (2007). Group therapy for self-esteem, using creative approaches and metaphor as clinical tools. *Behavioural and Cognitive Psychotherapy, 35*(3), 361–364. <https://doi.org/10.1017/S1352465806003389>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioural research*. New York: W. H. Freeman.
- Seligman, M. E. P., Abramson, L. Y., Semmel, A., & Von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology, 88*(3), 242–247. <https://doi.org/10.1037//0021-843X.88.3.242>

- Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*(1), 148–154.
<https://doi.org/10.1037/0033-2909.106.1.148>
- Soares, J. J. F., & Grossi, G. (2000). The relationship between levels of self-esteem, clinical variables, anxiety/depression and coping among patients with musculoskeletal pain. *Scandinavian Journal of Occupational Therapy*, *7*, 87–95. <https://doi.org/10.1080/110381200750018887>
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Swann, W. B. (1987). Identity negotiation: where two roads meet. *Journal of Personality and Social Psychology*, *53*(6), 1038–1051.
<https://doi.org/10.1037/0022-3514.53.6.1038>
- Tennen, H., Herzberger, S. (1987). Depression, self-esteem, and the absence of self-protective attributional biases. *Journal of Personality and Social Psychology*, *52*(1), 72–80. <https://doi.org/10.1037/0022-3514.52.1.72>
- Tennen, H., Herzberger, S., & Nelson, H. F. (1987). Depressive attributional style: The role of self-esteem. *Journal of Personality*, *55*(4), 631–660.
<https://doi.org/http://dx.doi.org/10.1111/j.1467-6494.1987.tb00456.x>
- Waite, P., McManus, F., & Shafran, R. (2012). Cognitive behaviour therapy for low self-esteem: A preliminary randomized controlled trial in a primary care setting.

Journal of Behavior Therapy and Experimental Psychiatry, 43(4), 1049–1057.

<https://doi.org/10.1016/j.jbtep.2012.04.006>

Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology*, 83(1), 185–197. <https://doi.org/10.1037//0022-3514.83.1.185>

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27(4), 552–565. <https://doi.org/10.1037/0012-1649.27.4.552>

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema Therapy: A Practitioner's Guide*. New York, NY, US: Guilford Press.

Part 3: Critical Appraisal

Critical Appraisal

Introduction

This appraisal includes a reflection on the process of undertaking the literature review and the empirical study of the doctoral thesis. A systematic review of domain specific self-esteem measures for adults was conducted as part of the literature review. The empirical study investigated the effectiveness of a domain specific self-esteem group intervention based on a model developed by Hollingdale (2015).

The critical appraisal begins with a discussion about definitional and measurement issues faced when performing the literature review. The commentary about the empirical study includes reflections on the background, theoretical issues, measurement issues, group experience and challenges faced. It concludes with a reflection on the clinical implications of implementing the domain specific self-esteem group intervention within the National Health Service (NHS).

Reflections on the Literature Review

Definition of Self-esteem

It became clear when I began work on my literature review that there was considerable debate around the definition of self-esteem (Byrne, 1996). What had originally started as a straightforward search quickly became a complex one as I attempted to familiarise myself with the various definitions surrounding the construct. For example, Shavelson, Hubner and Stanton (1976) had found 17 different conceptual dimensions of self-esteem. With the conflation of other constructs such as self-concept and self-efficacy, terms were often used interchangeably in the literature (Byrne, 1996).

I was fortunate to be able to refer to past literature reviews (e.g. Byrne, 1996; Shavelson et al., 1976) to understand better how they navigated through this

definitional maze. This was extremely helpful in deciding on a definition in order to search for related measurement instruments. Whilst this took considerable time and effort, it eventually gave me clarity as I proceeded to search for measurement instruments. This also enabled a smoother process of deciding on the appropriate measures to include and evaluate in the review.

The use of a measurement checklist

The COSMIN checklist was helpful in the process of evaluating the studies and the instruments. The COSMIN checklist was created to enable evidenced based instrument selection (Mokkink et al., 2010). It provided a clear rationale for ratings and simplified the complex process of comparing study methodology. I found the process of determining the quality of a study's methodology clear and straightforward to follow. This was especially helpful when the examined studies had numerous methodological differences.

Moreover, the quality of ratings was stringent which ensured that ratings for each psychometric component were valid. The method score was obtained by the lowest score rating ('worse score counts'; Mokkink et al., 2010). For example, if one item in the box 'Reliability' was scored poor, the methodological quality of that reliability study was rated as poor (Mokkink et al., 2010). This provided assurance in the methodological quality ratings of the measures examined. It also standardised the results in the study which can be used in the future to compare with other studies that utilised the same checklist.

However, given the complex nature of measuring the multidimensionality of self-esteem, deciding to use one instrument over another does not solely depend on the overall instrument ratings. Rather, researchers must be clear about the hypotheses

they are hoping to test and weigh the various advantages and disadvantages of the instruments suited for that purpose.

Reflections on the Research Paper

Background

Interest in clinical interventions as part of research

I was eager early on to undertake this research project as I had an interest in running a clinical intervention as part of my research. This stemmed from a hope to experience and understand the entire research process of piloting, running and evaluating a clinical intervention. In addition, testing a novel model and intervention that was developed by a senior trainee on our course intrigued me. We were given the opportunity to take Hollingdale's (2015) ideas forward to investigate in a clinical pilot study, which was meaningful and fulfilling. Being given the opportunity to collaborate with Hollingdale (2015) who conceptualised the model was helpful. This process enabled me to have a clear theoretical understanding of the intervention. While piloting and evaluating any intervention is a lengthy process, it was satisfying to be able to take a fellow trainee's ideas forward. My hope is that these ideas will continue to be developed in future research projects.

It was also fulfilling to facilitate the groups, as there were many direct opportunities to collaborate with participants to work towards their self-esteem goals. This made the research process meaningful, even during times when my co-facilitator and I felt exhausted. Recognising that the intervention might be beneficial to the participants kept us persevering through. Moreover, receiving positive feedback from some participants in the group was encouraging for me and helped to make the process a rewarding one.

Group processes contributing to recovery

It was an insightful experience to conduct a group intervention as part of the research project. Drawing ideas from Yalom and Leszcz's (2005) "theory and practice of group psychotherapy" was helpful in thinking about the process. Yalom and Leszcz (2005) identified 11 "therapeutic factors" in group therapy that influence the process of change and recovery in group therapies. This section will examine some of these factors that were relevant to the domain specific self-esteem group intervention.

Firstly, Yalom and Leszcz (2005) indicated universality as a factor in change and recovery. He suggested that most clients enter therapy feeling alone in their distress, but when they hear that others face similar problems and experiences, they begin to feel less alone, which aids recovery (Yalom & Leszcz, 2005). This was pertinent to the domain specific self-esteem group as participants had commented on the supportive group environment during the group sessions. Meeting new people and hearing their perspectives had helped them feel more understood. They referred to the group as a safe environment to share their experiences without being judged. Moreover, as participants were from the same university, they identified with the stresses and expectations placed upon them as students. This also relates with what Yalom and Leszcz (2005) called cohesiveness, a sense of "groupness" of being accepted and valued by the group; this satisfies one's need to belong (Yalom & Leszcz, 2005).

Secondly, altruism also appeared to be a factor for self-esteem improvements observed in the groups. Yalom and Leszcz (2005) explained that individuals who enter treatment often hold the belief that they have nothing to offer others. However, in a group setting, individuals learn that they are capable of helping others, resulting

in an increase in their self-esteem. I observed that many group members volunteered to share their experiences with others. For example, some participants shared their group therapy “blueprint” which included ways that they managed setbacks. The group setting had provided participants with a space to help others, which Yalom and Leszcz (2005) suggested might thereby build one’s self-esteem and promote recovery.

Finally, Yalom and Leszcz (2005) suggested that through the instillation of hope by seeing other group members get better, individuals also start to believe that they can get better. Every group session began with a discussion on the process of doing the “homework” from the previous session. The “homework” included positive strengths logs, thought diaries and behavioural experiments. During these group discussions, some participants would share positive experiences and lessons gained from these exercises, which might have instilled hope in others that change and recovery is possible. This might have further motivated others in the group to implement and engage in their own behavioural experiments.

Theoretical and Conceptual Discussions

Third Wave Cognitive Behavioural Therapy (CBT) approaches

The unsatisfactory self-esteem model was primarily based on CBT theory and framework, which drew on Fennell’s (1997) work on self-esteem. A key concept of the unsatisfactory self-esteem model is the value or importance that individuals place in particular domains in their life. During the second session of the group intervention, participants had the opportunity to obtain their own domain specific self-esteem profile and identify their valued domains. This became the basis for setting meaningful intervention goals to increase self-esteem in those domains.

On reflection, this concept possibly mirrors recent “third-wave” CBT approaches such as Acceptance and Commitment Therapy (ACT). In ACT, values are one of the key components of the model (Hayes, Strosahl, & Wilson, 2012). Values are defined as ‘qualities of action’ (Hayes et al., 2012). These might encompass values around domains such as work, relationships, or leisure. In ACT, one objective is to help individuals identify values and subsequently allow their values to exert an influence on their behaviours. Individuals commit to set goals that are in line with the values, which thereby brings purpose and meaning to them (Hayes et al., 2012). Similarly, valued domains were identified through the domain specific self-esteem profiles in the domain specific self-esteem group. The further exploration of goals that are in line with one’s values might be interesting to consider during the group intervention.

Some participants had struggled to ‘challenge’ firmly held beliefs around these valued domains in behavioural experiments. They believed that developing behavioural experiments to challenge these beliefs would not change how they thought about themselves and were therefore less motivated to do them. On reflection, the ACT principle of approaching these difficulties through acceptance might be advantageous. ACT suggests that individuals are often engaged in a relentless struggle to directly change, challenge or eliminate distressing thoughts which is referred to as experiential avoidance (Hayes et al., 2012). Acceptance refers to a proactive willingness to experience distressing feelings and thoughts without actively trying to get rid of them (Hayes et al., 2012; Luoma, Hayes & Walser, 2007). It might therefore be interesting to consider integrating these approaches in the unsatisfactory self-esteem model and examine whether these play a role in improving domain specific self-esteem.

Building on clients' strengths

The unsatisfactory self-esteem model aims better to conceptualise self-esteem through a multidimensional understanding of the construct. This focuses treatment on domains of self-esteem that are “unsatisfactory” for individuals’ at that particular point in time. However, the intervention simultaneously strives to enable participants to think holistically about their self-esteem and reflect on the domains they have developed “satisfactory” self-esteem in.

One cognitive distortion in CBT is discounting or disqualifying the positive, which refers to dismissing good things that one has done for some reason or another (Beck, 2011). Therefore, this was tackled in the group homework to help participants search for areas of strengths. Participants were encouraged to think of their positive qualities or speak to loved ones who would share these with them, in order to bring these qualities to the participant’s awareness. In the second session, participants were also tasked to reflect on how the self-esteem model could be relevant to domains that they have developed “satisfactory” self-esteem in. This included considering areas of strengths and skills participants developed to maintain “satisfactory” self-esteem in these domains. I believe that this process was helpful in encouraging participants to think holistically about their self-esteem in the various life domains.

Reflection of group experience

Understanding of CBT

As part of the introduction to the unsatisfactory self-esteem model, we had asked participants during the group sessions if they had previous knowledge of CBT. It was surprising that many participants had heard about CBT through lectures,

books or the internet; some were sufficiently knowledgeable to explain the relationship in CBT between one's thoughts, feelings and behaviours. This was helpful as it started meaningful conversations amongst group members around what CBT was and it also facilitated peer learning.

However, on reflection, one's knowledge about CBT might vary in a typical clinical setting. Therefore, while the group material introduces the model thoroughly, additional time to elaborate on it and answer queries might be necessary in other settings.

Joint Working

I thoroughly enjoyed the process of working jointly on a research project. Emily and I were able to "bounce off" ideas with each other, which thereby facilitated a fruitful brainstorming process. Planning and executing a group intervention was harder than imagined. Administrative and logistical tasks were more complicated than originally thought; this included tasks such as putting up posters up around the university campus, printing materials, preparing the presentation slides, booking group rooms and emailing participants. Sharing the workload made the whole process much less burdensome.

Joint working also provided a space for us to reflect about our experiences. This facilitated open and honest conversations about how the research process was for each of us. This was helpful for our emotional wellbeing, considering that we had to juggle various demands as part of our doctoral training.

It was also enjoyable running the groups with a co-facilitator. Facilitating a group together allowed us to better cater to the needs of the group members. Behavioural experiments come to mind. Planning a meaningful experiment for each participant took time and effort; we were able to do that effectively because two

facilitators were available. We spent five to ten minutes with each participant to plan behavioural experiments that were specific, meaningful and feasible to do. This would not have been conceivable without a co-facilitator. Moreover, we were able to pick up on one another's 'blind spots' during the facilitation of the group and answer difficult questions posed to us by group members. All of this made the entire experience of facilitating the groups an enjoyable and fulfilling process.

Challenges faced

Difficulties faced when planning behavioural experiments

Many participants were successful in implementing their planned behavioural experiments. However, there were a few participants who struggled to do so. Possible reasons are discussed below. Firstly, engaging some participants in behavioural experiments was difficult because some of them held beliefs that their negative assumptions of themselves were "facts" rather than opinions. It was therefore difficult to suggest to some participants to test these "facts" about themselves. Therefore, it might be helpful to take more time with these individuals to explore the possibility of viewing these "facts" as simply opinions through Socratic questioning. This would thereby provide a framework later on to test these opinions in behavioural experiments.

Secondly, a safe context for taking risks to try out new ways of thinking and behaving are important elements of behavioural experiments (Bennett–Levy et al., 2004). There might be some participants who felt that the group was not a safe place to "challenge" previously held beliefs. One example was a participant who had given feedback that they had difficulty disclosing to the other group members in the table. This might have resulted in a lack of openness to the behavioural experiment component of the intervention.

Dropout rates

The empirical paper indicated that the dropout rate was 23.5% from pre-intervention to post-intervention and 38.5% from post-intervention to the one-month follow up. While participants did not mention any group-pertaining reasons for dropping out, some hypotheses can be considered. While Yalom and Leszcz (2005) indicated that cohesiveness was a factor in group therapy that influenced recovery, a sense of “not belonging” to the group might have resulted in participant dropout. Although, as previously indicated, they belonged to the same university, participants in the groups comprised of undergraduate and post-graduate students with diverse backgrounds, ethnicities and cultures. Whilst there was a short ice-breaker exercise at the beginning of the group, some participants might have required more time to integrate within the group setting to feel comfortable.

Also, the exploration of possible painful experiences through the unsatisfactory self-esteem model might have created psychological distress in participants. Based on the domain specific self-esteem profiles, participants were given the opportunity to reflect on the ways they had developed unsatisfactory self-esteem through a longitudinal formulation. This process might have exposed some participants to emotional vulnerability, resulting from the exposure to painful early experiences that might have surfaced in the process. Further research is necessary to determine if these are valid reasons accounting for dropouts. If so, additional support should be provided to participants to increase a sense of belonging to the group and also a safe space where difficult emotions are managed and contained.

Response burden

A number of participants indicated that filling in the outcome measures was tedious, which might have resulted in participant response burden. While most participants completed the questionnaires in 20 minutes, some required considerably more time to do so. This might have also resulted in participant fatigue or the loss of engagement during the group sessions.

It is therefore important in the future to ensure a balance between capturing participant outcomes and reducing participant fatigue and response burden. It might be beneficial to pilot the questionnaires with a focus group to get feedback on the ease of completion of the measures. Prioritising measures will also help researchers decide which measures to include or exclude.

Measurement Issues

Choice of domain specific self-esteem measure

Based on the systematic review that was conducted in the literature review, a wide array of domain specific self-esteem measures were available from which to choose from. The Self-Perception Profile for College Students (SPP-CS; Neemann & Harter, 2012) was chosen based on psychometric and theoretical considerations. While the psychometric considerations were discussed in the empirical paper, this section further elaborates on the theoretical considerations which led to the decision of using the SPP-CS.

The theoretical stance of the measure chosen had to be aligned to the unsatisfactory self-esteem model in three areas. Firstly, the measure chosen had to capture the multidimensionality of self-esteem through various life domains (e.g. relationships, physical appearance etc.). Measures that examined only one domain (e.g. physical self-esteem) were not considered. Moreover, the measure had to capture an adequate range of different life domains. The SPP-CS had 12 domain

subscales that fit with this conceptualisation. The domains included creativity, intellectual ability, scholastic competence, job competence, athletic competence, appearance, romantic relationships, social acceptance, close friendships, parent relationships, humour and morality (Neemann & Harter, 2012). Conversely, an example of a multidimensional self-esteem measure that was not included due to its more abstract domain categories was the Six-Factor Self-Concept Scale (SFSCS; Stake, 1994). The SFSCS aimed to provide maximum generalisability across situations (e.g. work, relationships). Subscales therefore included more abstract categories such as Likability, Morality, Task Accomplishment, Giftedness, Power and Vulnerability (Stake, 1994). These did not fit with the conceptualisation of domains consistent with the unsatisfactory self-esteem model. Therefore, measures such as the SFSCS were excluded based on this.

Secondly, the unsatisfactory self-esteem model postulates that a key concept is the value or importance attached to a domain (Hollingdale, 2015). The importance placed in domains would determine if threats to self-esteem are perceived in those domains; only threats to domains of importance were hypothesised to affect self-esteem. Therefore, the domain specific self-esteem measure chosen had to incorporate importance ratings of the various life domains. For this purpose, the SPP-CS encompasses importance ratings to assess the importance of each domain to the individual.

Thirdly, the unsatisfactory self-esteem model posits that domain specific self-esteem fluctuates over the course of an individual's life. Developmental changes across the lifespan have been found to lead to changes in self-esteem that might be due to the changes in role demands, maturational changes, physical functioning and the individual's socioeconomic status (Orth, Trzesniewski & Robins, 2010).

Therefore, the chosen measure had to capture age appropriate domains for the population examined. As we were piloting the group intervention with university students, the measure chosen had to be appropriate for the study sample. The SPP-CS satisfied this by measuring self-esteem in domains that are relevant to a university sample.

Overall, the SPP-CS is not a perfect measure and has its limitations. However, it adequately met most of the theoretical considerations outlined above. It was therefore chosen as the domain specific self-esteem measure for the empirical study.

Valued domains identified by participants

Although most of the domains identified by the SPP-CS were relevant to the university sample in the group intervention, some participants had given feedback that they had valued domains that were not on the list. One example was family relationships. While parent relationships were included, some participants believed that their relationships with siblings and grandparents were important domains to consider. Other domains such as faith and spirituality that might have been relevant to an individual's self-esteem were also not included in the SPP-CS. It would be noteworthy for further research to explore the validity of these domains when examining domain specific self-esteem.

Clinical Implications: Improving Access to Psychological Therapies (IAPT)

Groups

Finally, implementing the domain specific self-esteem group within existing mental health service frameworks is important to consider. One possible consideration would be to incorporate the group in IAPT services alongside other evidenced based treatments. Since the inception of IAPT services, the demand for

mental health treatments has been increasing. Over 900,000 people access IAPT services each year (Clark, 2018). Therefore, delivering adequate and high-quality care in meeting patients' mental health needs is a top priority.

IAPT service provision is based on a stepped care framework (National Institute for Health and Care Excellence [NICE], 2011). The stepped-care framework is a model where each step represents an increased intensity of intervention and is used to organise the provision of services to help people find the most effective treatments (NICE, 2011). At these steps, IAPT services provide low-intensity (LI) and high-intensity (HI) psychological assessment and therapy for clients with depression and anxiety difficulties (NICE, 2011). Based on the NICE guidelines, CBT groups are typically situated at both steps 2 and 3 in the framework (NICE, 2011).

The unsatisfactory self-esteem group could be situated within this framework at step 2. LI therapists would facilitate and run the groups under supervision from a HI therapist for individuals struggling with unsatisfactory self-esteem. As a single pathway transdiagnostic approach, the group might cater for individuals with a broad range of psychiatric disorders such as depression and anxiety. This might be an appropriate pathway for individuals struggling with self-esteem and other mental health difficulties. Dixon (2018) found that the domain specific self-esteem group intervention also had a positive impact on depression and general wellbeing. Feasibility and acceptability studies in the IAPT setting could be conducted in the future to determine its utility. It would be beneficial to also investigate how a single pathway group intervention might complement evidenced based CBT therapies for specific disorders in mental health services.

Conclusion

This critical appraisal encapsulates my reflections around conducting the systematic review and the major research project as part of my Doctorate in Clinical Psychology. Through this appraisal, I reflected upon theoretical, measurement and other contextual issues of the thesis. I hope that highlighting some of these issues may be beneficial to others who seek to conduct research in the area of domain specific self-esteem.

Finally, I would highly recommend the opportunity to anyone interested in running a clinical intervention as part of their research project. It was a tremendously fulfilling experience to work directly with participants and to see the positive impact the group intervention had on some of them. Moreover, doing a joint project with another trainee made this research experience, which could have been highly stressful, an invaluable and enjoyable one.

References

- Beck, J. S. (2011). *Cognitive Behavior Therapy: Basics and Beyond*. The Guilford Press (2nd ed.). New York, NY, US: Guilford Press.
<https://doi.org/10.1017/CBO9781107415324.004>
- Bennett-Levy, J., Butler, G., Fennell, M., Hackman, A., Mueller, M., & Westbrook, D. (2004). *Oxford guide to behavioural experiments in cognitive therapy*. New York, NY, US: Oxford University Press.
<https://doi.org/http://dx.doi.org/10.1093/med:psych/9780198529163.001.0001>
- Byrne, B. M. (1996). *Measuring self-concept across the lifespan*. Washington, DC: American Psychological Association.
- Clark, D. M. (2018). A guide to Improving Access to Psychological Therapies services. Retrieved June 17, 2018, from <https://www.england.nhs.uk/blog/a-guide-to-improving-access-to-psychological-therapies-services/>
- Dixon, E. (2018). *A feasibility study of a Cognitive Behavioural Therapy group for domain-specific self-esteem and its impact on global self-esteem, depression, anxiety and psychological wellbeing*. (Unpublished doctoral thesis). University College London, London.
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change*. *Acceptance and commitment therapy: The process and practice of mindful change (2nd ed.)*. (2nd ed.). New York, NY, US: Guilford Press. Retrieved from

<http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc9&NEW S=N&AN=2012-00755-000>

Hollingdale, J. (2015). CBT for self-esteem: An Introduction to a new model.

Unpublished presentation given at University College London. University College London, London.

Luoma, J. B., Hayes, S. C., & Walser, R. D. (2007). *Learning ACT: An acceptance and commitment therapy skills-training manual for therapists*. Oakland, CA, US: New Harbinger Publications.

<https://doi.org/Doi:10.1017/S1352465808004670>

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D.

L., ... de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*(1), 22.

<https://doi.org/10.1186/1471-2288-10-22>

National Institute for Health and Care Excellence. (2011). Common mental health problems: identification and pathways to care. Retrieved from

<https://www.nice.org.uk/guidance/cg123/chapter/1-guidance>

Neemann, J., & Harter, S. (2012). *Self-Perception Profile for College Students:*

Manual and questionnaires. Unpublished manuscript, University of Denver, Colorado, USA.

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-Esteem development

From young adulthood to old Age: A cohort-sequential longitudinal study.

Journal of Personality and Social Psychology, 98(4), 645–658.

<https://doi.org/10.1037/a0018769>

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441.

<https://doi.org/10.3102/00346543046003407>

Stake, J. E. (1994). Development and validation of the Six-Factor Self-Concept Scale for Adults. *Educational and Psychological Measurement*, 54(1), 56–72.

Retrieved from

<http://journals.sagepub.com.libproxy.ucl.ac.uk/doi/pdf/10.1177/0013164494054001006>

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy* (5th ed.). New York, NY, US: Basic Books.

Appendices

Copyrighted material and material that might affect validity if freely available have
been removed

Appendix A

Researchers' contributions to the joint project

This project was a joint project with Emily Dixon. Ciping Goh's study examined the effectiveness of the domain specific self-esteem group in relation to changes in domain specific self-esteem in valued domains, discrepancy scores between perceived competence and importance, and attributional styles towards negative and positive events. In addition, Ciping Goh examined the relationship between domain specific self-esteem and attributional styles. Emily Dixon explored the relationship between domain-specific and global self-esteem and identified changes in global self-esteem. Her study also sought to explore whether the intervention would lead to improvements on scores of anxiety, depression and psychological wellbeing. In addition, participants' commented on their experience of the group through a feedback questionnaire. Both Emily Dixon and Ciping Goh separately identified themes from the qualitative data in the feedback questionnaire, which Emily Dixon subsequently compiled in the write-up.

The writing of the ethics amendment document, information sheets, and the guided mental imagery script and recording, were compiled jointly. All the group sessions were jointly run by both researchers, with each researcher taking different portions of the session. All practical tasks however were divided equally between the two researchers. For example, Ciping took the role in setting up the online questionnaires on Qualtrics, whilst Emily took the role of consolidating group numbers and emailing participants on the group sessions. The questionnaire data were consolidated and jointly coded by both researchers. All data analysis and write-up were conducted separately.

Appendix B

Ethics Approval Letter

UCL RESEARCH ETHICS COMMITTEE
ACADEMIC SERVICES



30th November 2016

Dr Henry Clements
UCL Research Department of Clinical, Educational and Health Psychology

Dear Dr Clements

Notification of Ethical Approval

Re: Ethics Application 9659/001: The effectiveness of a domain-specific self-esteem group

I am pleased to confirm in my capacity as Chair of the UCL Research Ethics Committee (REC) that your study has been ethically approved by the REC until **30th September 2018**.

Approval is subject to the following conditions:

Notification of Amendments to the Research

You must seek Chair's approval for proposed amendments (to include extensions to the duration of the project) to the research for which this approval has been given. Ethical approval is specific to this project and must not be treated as applicable to research of a similar nature. Each research project is reviewed separately and if there are significant changes to the research protocol you should seek confirmation of continued ethical approval by completing the 'Amendment Approval Request Form':

<http://ethics.grad.ucl.ac.uk/responsibilities.php>

Adverse Event Reporting – Serious and Non-Serious

It is your responsibility to report to the Committee any unanticipated problems or adverse events involving risks to participants or others. The Ethics Committee should be notified of all serious adverse events via the Ethics Committee Administrator (ethics@ucl.ac.uk) immediately the incident occurs. Where the adverse incident is unexpected and serious, the Chair or Vice-Chair will decide whether the study should be terminated pending the opinion of an independent expert. For non-serious adverse events the Chair or Vice-Chair of the Ethics Committee should again be notified via the Ethics Committee Administrator within ten days of the incident occurring and provide a full written report that should include any amendments to the participant information sheet and study protocol. The Chair or Vice-Chair will confirm that the incident is non-serious and report to the Committee at the next meeting. The final view of the Committee will be communicated to you.

Final Report

At the end of the data collection element of your research we ask that you submit a very brief report (1-2 paragraphs will suffice) which includes in particular issues relating to the ethical implications of the research i.e. issues obtaining consent, participants withdrawing from the research, confidentiality, protection of participants from physical and mental harm etc.

Yours sincerely



Professor John Foreman
Chair, UCL Research Ethics Committee

Cc: Emily Dixon & Ciping Goh

Academic Services, 1-19 Torrington Place (9th Floor),
University College London
Tel: +44 (0)20 3108 8216
Email: ethics@ucl.ac.uk
<http://ethics.grad.ucl.ac.uk/>

Appendix C

Study advertisement

Do you struggle with your Self-Esteem?

Do low levels of self-esteem sometimes make you feel anxious or depressed?

Are you a student at UCL?

If so, perhaps we can help. We are running a research study into the effectiveness of an **exciting new intervention for improving self-esteem**.

The intervention will consist of four, 2 hour group sessions that will run weekly at UCL, followed by a final 2 hour group, one month later. You will also be required to complete some short questionnaires on four separate occasions.

Course Credits are available for participation in this study.

The group will be run by Emily Dixon & Ciping Goh (Trainee Clinical Psychologists)

Please feel free to contact us on selfesteem2017@gmail.com with any questions.

The project will be supervised by Dr Henry Clements (Clinical Psychologist) henry.clements@ucl.ac.uk



If you are interested in participating in our study or would like to find out more please take a slip and go to the weblink provided.

ALL DATA WILL BE COLLECTED AND STORED IN ACCORDANCE WITH THE DATA PROTECTION ACT 1998



- goo.gl/MnfmYp

Appendix D

Participant information sheet



RESEARCH DEPARTMENT OF CLINICAL, EDUCATIONAL AND HEALTH PSYCHOLOGY

PARTICIPANT INFORMATION SHEET

Study Title: Study of a Domain-Specific Self-Esteem group

**This study has been approved by the UCL Research Ethics Committee (Project
ID Number):**

You are being invited to take part in a research study. Before you decide whether you would like to take part, it is important for you to know what the research is about and what it will involve. Please read this information sheet carefully and discuss with others if you wish. If there is anything that is not clear, or if you would like more information, you can contact us. Your participation in this study is completely voluntary and you may choose to withdraw at any time.

What is this study about?

This study forms part of University College London Doctorate of Clinical Psychology research theses by Emily Dixon (Trainee Clinical Psychologist) and Ciping Goh (Trainee Clinical Psychologist), and is supervised by Dr Henry Clements, Dr Sue Watson and Dr Sunjeev Kamboj.

The study aims to investigate the effectiveness of a group programme for people experiencing self-esteem difficulties. Currently, the majority of literature on self-esteem views it as a global evaluation of oneself (e.g. confidence in and respect for one's own worth or abilities). However, we believe that self-esteem is domain-specific, that is, it can vary within circumscribed domains. Thus, a person might experience self-esteem deficits in a particular domain(s) (e.g. appearance, academic achievement etc.) but not in others.

Additionally, we believe that self-esteem is on a spectrum and at times can become “unsatisfactory” for a person's needs, within specific domains or within a specific

time period. For example, a university student may value academic achievement highly, and perceived threats to this (e.g. failing an exam), will subsequently violate the individual's self-esteem in this area and so become unsatisfactory for that individual.

The study is a small scale study and we want to establish whether the group has any effect on self-esteem and also how it may be improved in the future to help people with self-esteem issues.

What happens in the group?

In the group, you will have the opportunity to explore your own valued domains, create your individualised domain-specific self-esteem chart and explore why you may have developed unsatisfactory self-esteem in some of these domains. Subsequently, you will plan individualised activities to engage in, with a view to develop a more satisfactory self-esteem in those domains.

Groups will consist of four, two hour sessions on a weekly basis, with a fifth follow-up session one-month later. The groups will be facilitated by ourselves, Emily Dixon and Ciping Goh. There will be approximately 10-12 people in each group.

During the sessions we will ask you to undertake a variety of activities, some of which you will also do between sessions: these may include, tracking your levels of self-esteem in domains important to you; keeping a thought diary; and planning experiments to test the validity of some of your thoughts.

Why have I been invited to take part?

This study is an open invitation to UCL students who would like to explore and work on self-esteem issues.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to give consent after reading through this information sheet.

If you decide to take part, you are still free to withdraw at any time without giving a reason.

What will happen if I take part?

If you are happy to take part in this study and have given consent, you will be asked to complete some online questionnaires regarding your self-esteem, and any possible depression and anxiety symptoms.

This will determine your eligibility for the study. If you are eligible, you will be required to do the following:

- Provide some demographic information (e.g. name, what you are studying etc.) and indicate the dates you are available to attend the group.
- Attend four weekly group sessions and one follow-up session (one month after the group ends) (each 2 hours long)
- Complete questionnaires that will be administered in the first and last session of the group and at follow-up. The questionnaires will include measures of global and domain-specific self-esteem, depression symptoms, anxiety symptoms and attributional style.

What will I be asked to do?

We ask that you attend all five group sessions as far as possible. You will then give yourself the opportunity to gain maximum benefit from the sessions.

You can carry on your everyday activities as normal while participating in the study.

Are there any risks in taking part?

Overall the risks of taking part in this study are minimal. The researchers conducting the group sessions have experience of working with adults with self-esteem issues in clinical settings. In addition, they will be working under supervision from qualified clinical psychologists. In the sessions, you will be encouraged but never forced to take part in any activity. However if being involved in this research really does not suit you, for example, should you find it distressing, you are free to withdraw at any point. We will also signpost you to other services if you need further support.

What are the potential benefits?

If you decide to participate in the study, we hope that you will find the sessions interesting, helpful and enjoyable.

The information gathered during this study will also help to inform our understanding of treatment for domain-specific self-esteem. We anticipate that this will be a step towards improving interventions for self-esteem difficulties in the future.

Will my taking part in the study be kept confidential?

All information collected about you over the course of the study will be kept confidential unless we became aware of something which makes us worry about you or someone around you, in which case we will discuss the issue with you. Once the study has finished, University College London (UCL) will keep the study data in a secure location. The data used for the study will be anonymised and it will not be possible to trace the results back to individual participants.

Your personal data given on this online platform is being handled by Qualtrics. Please refer to the following weblinks for the security and privacy statements.

<https://www.qualtrics.com/security-statement/>

<https://www.qualtrics.com/privacy-statement/>

What happens when the research study stops?

The results of the research study will be written up as part of Emily Dixon's and Ciping Goh's theses for the Clinical Psychology Doctorate at UCL. The report of the study could also be published in relevant journals outside UCL. You will not be identifiable from these results.

What if something goes wrong?

Every care will be taken in the course of this study to protect you. Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. You should contact Dr Henry Clements, who is the Chief Investigator for the research, and based at UCL.

Who is organising and funding the research?

The research has been organised by Emily Dixon and Ciping Goh, Trainee Clinical Psychologists. They are conducting this study as part of their Clinical Psychology Doctorates. The research will be funded by UCL.

Who can I contact for further information?

For more information about this research, please contact:

Emily Dixon and Ciping Goh

Research Department of Clinical, Educational and Health Psychology

UCL

Gower Street

WC1E 6BT

Email: emily.dixon.14@ucl.ac.uk; ciping.goh.15@ucl.ac.uk

Phone: TBC (we are waiting for phones specifically for the project)

Or if you have any concerns or complaints about this study please contact:

Dr Henry Clements

Research Department of Clinical, Educational and Health Psychology

University College London

Gower Street
London WC1E 6BT
Email: henry.clements@ucl.ac.uk
Phone: 07877 127863

**ALL DATA WILL BE COLLECTED AND STORED IN ACCORDANCE
WITH THE DATA PROTECTION ACT 1998.**

**THANK YOU FOR READING THIS INFORMATION SHEET AND FOR
CONSIDERING TAKING PART IN THIS RESEARCH.**

Appendix E

Participant consent form



**RESEARCH DEPARTMENT OF CLINICAL, EDUCATIONAL AND
HEALTH PSYCHOLOGY**

PARTICIPANT CONSENT FORM

Study Title: Study of a Domain-Specific Self-Esteem group

Name of Researchers: Emily Dixon and Ciping Goh

Please tick boxes

I confirm that I have read and understand the information sheet dated [insert date, insert version] for the above study, have had the opportunity to ask questions and have had these answered acceptably.	
I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.	
I understand that the information that I provide will be included in the researchers' doctoral thesis, may be published in a scientific journal, and may be presented at a national or	

international conference. I understand that all information included will be anonymised to protect my identity.	
I understand that all information given by me or about me will be treated as confidential by the research team. Such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.	
I agree to take part in the above study.	

By clicking the >> button below, I give consent to participate in the study.

Appendix F

Group session plan

Material removed

Appendix G

Domain Specific Self-Esteem Group Presentation Slides

Material removed

Appendix H

Self-Perception Profile for College Students (SPP-CS; Neemann & Harter, 2012)

Material removed

Appendix I
Attributional Styles Questionnaire

Material removed

Appendix J

AIC statistics for each mixed model analysis

A summary of the AIC statistics for each mixed model analysis, comparing the use of compound symmetry (CS) and unstructured matrix (UN). A lower AIC statistic represents a better model (Howell, 2015).

	Compound Symmetry (CS)	First order autoregressive model (AR1)
Perceived Competence Scores	-136.242 ^a	-135.229
Discrepancy Scores	146.596 ^a	150.113
ASQ CoNeg	451.775 ^a	457.474
ASQ CoPos	458.005	456.289 ^a

^a model used in the final analysis due to best model fit.

Appendix K

SPSS syntax for Howell's (2008) mixed-model analyses

MIXED PerceivedCompetence BY time

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)

SINGULAR(0.000000000001) HCONVERGE(0,

ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001,

ABSOLUTE)

/FIXED=time | SSTYPE(3)

/METHOD=REML

/PRINT=SOLUTION TESTCOV

/REPEATED=time | SUBJECT(ID) COVTYPE(AR1)

/EMMEANS=TABLES(time) COMPARE ADJ(BONFERRONI).