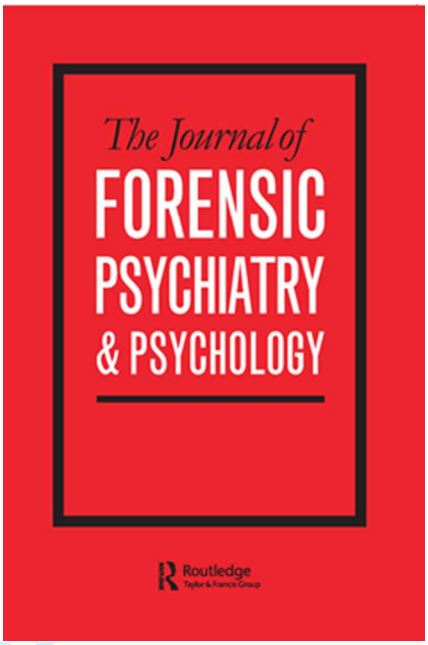

A systematic review on the relationship between self-esteem and interrogative suggestibility

Journal:	*Journal of Forensic Psychiatry and Psychology*
Manuscript ID	RJFP-2015-0011.R2
Manuscript Type:	Regular Article
Keywords:	criminal procedure, legal aspects of psychiatry, systematic review, risk factors

ABSTRACT

Some factors, such as age, learning disability and mental health difficulties, have been identified as making police suspects more vulnerable to suggestibility and false confessions during interview. However, there has been no systematic review on the association between self-esteem and suggestibility.

Seven electronic bibliographic databases and reference lists of previous literature reviews of suggestibility in children were searched. Selected studies were quality assessed using pre-defined criteria before data were extracted.

Electronic searches yielded 1914 hits. Of these, 685 duplicates, 1181 irrelevant references and 39 references that did not meet the inclusion criteria were removed. Nine publications were included in the review.

Significant correlations between self-esteem and suggestibility, most notably on the Yield 1 subscale of the GSS, were found but four of the nine studies found no significant correlation. The prevalent use of self-report measures and lack of clarity in defining self-esteem limit the validity of those studies.

1. Introduction

1.1 Defining and measuring interrogative suggestibility

Suggestibility has been defined as "the influence of one person on another without his or her consent, the implanting of an idea, possessing a submissive tendency, and appealing to the unconscious" (Marcuse, 1976, cited in Wagstaff, 1991, p. 132). More recently, this has been divided into two distinct concepts, suggestibility and compliance. Interrogative suggestibility refers to the extent to which an individual comes to accept a message communicated by another person as fact (Gudjonsson & Clark, 1986) and integrates this into their own knowledge and behaviour.

Gudjonsson and Clark (1986) noted three components as prerequisites to the process of interrogative suggestibility: uncertainty, trust (in the interviewer) and expectation (interviewee belief they should know the answer). In contrast, compliance does not require the private acceptance of the message (Gudjonsson, 1997), but rather concerns a conscious decision to carry out the behaviour requested. The concepts are overlapping in that both are prompted in an effort to avoid conflict or confrontation, or in an effort to please the other person.

The most predominantly used tool for measuring interrogative suggestibility remains the Gudjonsson Suggestibility Scales (GSS; Gudjonsson, 1984, 1997). The GSS comprises a narrative containing forty distinct ideas which is of sufficient length that no respondent is able to remember all of the material. This is followed by a series of questions about the story which are read to the respondent by the interviewer. These questions include fifteen suggestive and five "true" questions. Measures include recall (Immediate Recall and Delayed Recall subscales), response to leading questions (Yield 1 and Yield 2 subscales) and response to negative feedback (Shift subscale). A Total Suggestibility score is calculated from the Yield 1 and Shift subscales.

Interrogative suggestibility, as outlined above, should be distinguished from hypnotic suggestibility, as measures of these concepts are not found to correlate significantly (Register & Kihlstrom, 1988).

1.2 The importance of recognizing and managing interrogative suggestibility

Interrogative suggestibility has been of relevance in cases of false confession during police interview. In 2004, Drizin and Leo compared 125 cases of false confession (proven through DNA) in the United States from 1971 to 2002. Of these, they found that 93% were made by males, with 81% of the false confessions occurring within cases of murder. 63% of those who confessed were aged under 25, and 80% of those who confessed falsely and went to trial were convicted of the offence they had admitted to. Realistically, it is difficult to ascertain the actual numbers of false confessions made. Previous research has found percentages from 7% to as high as 28% where false confessions are self-reported by participants (Gudjonsson et al, 2007a, 2007b, 2010, 2012; Redlich, Summers & Hoover, 2010). It must be noted, however, that where false confessions are self-reported they have rarely been backed by definitive evidence that the confession has been false. Equally, these reports often relate to low-level offences. With this in mind, the validity of such statistics should be considered.

Despite the large number of studies within the area of suggestibility, this evidence tends not to be used in practice for reducing false confessions – for example, whilst it has had some effect in shaping police interviewing techniques with eyewitnesses, it has had little effect on the suspect interview. Although the provision of Appropriate Adults for vulnerable detainees was entrenched into the 1984 Police and Criminal Evidence Act (PACE) in an effort to reduce the high incidence of false confessions within this population, there has been limited guidance for police in identifying the characteristics which make a suspect

"vulnerable" for the purposes of interview and which would therefore allow their identification by custody officers at the time of booking in. At the current time in England and Wales, vulnerability is identified in terms of age, learning disability and mental health difficult/illness, and is supported by research by Conley, Luckasson and Bouthilet (1992), Gudjonsson, Clare, Rutter and Pearse (1993) and Redlich (2004). Literature reviews of suggestibility research have, however, indicated a number of other possible factors, and it is possible that important, but more subtle, factors are being missed by custody staff in the identification of vulnerable detainees.

1.3 Self-esteem and interrogative suggestibility

Blascovich and Tomaka (1991) noted that that throughout the history of research on self-esteem, the concept has remained poorly defined and therefore badly measured. Coopersmith (1967) defined self-esteem as "the extent to which an individual believes himself to be capable, significant, successful and worthy" (pp. 4-5), whilst Baumeister (1998) considered it to be the evaluative aspect of the self-concept that corresponds to an overall view as worthy or unworthy. One of the more popular definitions of self-esteem, however, comes from Rosenberg (1965), who described it as a favourable or unfavourable attitude towards the self (p. 15). More recently, Brown and Marshall (2006) suggested that the confusion surrounding the definition of self-esteem is grounded in a lack of agreement regarding the construct itself (p. 4). They highlighted three different uses of the term "self-esteem", to describe global self-esteem, feelings of self-worth, or self-evaluations.

Ziegler-Hill (2014) noted "it is difficult to estimate the prevalence of low (or high) levels of self-esteem in the population because self-esteem is almost always conceptualized as a dimensional construct rather than as discrete categories" (p. 268). With few

conceptualizations of what constitutes 'high' or 'low' self-esteem, there are few estimations of the commonality of self-esteem problems within the general population.

Difficulties in operational definition aside, the development of specific psychometrics focussing on self-esteem have brought with them the potential for a common understanding of this concept and the replication and generalisation of its measurement. It is through the development of these self-esteem scales – and comparison with similar measures of interrogative suggestibility – that the relationship between these two concepts can be studied. What is common between these scales and operational definitions are the two ideas, firstly, that this concept clearly concerns the self, and secondly, that this concept concerns positive and/or negative views.

Scoping revealed a number of studies where self-esteem had been considered as a factor relating to suggestibility (Baxter, Jackson & Bain, 2003; Numoja & Bachmann, 2008; Drake, Bull & Boon, 2008). A significant negative relationship between these two concepts (i.e. indicating that an individual with lower self-esteem may experience increased suggestibility) may have implications for police interviewing procedure. Self-esteem is not a factor currently considered as causing suspects in police interview to be vulnerable to suggestibility and subsequent false confession. As such, interviewees presenting in custody with low self-esteem would not currently be afforded measures to manage this, such as the engagement of an Appropriate Adult to ensure that their rights are upheld and that communication between suspect and police is facilitated effectively.

1.4 Existing reviews and meta-analyses

No previous systematic literature reviews or meta-analyses focusing specifically on the association between self-esteem and suggestibility have been published. Whilst there is an abundance of literature reviews published focusing on factors associated with suggestibility,

none of these have used systematic principles, and rather provide an overview and exploration of previous research.

Several reviews have been conducted into the factors associated with suggestibility in children. Ceci and Bruck (1993) conducted a review of the suggestibility in relation to child witnesses. Findings identified three 'families' of factors in suggestibility: Cognitive, social and biological, and it was suggested that despite age differences in suggestibility, even very young children are able to recall relevant details. Bruck and Melnyk (2004) also explored the individual differences in children's suggestibility. 69 studies were synthesized and divided into demographic factors, cognitive factors and psychosocial factors. The highest correlations for psychosocial factors included self-concept/self-efficacy. Additional reviews have focused on the relationship between intelligence (learning disability) and suggestibility (Kebbell & Hatton, 1999). Drake and Bull (2011) noted "adult interrogative suggestibility has so far received relatively little consideration from psychologists" (p. 677).

1.5 Aims and objectives

This review aims to systematically and comprehensively explore the association between self-esteem and suggestibility in individuals of criminally responsible age in England and Wales (≥10 years) in whom alternative strongly predictive factors of suggestibility (intelligence and mental health issues) do not exist. This review seeks to explore whether a relationship between self-esteem and suggestibility exists, and if so, the nature of such a relationship. The value of self-esteem in predicting suggestibility will also be considered.

2. Method

2.1 Search strategy: Sources of literature

For the identification of primary studies on the association between self-esteem and suggestibility, a number of electronic bibliographic databases were searched, including PsychINFO (1806 – week 3 Dec 2014), MEDLINE (1948 – week 3 Dec 2014), EMBASE (1980 – week 3 Dec 2014), ASSIA (1987 – week 3 Dec 2014), Web of Science (1900 – week 3 Dec 2014), NCJRS Abstracts Database (1975 – week 3 Dec 2014) and EThOS (earliest – week 3 Dec 2014).

Researchers additionally searched the Cochrane and Campbell libraries and PROSPERO for relevant reviews with no results. One meta-analysis and two literature reviews (identified above) were found during scoping, and the references of these were hand-searched for additional relevant publications. Time constraints meant that researchers were unable to make contact with experts in the field.

2.2 Search strategy: Search terms

The following is a guide to the search terms that were used in all databases. These were modified to meet the specific requirements and parameters of each database (available upon request).

suggestibility/compliance/misinformation/cross-examination

AND

self-esteem/self-concept/self-perception/self-confidence

2.3 Study selection

Irrelevant studies retrieved through the searches were identified from their titles and abstracts and removed from the sample. Inclusion and exclusion criteria were then applied to the remaining studies using a pre-defined form (available upon request). Studies were

selected based on their adherence to all of these inclusion criteria. A list of excluded studies and reasons for exclusion is available upon request.

Studies that met the following criteria were included in the review:

Population:	Adults or young people, where the mean age of the sample is 10 years or older.
Exposure/Issue:	Self-esteem measured as below average by psychometric assessment, rated as "low" by researchers, or measured as part of a scale.
Comparator:	Self-esteem measured as above average by psychometric assessment, rated as "high" by researchers, or measured as part of a scale.
Outcome:	Suggestibility measured by psychometric assessment, response to leading/misleading questions or response to misinformation.
Study type:	Cohort, case control or cross-sectional studies
Exclusion:	Studies which focused only on individuals with an identified learning disability, individuals in psychiatric hospitals or with identified mental health issues, or where the mean age of the sample was less than 10 years old. Studies where no measurement of self-esteem or suggestibility was conducted. Studies which considered social conformity, social influence, hypnotic suggestibility or persuadability. Narrative reviews, qualitative studies, editorials, opinion papers, commentaries and book chapters.
Language:	English language only.

The population was limited to individuals above the age of criminal responsibility in England and Wales to enable findings to be applied to potential police suspects and to link with the provision of the Appropriate Adult for vulnerable suspects. Studies where included participants had a mean age of 10 years old were included, provided that the data of participants of the appropriate age (≥10 years) could be separated from those who were too young. Studies which only included participants with a learning disability or with mental health difficulties were excluded, as these factors have been strongly associated with suggestibility (Conley, Luckasson & Bouthilet, 1992; Gudjonsson, Clare, Rutter & Pearse, 1993; Redlich, 2004) and might be considered as mediating variables. No specific standardized assessments of self-esteem or suggestibility were outlined as being necessary for inclusion, as limitations of specific measurements would be taken into account in both the quality assessment and subsequent analysis stages. Studies which measured concepts similar to compliance were excluded, as this is considered a different concept from suggestibility because it does not rely on the internalization of information (Gudjonsson, 1989). This term was included in the search strategy, however, to allow sensitivity to differences in vocabulary and keywords used within studies. Studies exploring hypnotic suggestibility were also excluded as this has been found to differ significantly from interrogative suggestibility (Gudjonsson, 1987a). No limits were set on language during the search stage, but studies could only be included within the final review if they could be sourced in the English language.

2.4 Quality assessment

The quality of each study was assessed using pre-defined criteria (available on request) adapted from the CASP critical appraisal checklists. These checklists assist researchers in examining bias (selection, performance, detection and attrition) in

methodology. Quality criteria allowed researchers to appraise individual bias items as present or absent. Researchers applied structured judgment of the number of quality criteria met and their relative importance to qualify studies as high, reasonable or low quality.

Quality assessment was carried out on all of the studies independently by the researcher and another reviewer, both of whom were engaged in a professional doctoral degree for trainee forensic psychologists. The percentage of agreements between the two reviewers was 97%. An inter-rater reliability analysis using the Kappa statistic was also performed. An intra-class correlation coefficient (ICC) of .817 was achieved between the two assessors, which can be considered 'excellent' according to guidelines given by Fleiss (1986). Disagreements in ratings were resolved by discussion between the two reviewers, where each put forward reasoning for their rating and a compromise was effectively reached.

2.5 Data extraction

A pre-defined form (available upon request) was used to extract data from the included studies prior to synthesis. Relevant data such as the sample size and details, the measures used and the findings were extracted from the publications. In cases where information was unclear, this was recorded as unknown.

3. Results

3.1 Description of studies

The full search yielded 1914 publications. Of these, 685 duplicates and a further 1181 irrelevant references were removed. When inclusion criteria were applied to the remaining 48 publications, 37 were excluded for not meeting these, including 1 meta-analysis, with an additional 1 removed due to unavailability and another 1 removed as it was non-English language. The remaining 9 papers were included in the review, and references of these were

hand-searched but yielded no additional results. No minimum quality threshold was set and this was taken into account during analysis. Figure 1 demonstrates the selection process.

3.2 Characteristics of included studies

The characteristics and findings of all the studies in this review are summarized and arranged according to measures of self-esteem and suggestibility in Table 1. Each study is numbered in superscript in the Table and referred to by their study number in the synthesis.

The number of participants considered within this review of nine studies is 631 (*M*=70.1, range=30–120), with all studies treated as having separate participants. Of these 631, 73 cases did not meet the inclusion criteria of this review, with one participant group falling below 10 years old[5] and another having Autistic Spectrum Disorder (ASD)[6]. Both of these studies were included in this review due to additional appropriate participant groups which were clearly identified and whose data was analysed separately from those who could not be included. The actual number of included participants in this study is therefore 558 (*M*=62.0, range=30–120) and data synthesis is based only on these participants. Only one of the nine studies involved a sample of young people with a mean age under 18 years old[5].

As some researchers have been involved in more than one of the studies included, with similar recruitment methods and locations, it is possible there may have been some overlap of participants[1,2,8,9]. At most, 78 of the participants (14.0%) may have taken part in more than one study. It was not possible to identify the degree of overlap therefore all included studies were treated as separate studies.

Four studies did not contain include enough participants for sufficient statistical power[1,4,8,9]. Samples also tended to be drawn from a specific population (for example, undergraduate students, nurses) affecting the applicability of their results to wider

populations. The countries where studies took place included the UK (n=7), the USA (n=1) and Estonia (n=1).

Five (55.6%) of the nine studies[3,4,7,8,9] reviewed were of a cross-sectional design, and involved the examination of the relationship between self-esteem and suggestibility within a defined population at one point in time. Four (44.4%) of the studies[1,2,5,6] were of case-control design, comparing the level of suggestibility between individuals with differing levels of self-esteem. The majority of studies[3-9] (n=7) adopted a correlational approach, with the remaining studies[12] using ANOVA to make a comparison of means.

Only one of the nine studies[6] investigated only self-esteem and suggestibility, whilst the other eight considered additional factors such as interviewer behaviour or the impact of negative life events. Eight of the nine studies[1,2,3,4,5,7,8,9] used the Gudjonsson Suggestibility Scale (Gudjonsson, 1984, 1997) to measure suggestibility. The only other measure used was created specifically for the study in question[6], and calculated suggestibility scores based on incorrect responses to (mis)leading questions.

3.3 Quality of included studies

The predominant use of cross-sectional design and correlational analysis within the included studies meant that no causal relationships between self-esteem and suggestibility were established. Conclusions drawn, therefore, could only be with regard to an association between self-esteem and suggestibility. The methodological aspects of the included studies are summarized in Table 2.

Whilst all of the studies included a clear operational definition of suggestibility, only one study clearly defined self-esteem[6]. Several studies included small sample sizes[1,4,8,9], and the lack of consideration for additional background factors and demographics reduce the ability to generalize findings beyond the original populations tested and establish a real

association (or lack of such) between self-esteem and suggestibility. With regard to measurement, the assessment of both self-esteem and suggestibility appears to have been carried out consistently within studies, and in the majority of studies the same psychometric assessments have been used. There is, however, a heavy reliance on self-report in the measurement of self-esteem and this is not validated by objective observations or independent raters. This is, perhaps, more of a critique of available measures rather than of the studies themselves, but might be thought to affect the overall quality of their findings. The Gudjonsson Suggestibility Scales (GSS), both the original (Gudjonsson, 1984) and revised version (Gudjonsson, 1997) as well as the parallel form (Gudjonsson, 1997) have been used in all but one of the included studies. The GSS possesses a robust and rigorous research base and relies on a carefully constructed theoretical underpinning. However, there is a relatively small amount of independent research into the various aspects of validity and reliability of the tool. There are some difficulties with score interpretation, notably the large standard errors and lack of classifications with regard to clinically significant scores. Some flaws in the design are also identified, and particularly in the use of a narrative scenario which the respondent has not personally experienced and an outcome in which they are not particularly invested (White & Willner, 2005). These criticisms aside, the widespread use of the GSS in research might be reflective of practitioners' perceived strength of the assessment. Within the included studies, blinding of participants and assessors is not clarified and is therefore for the large part unknown. None of the studies state refusal or attrition rates, and it is unclear as to whether this is due to no difficulties in this area or lack of reporting.

3.4 Descriptive data synthesis

The diversity of the samples, measures employed and divergent design and quality of the included studies made quantitative data synthesis (meta-analysis) unsuitable, and therefore only qualitative data analysis was carried out.

Self-esteem measures differed, with five measures being used across the nine studies. Most prevalent was the use of the Culture-Free Self-Esteem Inventory (CFSEI) (Battle, 1981) in three studies, with the Rosenberg Self-Esteem Scale (Rosenberg, 1965) and the Semantic Differential technique (Osgood, Suci & Tannenbaum, 1957) each being used in two studies. Additional measures included the Behavioural Academic Self Esteem Scale (Coopersmith & Gilberts, 1982) and the Self-Perception Profile for College Students (Neeman & Harter, 1986). None of these tools, therefore, have been robustly tested for correlation with suggestibility, and with differences between the measures in terms of process and final outcome, overall conclusions drawn can only be tentative. In contrast, the majority of the studies employed the Gudjonsson Suggestibility Scales (Gudjonsson, 1984, 1997) in the measurement of suggestibility, making these scores directly comparable.

Quality for cross-sectional studies tended to be deemed as 'reasonable'[4,7,9], with one deemed as 'high'[3] and one deemed 'low'[8]. The highest quality was observed in a study which used the Culture-Free Self-Esteem Inventory (CFSEI), whilst the lowest was observed in a study which used the Semantic Differential technique. For case control studies, quality ranged from 'reasonable'[6] to 'high'[1,2,5]. The study identified as having 'reasonable' quality used the Behavioural Academic Self-Esteem Scale (BASE), whilst studies observed to have 'high' quality used the (CFSEI) or the Rosenberg Self Esteem Scales.

Most pertinent to the concept of suggestibility in this review is the Yield 1 subscale on the GSS, which measures the effects of (mis)leading questions. Mean score on this subscale ranged from 1.67 to 7.90 (out of 15) over the seven studies which employed the GSS as a measure and used this subscale (only Total suggestibility score was used in Peiffer &

Trull, 2000). The other study[6] used response to misleading questions as a measure of suggestibility, and found that these were answered incorrectly at a rate of 52%. Additional subscales found within the GSS are the Shift, Yield 2 and Total suggestibility. The Shift subscale measures the extent to which participants change their answers following negative feedback. Mean scores on this subscale of ranged from 1.72 to 5.50 (out of 20). The Yield 2 subscale measures the extent to which participants yield to misleading questions following negative feedback, and mean scores ranged from 1.31 to 8.10 (out of 15) within the five studies which included this subscale. Total suggestibility represents an overall score, calculated by summing Yield 1 and Shift scores. Within this review, seven studies included this subscale, with mean scores ranging from 3.36 to 13.60 (out of 35). No guidance is given within the GSS manual (Gudjonsson, 1997) for what constitutes an elevated score, but using the rule of more/less than one standard deviation from the mean, norms are shown in Table 3 with mean and standard deviation scores taken from the manual (Gudjonsson, 1997). This table also shows the mean scores on each of the subscales found within this review.

Mean suggestibility scores on each of the subscales within the GSS for the studies within this review fall within one standard deviation of the mean for adults in the general population. These scores also fall within (and often less than) one point of the mean scores given within the GSS manual. This suggests that the overall sample included within this review is comparable in terms of level of suggestibility to the normed sample for the GSS.

Mean self-esteem scores are not comparable between publications in this review due to the diverse nature of the measures used and the designs of the studies.

Two of the case control studies[56] included only one group of participants (controls) who met the inclusion criteria for this review. The data extracted from these studies was therefore analysed alongside data from the cross-sectional studies in terms of correlations. Data from the groups who did not meet the inclusion criteria for this review (in the first of

these studies the cases group was too young, in the second the cases group all had a Learning Disability) were not analysed in this study.

Three of the five cross-sectional studies[4,8,9] and one of the four case control studies[6] found a significant correlation between self-esteem and at least one aspect of suggestibility. Two more case control studies[12] found a main effect of self-esteem on at least one aspect of suggestibility. Remaining studies found no significant correlation between self-esteem and any aspect of suggestibility (Yield 1, n=4; Yield 2, n=2; Shift, n=3; Total suggestibility, n=4).

In terms of response to misleading questions, significant correlations (at the $p<.05$ level) were found in three studies. One of these studies[6] was the only in this review involving children (aged under 18 years) and found a correlation coefficient of .79 for the relevant sample (aged 10-11 years). Dancey and Reidy (2004) offer a rule-of-thumb for strengths of correlation (zero=0; weak=0.1 - 0.3; moderate=0.4 - 0.6; strong=0.7 - 0.9; perfect=1) and this coefficient might therefore be regarded as strong.

The Yield 2 subscale measures response to misleading questions following negative feedback. One cross-sectional study[4] found a significant correlation ($p<.05$) between this aspect of suggestibility and self-esteem, with a correlation coefficient of -.32 (weak negative correlation). This suggested that as self-esteem decreased, response to misleading questions following negative feedback increased.

Two further studies[8,9] found significant correlations between self-esteem and response to misleading questions, with both using the Semantic Differential technique. Factor analysis used in both of these studies revealed slightly different components contributing to self-esteem, although there was some overlap. Response to misleading questions was significantly correlated ($p<.05$) with the 'Competence' (correlation coefficients .59 and .66) and 'Potency' (correlation coefficient .51 and .40) aspects of self-esteem. As the perceived distance between self and experimenter increased, so too did the level of suggestibility. Similar findings were

also presented between the Shift subscale and Total suggestibility subscale in terms of these dimensions of self-esteem.

Two case control studies[1,2] found a main effect of self-esteem on the Shift subscale of the GSS at the $p<.05$ level. The first of these additionally found a significant main effect of self-esteem on the Yield 1, Yield 2 and Total suggestibility subscales at the $p<.001$ level, with lower self-esteem being associated with higher suggestibility.

Studies which reported at least one significant correlation between an aspect of self-esteem and an aspect of suggestibility were deemed to be of 'low' quality (n=1), 'reasonable' quality (n=3) or 'high' quality (n=2). Studies which found no significant correlations were deemed as 'reasonable' quality (n=1) or 'high' quality (n=2).

4. Discussion

The main aims of this systematic review were to comprehensively explore the association between self-esteem and suggestibility, with regard to whether a correlative relationship exists and, if so, the nature of this (positive or negative). In contrast to previous reviews, the current review takes a systematic approach. In addition, it focusses on the role of suggestibility within interviews for police suspects, and as such includes studies relating to those at or above the age of criminal responsibility for England and Wales (over 10 years) rather than on children specifically. This systematic literature review sought to focus on only one specific factor, self-esteem, in an effort to explore whether a revision of the currently recognised factors for vulnerability of age, learning disability and mental health difficulty should be extended to encompass more obscure factors such as self-esteem.

Only nine studies were found to research this area directly after inclusion criteria were applied. The bias generated by the proportion of studies originating from the UK

(encompassing 74% of participants) means that the conclusions drawn from this review can only be tentatively applied to setting and practices in other countries.

Of interest in this review was the association between self-esteem and interrogative suggestibility in a 'typical' population – that is, a population possessing none of the factors currently considered as strongly related to suggestibility and which appear in the Home Office (2014) guidance. This includes age, learning disability and mental health issues (Gudjonsson, 1988; Tully and Cahill, 1984; Warren, Hulse-Trotter and Tubbs, 1991; Redlich, 1999;). Mean scores on each of the GSS subscales were calculated overall, and suggested that the total sample included in this review closely reflected that of the general adult population used to calculate means and standard deviations for the scales themselves (reported in Gudjonsson, 1997).

Reviewed publications demonstrated mixed findings, with some aspects of suggestibility, most notably the response to misleading questions, being significantly associated with suggestibility whilst others showed no significant correlations. Findings were not consistent between studies, and the ability of researchers to explore these as a whole was limited by the vast differences in self-esteem measures employed. Whilst there is some evidence for an association between the two concepts, this is far from definitive and further specific research is certainly required to develop the understanding of the relationship between them.

4.1 Methodological limitations of included studies

4.1.1 Definition and measurement of self-esteem

Many of the studies included lacked definition of self-esteem. Self-esteem as a concept can vary widely depending on the assessment measure or focus. Indeed, some psychometric assessments have gone as far as to specify areas of self-esteem within different

settings, such as the Culture-Free Self-Esteem Inventory (CFSEI-3; Battle, 2002) child version, which considers academic, general, parental/home, social and personal self-esteem as aspects contributing to overall (or 'global') self-esteem. With this in mind, a clear operational definition of self-esteem is of great importance when considering the reach of the results and in applying these to other contexts. Of note might be the relationship between the individual subscales or aspects of self-esteem with suggestibility, and further research might reveal a more significant association between, for example, personal or social self-esteem and interrogative suggestibility.

Included studies used a wide range of self-esteem measures. Although the majority of these were self-report, these were not consistent in terms of the aspects of self-esteem measured. One third of the studies stated use of the Culture-Free Self-Esteem Inventory (CFSEI) (Battle, 1981), although an updated version (Battle, 2002) exists. Given the date of these studies (Baxter, Jackson & Bain, 2003; Bain, Baxter & Fellowes, 2004; Drake, Bull & Boon, 2008) it is possible that the more recent version of this psychometric was used. The CFSEI, which provides a measure of a self-esteem across a number of dimensions and has been validated for use across a wide range of client groups, might be the most appropriate tool for use in relation to such research. The various dimensions of self-esteem measured by this psychometric as well as the Global Self-Esteem Quotient could be compared directly with the subscales of the GSS to provide further analysis of the relationships between each of these, and potentially highlight specific areas in which to provide support or intervention to potentially reduce an individual's suggestibility at a given point in time.

The Rosenberg Self-Esteem Scale (Rosenberg, 1965) was used in two of the publications, although one of these was the Estonian version, validated for use by Pullmann and Allik (2000). Both of these scales allow researchers to calculate an overall score of self-esteem, measured across a variety of contexts and behaviours.

In contrast, the Semantic Differential technique used in two of the studies (Gudjonsson & Lister, 1984; Singh & Gudjonsson, 1984) requires participants to rate their self-perceptions and perceptions of the experimenter, with scores calculated from the distance between these concepts. Although the Semantic Differential technique has been used to measure self-esteem in other studies (Julian, Bishop & Fiedler, 1966; Franks & Marolla, 1976; Tafarodi & Swann, 1995), this has tended to be a calculation of the difference between *How I am generally* and *How I would like to be*, rather than a measure of distance between self and (identified) others. It is arguable as to whether the Semantic Differential technique, as used in the two studies included in this review, is an accurate measure of self-esteem or whether differences in self-perceptions and perceptions of the experimenter might be attributed to other factors.

4.1.2 Measurement of suggestibility

The majority of the studies employed the Gudjonsson Suggestibility Scales (Gudjonsson, 1984, 1997) in the measurement of suggestibility. Whilst alternative versions were used (the original version, the revised version, or the parallel version) by different publications, the process and scoring remains consistent between these and they have been demonstrated to be psychometrically similar in terms of internal consistency (Gudjonsson, 1984; Gudjonsson, 1992) and inter-rater reliability (Richardson & Smith, 1993; Clare, Gudjonsson, Rutter & Cross, 1994) and correlations between the two measures have been acceptable (>.70) (Gudjonsson, 1987b).

The GSS does, however, possess some limitations. Research has indicated that interviewer behaviour can have a significant effect on suggestibility scores (Bain and Baxter, 2000; Baxter and Boon, 2000; Baxter, Boon and Marley, 2006) and, with the exception of two studies that directly investigated this issue, this was not controlled for in the majority of

the included studies. There is also a lack of clarity in the GSS manual in terms of score interpretation, and this reduces the ability of researchers to clarify whether an elevated score is a problematic score (i.e. clinical significance).

The GSS notes within its guidance the importance of participants being blind to the true purpose of the assessment. Conceptually, if a participant knew they were being asked misleading questions and given (inaccurate) negative feedback, this would affect their performance within each of these domains. The majority of the include studies were unclear as to whether participants were blind to the aims and purpose of the studies. If this were not the case, serious questions about the validity of the results would be raised in terms of the suggestibility scores.

4.2 Limitations of the current review

The current inclusion criteria identified papers which studied suggestibility specifically, and on this basis a number of studies were excluded as they studied interrogative compliance instead. These two structures, whilst notably different, have been found to be significantly associated (Gudjonsson, 1989). A further review might therefore be appropriate specifically investigating the association between self-esteem and compliance.

Findings of this review, whilst comprehensive in terms of available publications, are also restricted by methodological and design limitations inherent in the reviewed studies. A limitation of this review developed from the methods used to measure self-esteem in the included publications. With a diverse range of self-esteem measures available, very few of the studies were directly comparable. In contrast, very few instruments are available which directly measure suggestibility, and therefore the majority of publications had made use of the GSS. Whilst this made study results comparable, weaknesses inherent within the GSS limit the findings of the review.

4.3 Conclusions and recommendations

The lack of consistency between findings makes interpretation of these studies difficult, and therefore no firm conclusions can be drawn as to whether self-esteem and suggestibility are significantly associated. Further research in this area which utilizes larger and more representative samples as well as remaining consistent (or at least comparable) in terms of self-esteem measures may provide further insight and clarification.

Self-esteem can be seen as multidimensional in nature, and the measurement of its individual aspects, such as those provided by the CFSEI, offer an opportunity for researchers to carefully evaluate any differences in the relationships between each of these and suggestibility. However, the existing and ongoing debate regarding the definition of self-esteem requires that significant caution be employed in drawing together the findings of separate studies to develop an overall hypothesis about the relevance and relationship of this concept to suggestibility. Research papers in this area which provide a clear and specific definition of self-esteem should be encouraged to enable direct comparisons to be drawn in a more substantial and reliable way.

Traditionally, the GSS is not used specifically to inform police interviews in England and Wales, although may sometimes be used to inform Court proceedings if a suspect is charged with an offence. Time constraints inherent within the judicial system, and particularly with regard to the length of time suspects may be held in police custody, restrict the opportunity for expert opinion about a suspect's potential vulnerability to suggestibility to be sought. The GSS, as an instrument only to be used by specifically qualified professionals, is thus unsuitable for use by either police custodial staff or by (the majority) of Appropriate Adults. However, the GSS does to some extent accurately reflect the circumstances of police interview, where they are asked to recall events and then answer specific questions about

their narrative, and is therefore suitable for use in research scenarios considering the impact or association of individual factors in relation to suggestibility in investigative interviewing.

Interrogative suggestibility is a key issue in the interviewing of both police suspects and witnesses. Whilst some measures have been taken to reduce the incidence of false confessions from suspects, such as the introduction of the Appropriate Adult role, these are often only applied where suspects are considered vulnerable due to age, intelligence (learning disability) or the presence of mental health issues. With a developing body of research into the area of interrogative suggestibility, there is an increasing number of emerging factors. Should further factors be identified as strongly related to suggestibility, the current practices of the police with regards to the identification of 'vulnerable' suspects and the consequential provision of Appropriate Adults might be questioned. Preliminary steps towards an extension of the definition of 'vulnerable', initially by providing additional training to appropriate professionals within the custody environment in order to better identify those with difficulties, might be beneficial in managing this continuing problem. Further reviews summarising the wide research base of other emerging factors may also be a positive step towards change in this area.

References

Bain, S. A., & Baxter, J. S. (2000). Interrogative suggestibility: The role of interviewer behaviour. *Legal and Criminological Psychology, 5,* 123-133.

Bain, S. A., Baxter, J. S., & Fellowes, V. (2004). Interacting influences on interrogative suggestibility. *Legal and Criminological Psychology, 9*(2), 239-252.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(6)*,* 5-17.

Battle, J. (1981). *Culture-free SEI: Self-esteem inventories for children and adults.* Seattle: Special Child Publications.

Battle, J. (2002). *Culture-Free self-Esteem inventories examiner's manual.* Austin, TX: PRO-ED Inc.

Baumeister, R. F. (1998). The self. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 680–740). New York, NY: Random House.

Baxter, J. S., & Boon, J. C. W. (2000). Interrogative suggestibility: The importance of being earnest. *Personality and Individual Differences, 28,* 753-762.

Baxter, J. S., Boon, J. C. W., Marley, C. (2006). Interrogative pressure and responses to minimally leading questions. *Personality and Individual Differences, 40,* 87-98.

Baxter, J. S., Jackson, M., & Bain, S. A. (2003). Interrogative suggestibility: Interactions between interviewees' self-esteem and interviewer style. *Personality and Individual Differences, 35*(6), 1285-1292.

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J.P. Robinson & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes* (pp. 115-160). San Diego, CA: Academic Press.

Brown, J. D., & Marshall, M. A. (2006). The three faces of self-esteem. In M. Kernis (Ed.), *Self-esteem: Issues and answers* (pp. 4-9). New York: Psychology Press.

Bruck, M., & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology, 18*(8), 947-996.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*(3), 403-439.

Clare, I. C. H., Gudjonsson, G. H., Rutter, S. C., & Cross, P. (1994). The inter-rater reliability of the Gudjonsson Suggestibility Scale (Form 2). *British Journal of Clinical Psychology, 33,* 357-365.

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

Coopersmith, S., & Gilberts, R. (1982). *Manual for the Behavioural Academic Self Esteem scale.* Palo Alto, CA: Consulting Psychologists Press.

Conley, R., Luckasson, R., Bouthilet, G (1992). *The criminal justice system and mental retardation: Defendants and victims.* Baltimore: Paul H. Brookes.

Dancey, C., & Reidy, J. (2004). *Statistics without maths for psychology: Using SPSS for Windows.* London: Prentice Hall.

Drake, K., & Bull, R. (2011). Individual differences in interrogative suggestibility: Life adversity and field dependence. *Psychology, Crime & Law, 17*(8), 677-687.

Drake, K. E., Bull, R., & Boon, J. C. W. (2008). Interrogative suggestibility, self-esteem, and the influence of negative life-events. *Legal and Criminological Psychology, 13*(2), 299-307.

Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in post-DNA world. *North Carolina Law Review*, *82*, 891-1007.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder and Stoughton.

Fenigstein, A., & Vanable, P. A. (1992). Paranoia and self-consciousness. *Journal of Personality and Social Psychology, 62*(1)*,* 129-138.

Fleiss, J. L. (1986). *Design and analysis of clinical experiments.* NewYork: John Wiley & Sons.

Franks, D., & Marolla, J. (1976). Efficacious action and social approval as interacting dimensions of self-esteem: A tentative formulation through construct validation. *Sociometry, 39*(4), 324-341.

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*(3), 303-314.

Gudjonsson, G. H. (1987a). Historical background to suggestibility: How interrogative suggestibility differs from other types of suggestibility. *Personality and Individual Differences, 8,* 347-355.

Gudjonsson, G. H. (1987b). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology, 26,* 215-221.

Gudjonsson, G. H. (1988). The relationship of intelligence and memory to interrogative suggestibility: The importance of range effects. *British Journal of Clinical Psychology, 27,* 185-187.

Gudjonsson, G. H. (1989). Compliance in an interrogation situation: A new scale. *Personality and Individual Difference, 10,* 535-540.

Gudjonsson, G. H. (1992). Interrogative suggestibility: Factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences, 13,* 479-481.

Gudjonsson, G.H. (1997). *The Gudjonsson Suggestibility Scales manual.* Hove, UK: Psychology Press.

Gudjonsson, G. H. (1999). The making of a serial false confessor: The confessions of Henry Lee Lucas. *Journal of Forensic Psychiatry, 10*(2), 416-426.

Gudjonsson, G.H., Clare, I., Rutter, S., & Pearse, J. (1993). *Persons at risk during interviews in police custody: The identification of vulnerabilities (Royal Commission on Criminal Justice Research Study No. 12).* London: Her Majesty's Stationery Office.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1,* 83-104.

Gudjonsson, G. H., & Lister, S. (1984). Interrogative suggestibility and its relationship with self-esteem and control. *Journal of the Forensic Science Society, 24,* 99-110.

Gudjonsson, G.H., Sigurdsson, J.F., Asgeirsdottir, B.B. & Sigfusdottir, I.D. (2007a). Custodial interrogation: What are the background factors associated with claims of false confession to police?, *Journal of Forensic Psychiatry & Psychology, 18*(2), 266-275.

Gudjonsson, G.H., Sigurdsson, J.F., Einarsson, E., Bragason, O.O. & Newton, A.K. (2010). Inattention, hyperactivity/impulsivity and antisocial personality disorder. Which is the best predictor of false confessions? *Personality and Individual Differences, 48,* 720-724.

Gudjonsson, G.H., Sigurdsson, J.F., Sigfusdottir, I.D. & Young, S. (2012). False confessions to police and their relationship with conduct disorder, ADHD, and life adversity. *Personality and Individual Differences, 52,* 696-701.

Home Office (2014). *Police and Criminal Evidence Act 1984: Codes of Practice: Code C* (revised edition). London: HMSO.

Julian, J. W., Bishop, D. W., & Fiedler, F. E. (1966). Quasi-therapeutic effects of intergroup competition. *Journal of Personality and Social Psychology, 3,* 321-327.

Kebbell M. R., & Hatton C. (1999). People with mental retardation as witnesses in court: A review. *Mental Retardation, 37,* 179–187.

Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin, 9,* 371-375.

Maras, K. L., & Bowler, D. M. (2012). Brief report: Suggestibility, compliance and psychological traits in high-functioning adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6,* 1168-1175.

Neeman, J., & Harter, S. (1986). *Manual for the Self-Perception Profile for College Students.* Denver, CO: University of Denver, Department of Psychology.

Nolvak, A., & Pullmann, H. (2002). *Isiksusekusimustik NEC, v2. Testi manual.* Tripod Grupp OU.

Norbeck, J. S. (1984). Modification of life event questionnaires for use with female respondents. *Research in Nursing Health, 7,* 61-71.

Numoja, M., & Bachmann, T. (2008). On the role of trait-related characteristics in interrogative suggestibility: An example from Estonia. *TRAMES: A Journal of the Humanities & Social Sciences, 12*(4), 371-381.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. (1957). *The Measurement of Meaning.* Illinois: University of Illinois Press.

Peiffer, L. C., & Trull, T. J. (2000). Predictors of suggestibility and false-memory production in young adult women. *Journal of Personality Assessment, 74*(3), 384-399.

Pullmann, H., & Allik, J. (2000). The Rosenberg self-esteem scale: Its dimensionality, stability and personality correlated in Estonia. *Personality and Individual Differences, 28,* 701-715.

Redlich, A. D. (1999). *False confessions: The influence of age, suggestibility, and maturity.* Unpublished Ph.D. Thesis, University of California.

Redlich, A. (2004). Mental illness, police interrogations, and the potential for false confession. *Law and Psychiatry, 55,* 19 –21.

Redlich, A.D., Summers, A. & Hoover, S. (2010). Self-reported false confessions and false guilty pleas among offenders with mental illness. *Law and Human Behavior, 34,* 79-90.

Register, P. A., & Kihlstrom, J. F. (1988). Hypnosis and interrogative suggestibility. *Personality and Individual Differences, 7,* 549-557.

Richardson, G., & Smith, P. (1993). The inter-rater reliability of the Gudjonsson Suggestibility Scale. *Personality and Individual Differences, 14,* 251-253.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 803-814.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rotter, J. B. (1966). Generalized experiences for internal versus external control of reinforcement. *Psychological Monographs, 80*(1), 1-28.

Singh, K. S., & Gudjonsson, G. H. (1984). Interrogative suggestibility, delayed memory and self-concept. *Personality and Individual Differences, 5*(2), 203-209.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory (STAI).* Palo Alto, CA: Consulting Psychologists Press.

Tafarodi, R. W., & Swann, W. B. Jr. (1995). Self-linking and self-competence as dimensions of global self-esteem: Initial validation of a measure. *Journal of Personality Assessment, 65*(2), 322-342.

Tully, B., & Cahill, D. (1984). *Police interviewing of mentally handicapped persons: An experimental study.* London: Police Foundation of Great Britain.

Vrij, A., & Bush, N. (2000). Differences in suggestibility between 5-6 year olds and 10-11 year olds: The relationship with self confidence. *Psychology, Crime & Law, 6,* 127-138.

Wagstaff, G. (1991). Suggestibility: A social psychological approach. In J. F. Schumaker (Ed.), *Human suggestibility*. New York & London: Routledge.

Warren, A., Hulse-Trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility. *Children, Law and Human Behaviour, 15,* 273-285.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule – Expanded Form.* Unpublished manuscript, University of Iowa, Iowa City.

White, R., & Willner, P. (2005). Suggestibility and salience in people with intellectual disabilities: An experimental critique of the Gudjonsson Suggestibility Scale. *The Journal of Forensic Psychiatry & Psychology, 16*(4), 638-650.

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67,* 555-561.

Ziegler-Hill, V. (2014). Self-esteem: Research. In L. Grossman & S. Walfish (Eds.), *Translating psychological research into practice* (pp. 267-269). New York, NY: Springer Publishing Company.

Table 1. Summary of included studies

Authors and year of study [study type]	Sample details	Exposure measures	Outcome measures	Findings	Quality assessment
Self-esteem measured by Culture-Free Self-Esteem Inventory (Battle, 1981)					
Baxter, Jackson and Bain (2003)[1] [case control]	N = 48 (14 male, 34 female) Age (years): *M*=19.12, *SD*=1.68, range=17-23 All undergraduate psychology students. 200 participants initially recruited. 168 of these responded. The 24 participants with the highest and lowest self-	Culture-Free Self-Esteem Inventory (CFSEI) (Battle, 1981). Participants allocated to either 'abrupt' or 'friendly' interviewer group.	Gudjonsson Suggestibility Scales[ii] (GSS) (Gudjonsson, 1997).	Yield 1: $F(1,44)=32.3$, $p<.001$ Yield 2: $F(1,44)=19.2$, $p<.001$ Shift: $F(1,44)=12.1$, $p<.001$ Total suggestibility: $F(1,44)=28.7$, $p<.001$ An interaction effect was found between self-esteem and interviewer behaviour for Yield 2 score ($F(1,44)=6.4$, $p=.015$) and Shift score ($F(1,44)=8.9$, $p=.004$). Participants with low self-esteem obtained lower scores on these two subscales in the friendly conditions, and higher on these subscales in the abrupt condition (and vice versa for high self-esteem participants).	High

	esteem were selected for this study.				
Bain, Baxter and Fellowes (2004)[2] [case control]	N = 120 (33 male, 87 female) Age (years): *M*=20.12, *SD*=4.96, range=16-47 All first-year undergraduate psychology students. 450 participants initially recruited. 387 of these responded. The 60 participants with the highest and lowest self-esteem were selected for this study.	Culture-Free Self-Esteem Inventory (CFSEI) (Battle, 1981). Participants allocated to either 'abrupt' or 'friendly' interviewer group. Participants also allocated to either 'warning' or 'no warning' group	Gudjonsson Suggestibility Scales[ii] (GSS) (Gudjonsson, 1997).	Main effect of self-esteem only on Shift scores ($F(1,112)=4.3$, $p<.05$). Participants with low self-esteem made significantly more shifts than those with higher levels of self-esteem.	High

Drake, Bull and Boon (2008)[3] [cross-sectional]	N = 60 (27 male, 33 female) Age (years): *M*=26.9, *SD*=11.32, range=18-65 "Randomly selected" but no details as to how. Participants from a variety of occupations.	Culture-Free Self-Esteem Inventory (CFSEI) (Battle, 1981).	Gudjonsson Suggestibility Scales[ii] (GSS) (Gudjonsson, 1997).	Yield 1: r=-.103, p> .05 Yield 2: r=-.084, p>.05 Shift: r=-.202, p>.05 Total suggestibility: r=-.199, p>.05	High
Self-esteem measured by Rosenberg Self-Esteem Scale (Rosenberg, 1965)					
Numoja and Bachmann (2008)[4] [cross-sectional]	N = 61 (20 male, 41 female) Age (years): *M*=20.6, *SD*=2.98, range=18-35 Undergraduate students	(Estonian) Rosenberg Self-Esteem Scale (ERSES) (Pullman & Allik, 2000).	Gudjonsson Suggestibility Scales[iii] (GSS 2) (Gudjonsson, 1997).	Yield 2: r_s=-.32, p< .05 Shift: r_s=-.29, p< .05 Free recall: r_s=.08, p>.05 Yield 1: r_s=-.19, p>.05 Total suggestibility: r_s=-.24, p>.05	Reasonable

	from different universities in one area of Estonia.				
Maras and Bowler (2012)[5] [case control]	N = 62 ASD: n = 32 (24 male, 8 female) No-ASD: n = 30 (22 male, 8 female)	Rosenberg Self-Esteem Scale (Rosenberg, 1965).	Gudjonsson Suggestibility Scales[iii] (GSS 2) (Gudjonsson, 1997).	ASD group: Yield 1: r_s=-.26 Yield 2: r_s=-.23 Shift: r_s=-.01 Total suggestibility: r_s=.13 No-ASD group: Yield 1: r_s=-.34 Yield 2: r_s=-.18 Shift: r_s=-.23 Total suggestibility: r_s=-.29 None significant at the p<.05 level.	High
Self-esteem measured by Behavioural Academic Self Esteem Scale (Coopersmith & Gilberts, 1982)					
Vrij and Bush (2000)[6]	N = 97 Group 1: aged 5-6	Behavioural Academic Self Esteem Scale (BASE) (Coopersmith &	% of incorrect responses to four (mis)leading questions.	Overall: r(97)=-.78, p<.01 Group 1: r(41)=.75, p<.01	Reasonable

[case control]	(n = 41) (20 male, 21 female) Group 2: aged 10-11 (n = 56) (22 male, 34 female) Children from one school. Causasian, middle class background.	Gilberts, 1982) rated by the teacher of the participant.		Group 2: $r(56)=-.79$, $p<.01$	
Self-esteem measured by the Self-Perception Profile for College Students (Neeman & Harter, 1986)					
Peiffer and Trull (2000)[7] [cross-sectional]	N = 103 Females only. Age (years): *M*=19.50, *SD*=.83 All participants single and students from one	Self-perception Profile for College Students (SPCS) (Neeman & Harter, 1986).	Gudjonsson Suggestibility Scales[ii] (GSS) (Gudjonsson, 1997).	Total suggestibility: $r=.04$, $p>.05$	Reasonable

	course at a university. Ethnicity: White (91.4%), African American (3.8%)				
Self-esteem measured by the Semantic Differential technique (Osgood et al., 1957)					
Singh and Gudjonsson (1984)[8] [cross-sectional]	N = 30 (15 male, 15 female) All nurses in a psychiatric hospital.	Semantic Differential technique (Osgood et al., 1957).	Gudjonsson Suggestibility Scales[i] (GSS) (Gudjonsson, 1984).	*Myself generally*: 'Competence' & Yield (immediate): r=.66; p<.001 'Competence' & Total (immediate): r=.64; p<.001 'Competence' & Yield (week delay): r=.45; p<.05 'Competence' & Total (week delay): r=.40; p<.05 *Myself during experiment*: 'Potency' & Yield (immediate): r=.40; p<.05 'Potency' & Shift (immediate): r=.37; p<.05	Low

				'Potency' & Total (immediate): *r*=.50; *p*<.01 *Experimenter*: 'Potency' & Shift (week delay): *r*=-.31; *p*<.05	
Gudjonsson and Lister (1984)[9] [cross-sectional]	N = 50 (25 male, 25 female) Age (years): Male: *M*=26.2 *SD*=10.2 Female: *M*=34.2 *SD*=13.9 From variety of occupations.	Semantic Differential technique (Osgood et al., 1957).	Gudjonsson Suggestibility Scales[i] (GSS) (Gudjonsson, 1984).	Males: Competence & Yield: *r*=.59; *p*<.001 Evaluative & Yield: *r*=.09; ns Potency & Yield: *r*=.51; *p*<.01 Competence & Shift: *r*=.62; *p*<.001 Evaluative & Shift: *r*=.32; ns Potency & Shift: *r*=.16; ns Competence & Total: *r*=.75; *p*<.001 Evaluative & Total: *r*=.23; ns Potency & Total: *r*=.48; *p*<.01 Female:	Reasonable

			Competence & Yield: *r*=.31; ns Evaluative & Yield: *r*=.17; ns Potency & Yield: *r*=.01; ns Competence & Shift: *r*=-.05; ns Evaluative & Shift: *r*=.08; ns Potency & Shift: *r*=-.05; ns Competence & Total: *r*=.21; ns Evaluative & Total: *r*=.13; ns Potency & Total: *r*=-.19; ns

[i] Original version of the GSS [ii] Revised version of the GSS [iii] Parallel version of the GSS

Table 2. Quality of included studies

Study	Study type	Clear definition of self-esteem	Clear definition of suggest-ibility	Adequate total sample	Measures used for self-esteem adequate	Measures used for suggest-ibility adequate	Assessor blinding	Participant blinding	Measurement consistent across participants	Confoun ding factors dealt with	Missing informati on dealt with	Appropriate statistical analysis
Baxter, Jackson & Bain (2003)	Case control	P	Y	P	Y	Y	U	U	Y	U	U	Y
Bain, Baxter & Fellowes (2004)	Case control	P	Y	Y	Y	Y	U	U	Y	U	U	Y
Drake, Bull & Boon (2008)	Cross-sectional	N	Y	U	Y	Y	U	Y	Y	U	U	Y
Numoja & Bachmann (2008)	Cross-sectional	N	Y	Y	Y	Y	U	Y	Y	U	U	Y
Maras & Bowler (2012)	Case control	N	Y	Y	Y	Y	U	U	Y	U	U	Y
Vrij & Bush	Case	Y	Y	N	Y	Y	Y	Y	Y	U	U	Y

(2000)	control											
Peiffer & Trull (2000)	Cross-sectional	P	Y	Y	Y	Y	U	U	Y	U	U	Y
Singh & Gudjonsson (1984)	Cross-sectional	N	Y	N	N	Y	N	Y	Y	U	U	Y
Gudjonsson & Lister (1984)	Cross-sectional	P	Y	U	N	Y	U	Y	Y	U	U	Y

Y = Yes N = No P = Partial U = Unclear

Table 3. Mean and standard deviation scores on the GSS and for total participants within this review

GSS subscale	GSS norms			Current review
	M	*SD*	Normal range	*M* (no. of studies)
Immediate recall	21.3	7.1	14.2 – 28.4	20.8
Yield 1	4.6	3.0	1.6 – 7.6	3.8
Yield 2	5.6	3.8	1.8 – 9.4	4.7
Shift	2.9	2.5	0.4 – 5.4	3.5
Total suggestibility	7.5	4.6	2.9 – 12.1	7.3

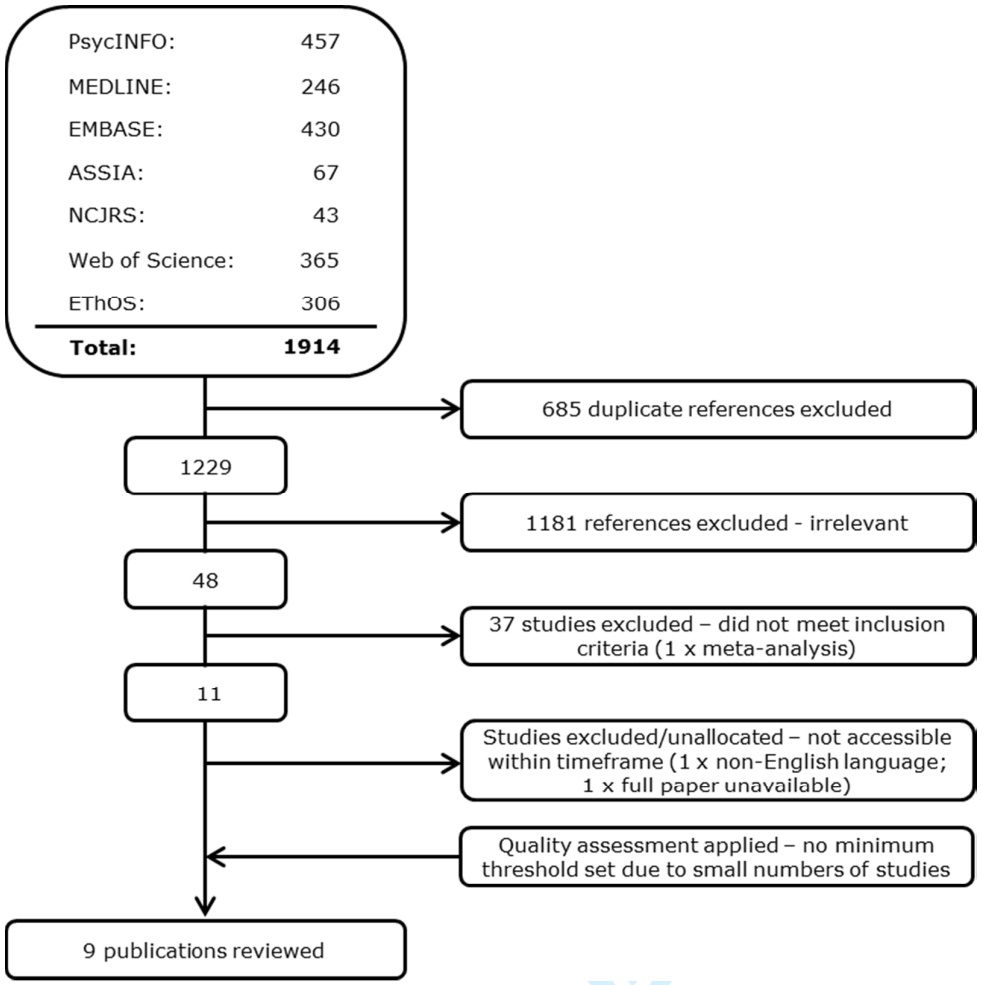

Fig 1. Study selection process