

# Critique of a Measure of interrogative Suggestibility for Children

Rowsell, Kathryn; Colloff, Melissa

DOI:  
[10.1002/acp.4145](https://doi.org/10.1002/acp.4145)

License:  
Creative Commons: Attribution (CC BY)

*Document Version*  
Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*  
Rowsell, K & Colloff, M 2023, 'Critique of a Measure of interrogative Suggestibility for Children: The Bonn test of Statement Suggestibility', *Applied Cognitive Psychology*, pp. 1-8. <https://doi.org/10.1002/acp.4145>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy



While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

## RESEARCH ARTICLE

WILEY

# Critique of a measure of interrogative suggestibility for children: The Bonn test of statement suggestibility

Kathryn Rowsell  | Melissa F. Colloff 

School of Psychology, Centre for Applied Psychology, University of Birmingham, Birmingham, UK

## Correspondence

Kathryn Rowsell, School of Psychology, Centre for Applied Psychology, University of Birmingham, Birmingham, B15 2TT, UK.  
Email: [k.rowsell@bham.ac.uk](mailto:k.rowsell@bham.ac.uk)

## Funding information

British Academy; Economic and Social Research Council, Grant/Award Number: ES/T01475X/1

## Abstract

Professionals supporting child witnesses are concerned with susceptibility to suggestion, that is, the impact of suggestive questioning on the information that children report. The Bonn test of statement suggestibility (BTSS; Endres, *The Journal of Credibility Assessment and Witness Psychology*, 1997, 1, 44–67; Endres et al., *Psychologie der zeugenaussage*, 1998) is a measure of interrogative suggestibility. It focuses on the interaction between the individual child and their environment, which may make a child vulnerable to suggestion. We aim to examine the psychometric properties of the BTSS tool by critiquing the scientific construction of the elements and the robustness of the measure. We conclude that the limited available research allows for a tentative conclusion of suitable levels of reliability and validity. However, we propose that the BTSS should not be administered in isolation to determine suggestibility but instead there should be a wider assessment completed with consideration of the broader aspects of the child's individual situation.

## KEYWORDS

child witnesses, forensic interviewing, suggestibility

## 1 | INTRODUCTION

A principle component of the child witness credibility literature is the measurement of suggestibility. The American Psychological Association (2023) defines suggestibility as an inclination to readily and uncritically adopt the ideas, beliefs, attitudes or actions of others. Within the forensic context, one field of research has focused specifically on the suggestive influence of forensic interviewers on interviewees, known as interrogative suggestibility. Gudjonsson and Clark (1986) define interrogative suggestibility as “the extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as a result of which their subsequent behavioural response is affected” (p. 4). A goal of this body of research has been to determine if there are individuals who are more likely to give inaccurate accounts (trait suggestibility) under perceived external pressure and to explore the mechanisms

underlying this process (e.g., cognitive factors, Milne & Bull, 2003). Researchers have also been particularly interested in the interaction between trait suggestibility, that is the concept that “some people are inherently more suggestible than others” (Ridley, 2013, p. 2) and the environment in which the interview is conducted (social influences, Vrij & Bush, 2000), as reducing exposure to suggestion (e.g., through interviewer training) has been a key goal.

It is crucial for the well-being of child witnesses and the fair administration of justice, to ensure that testimony is not impacted by error, deception or contamination. One method psychologists use to measure the possible suggestibility of children is through psychometric measures; objective and quantifiable tests which are grounded in statistical analysis. One such measure is the Bonn test of statement suggestibility (BTSS, Endres, 1997; Endres et al., 1998) which was created to measure individual differences in suggestibility of children aged 4 to 10 years old. This scale was developed in response to

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Applied Cognitive Psychology* published by John Wiley & Sons Ltd.

criticism that most suggestibility scales designed for adults (such as the Gudjonsson suggestibility scale, GSS, Gudjonsson, 1997) are too complex for young children. The BTSS and GSS are the most commonly used measures of interrogative suggestibility. In this paper, we examine the psychometric properties of the BTSS tool by critiquing the scientific construction and the robustness of the measure. A reliable and valid measure of children's interrogative suggestibility is important both for the fair administration of justice, and also for theory development.

## 2 | OVERVIEW

### 2.1 | The background of interrogative suggestibility

Historically, the suggestibility literature is broadly split into two approaches; the experimental approach (Schooler & Loftus, 1986) and the individual differences approach (Gudjonsson & Clark, 1986). The experimental approach posits that it is possible to mislead people into remembering events incorrectly by introducing certain circumstances, for example, through manipulating the delay between the event and the moment of recall (Schooler & Loftus, 1986). Conversely, the individual differences approach focuses on the differences between witnesses and why these differences might make some more suggestible than others.

Considering forensic practice in criminal courts, Gudjonsson and Clark (1986) created an individual differences model which posits that interrogative suggestibility is a product of the person, the environment, and others within that environment. Thus, when faced with the external conditions of either uncertainty or expectation in forensic interview, interrogative suggestibility is measured by the coping mechanisms the individual can implement at that moment (Gudjonsson, 2013). This model puts forward two main types of suggestibility which can be influenced by individual factors: 'Yield' and 'Shift'. Yield relates to the likelihood of the witness acquiescing to the interviewer's leading questions and Shift relates to the measurement of reactions to interrogative pressure techniques (such as those seen in forensic interviews and cross-examination, like repeated questioning). These two concepts are the basis for the interrogative suggestibility research conducted by Gudjonsson and Clark (1986); the theory was subsequently used to develop a scale for adult witnesses (the GSS) to measure suggestibility in interview and trial situations.

The BTSS adopts the interrogative suggestibility approach, applying the same principles of Yield and Shift to the child witness population. Endres (1997) states the main need for the instrument was because methodological weaknesses had been identified in other suggestibility scales. As mentioned above, it was not deemed appropriate to use adult scales (e.g., the GSS) with children. Other tests attempting to measure suggestibility in children have been criticised for appearing to show reliability and validity limitations. For example, the test of statement suggestibility (Burger, 1971) lacks criterion validity due to absence of significant correlation with other similar tests and the

Würzburg suggestibility test (WST, Bottenberg & Wehner, 1972) is designed as a group task meaning it does not replicate the forensic interview in real life. As such, Endres (1997) and Endres et al. (1998) identified that there was a gap in the suggestibility field for a valid and reliable test for children. The critical question is not whether children are susceptible, but rather to what extent a child with specific individual factors is more susceptible to interrogative suggestibility compared to their peers. The aim of the BTSS therefore, is to provide an instrument which can help to measure this. To this end, we first outline the structure and application of the BTSS, and then consider the tool's psychometric properties.

### 2.2 | Structure and administration of the BTSS

The BTSS adopts the interrogative suggestibility approach, applying the same principles of Yield and Shift to the child witness population. Endres (1997) states the main need for the instrument was due to identified methodological weaknesses in other suggestibility scales (reliability and validity limitations). The answers given by the child on the BTSS are used to indicate their position on Yield, Shift and Total Suggestibility scales. The Yield scale (18 items) measures the child's acquiescence to incorrect information given by an interviewer and comprises of two question types. For yes/no questions (10 items), the child is presented an incorrect fact in an affirmative manner. For alternative questions (8 items), the child is presented two incorrect facts in a manner which demands a forced choice. The Shift scale (8 items) measures whether the child changes their answers in response to negative feedback from the interviewer and is measured by asking children repeated questions; these are immediately presented after the yes/no questions. The overall suggestibility score is the sum of the Yield and Shift scales (26 items) and indicates the extent to which the child is suggestible. It does not include the additional 5 distractor questions which total the 31 questions in the BTSS.

The BTSS is a standardised test with clear and defined instructions for administration. The test does not require a specific level of training for administration and scoring and the results should therefore be similar between different administrators. The advantages to such an approach are that the results remain objective and simple to communicate to others; it gives a clear indication of suggestibility for an individual child. This is an advantage in the forensic domain because the results are often reported to individuals in the criminal justice system (such as judiciary, prosecutors, and defenders) who do not possess formal psychological training.

## 3 | PSYCHOMETRIC PROPERTIES OF THE BTSS

### 3.1 | Internal reliability

Internal reliability is concerned with whether items within the same test are measuring the same construct (Kline, 1998). The statistic often

used to measure this is the Cronbach Alpha Coefficient, denoted by  $\alpha$  (Cronbach, 1951). Psychology is, at best, able to measure behaviour with a margin of standard error and the acceptable Cronbach Alpha Coefficient is generally recognised to be above 0.7 (Kline, 1998).

The initial published reliability findings for the BTSS (Endres, 1997), which sampled 62 children aged between 4 and 10 years old, demonstrated that the three question types (yes/no, repeated and alternative) and Total Suggestibility scales had acceptable ratings for internal reliability indicating consistency throughout the measure (yes/no,  $\alpha = .74$ ; alternative,  $\alpha = .77$ ; repetition,  $\alpha = .70$ ; total suggestibility,  $\alpha = .85$ ; Endres, 1997). Candel et al. (2000) replicated these promising findings when examining the Dutch version of the BTSS with a sample of 48 primary school children aged between 4 and 10 years old. In this study, the yes/no and alternative questions were combined into the Yield scale ( $\alpha = .78$ ) and the repetition questions into the Shift scale ( $\alpha = .82$ ). The Total Suggestibility scale also produced scores within the acceptable range for internal reliability ( $\alpha = .87$ ). Given these results are limited to two studies, we can cautiously conclude that the questions on the Yield, Shift and Total Suggestibility scales are all measuring the same construct; although further empirical research would be of benefit.

### 3.2 | Test-retest reliability

The test-retest principle relates to a test's consistency over time. An individual taking the same test at two different times should produce the same results when no intervention has taken place between the testing phases (Kline, 1998). During the initial testing of the BTSS, Endres (1997) and Endres et al. (1998), delivered the BTSS on two occasions to 62 children aged between 4 and 10 year-olds with a period of several weeks in between. The results demonstrated nonsignificant but positive correlations for Total Suggestibility scores ( $r = .66$ ), and for scores on the yes/no ( $r = .67$ ) and alternative ( $r = .65$ ) questions. Scores for repeated questions produced the weakest nonsignificant correlation ( $r = .32$ ), demonstrating relatively weak correspondence between the same child's score over time.

However, in an exploration of the construct scales, as opposed to question type, Candel et al. (2000) delivered the BTSS to 4 to 10-year-olds ( $n = 48$ ) and retested them on the same story 6 weeks later. They found much stronger retest correlations for total suggestibility scores ( $r = .90$ ,  $p < .05$ ), and for Yield ( $r = .90$ ,  $p < .05$ ) and Shift ( $r = .78$ ,  $p < .05$ ) scores. It is worth interpreting these results with caution because a 6-week retest period may have allowed recall from the previous testing phase. Alternatively, recall from the previous testing phase may not be problematic in this case, because 6 weeks is a long time for a young child to remember details. O'Neill and Zajac (2013), for example, have shown that 5 and 6-year-olds demonstrate a particular detriment to their recall after a time delay of 1 month. Therefore, overall it appears that the construct scales of the BTSS show sufficient test-retest reliability according to the most recent study that has assessed this (Candel et al., 2000). Yet, the evidence for sufficient test-retest reliability for the question types is weak and more research is required.

### 3.3 | Validity

Validity is defined as the concern of "whether an instrument is indeed measuring what it purports to evaluate, that is, the construct of actual interest" (Raykov & Marcoulides, 2011, p. 183). An instrument which is valid is useful as it is able to make inferences about the construct being measured beyond the testing conditions. Defining validity is not an all or nothing science; Raykov and Marcoulides (2011) argue the aim is to develop a strong case for the existence of validity using instrument validation rather than presuming it to be present or not.

#### 3.3.1 | Face validity

Face validity relates to whether an instrument appears to be measuring what it sets out to measure. Being a standardised test, the BTSS aims to measure the extent the participant will concede to suggestive questioning about their memory of a story. Malingering is unlikely, as the participant is shielded from the true nature of the test. Endres (1997) and Endres et al. (1998) added filler questions to distract children from the true nature of the instrument (i.e., suggestibility to misleading information). These filler questions are posed in a leading manner but supply correct information (e.g., "And Bettina's friend was called Michaela, wasn't she?"). Therefore, the BTSS appears to be measuring what it set out to measure, namely suggestibility in the form of participant acquiescence to the presence of external pressures.

#### 3.3.2 | Criterion based validity

Criterion based validity is concerned with the extent to which one can predict a subject's score on another variable or criterion of interest from scores on the instrument (Raykov & Marcoulides, 2011). This is important in circumstances where we wish to evaluate potential future performances. For example, the BTSS is concerned with predicting how likely a child will be susceptible to suggestibility when put in a high demanding situation such as cross-examination in a criminal trial. Criterion validity comprises two elements; concurrent and predictive validity. Endres (1997), at the time of publishing the initial findings into the BTSS, was not able to demonstrate that the BTSS was able to predict subsequent suggestibility in real investigations and it has not been possible to identify any literature which has explored this relationship in this present paper. We are therefore unable to comment on the predictive validity of the BTSS; however, comparable adult scales, such as the GSS, appear to hold acceptable levels of predictive validity (Merckelbach et al., 1998).

#### 3.3.3 | Concurrent validity

Concurrent validity explores how well the test correlates with other tests claiming to measure the same construct; new tests can be evaluated against existing tests to establish if a correlation exists. Roma

et al. (2011) compared the BTSS with the GSS (the adult scale on which the BTSS is based) with 84 children aged between 8 and 10 years old. The authors found strong correlations between scores on the two tests for the Yield and Total Suggestibility subscales ( $r = .71$ ;  $p < .001$  and  $r = .72$ ;  $p < .001$  respectively) but a weaker correlation for the Shift scale ( $r = .33$ ;  $p < .05$ ). This weaker correlation is possibly because the BTSS and GSS assess the Shift variable differently; the BTSS repeats eight of its questions and does so immediately after the child's original response, whereas the GSS repeats 20 questions after a delay. Roma et al. (2011) state that the BTSS replicates real life; forensic interview and cross-examination questioning patterns are more likely to repeat only some of the questions and do so in an immediate fashion.

Roma et al. (2011) compared the BTSS with the similarities and vocabulary subtests of the Wechsler Intelligence Scale for Children Revised (WISC-R, Wechsler, 1974) and Raven's colored progressive matrices test (CPM, Raven et al., 1998). The results showed that the Yield (vocabulary  $r = -.34$ ;  $p < .05$ , similarities  $r = -.38$ ;  $p < .05$ , and CPM  $r = -.24$ ;  $p < .05$ ) and Total Suggestibility (vocabulary  $r = -.24$ ;  $p < .05$ , similarities  $r = -.37$ ;  $p < .05$  and, CPM  $r = -.29$ ;  $p < .05$ ) scales correlated negatively with the cognitive tests with small to medium effect sizes. The correlations for the Shift scale were small to medium effect sizes, but were not statistically significant (vocabulary  $r = -.13$ , similarities  $r = -.26$ , and CPM  $r = -.26$ ). This suggests that the Yield and Total Suggestibility subscales of the BTSS may be associated with cognitive factors, where better cognitive abilities may mean the child is less likely to acquiesce to leading questioning. It is possible that the Shift scale is less associated with cognitive factors, and perhaps more associated with other factors (e.g., situational), but this is yet to be tested. It is also possible, however, that the Shift scale has poorer concurrent validity than the Yield and Total Suggestibility scales.

### 3.3.4 | Content related validity

Content validity is concerned with whether the instrument measures all possible aspects of the construct being measured. This allows for a full and accurate measurement to be made; if elements of the construct are missing the scale would not provide an accurate picture. Interrogative suggestibility theory underpinning the BTSS conceives suggestibility as an interaction between the motivational aspects, cognitive abilities, and individual suggestiveness of the witness. That is, the witness has certain predispositions to interrogative questioning which, under specific circumstances, are likely to heighten subsequent suggestibility. For example, the developmental level of a child (cognitive factor) was examined by Volpini et al. (2016) who, using a sample of 92 children aged between 3 and 6 years old, demonstrated that the Shift subscale was related to situational factors (the desire to appear socially compliant increasing with the age of the child) and the Yield subscale was associated with cognitive factors (the younger the child the less their cognitive facilities have developed and the more likely they will show acquiescence bias). The research by Volpini et al. (2016) indicates that suggestibility is influenced by both cognitive factors (e.g., developmental level, age,

intellectual functioning) and situational factors (e.g., interrogative questioning style, social conformity). The BTSS does not measure either situational or cognitive factors; rather it measures the extent of suggestibility. As such, it is not possible to conclude the BTSS definitively measures the construct of suggestibility as a whole entity; merely the presence of one aspect of it (i.e., the situational factors in forensic interview scenarios). Therefore, for a more thorough measurement of suggestibility, the BTSS may require supplementation with other assessments including clinical interviews with the child, their caregivers or other adults such as teachers who have observed the child over a period of time, and cognitive testing (e.g., WISC-IV, Wechsler, 2003; or CPM, Raven et al., 1998).

### 3.3.5 | Construct related validity

Construct validity, often considered the gold standard of validity, is concerned with the correlation of the construct under investigation (interrogative suggestibility) with other variables known to be linked. This can tell us to what degree performance on a task is influenced by both the construct under investigation and by other related constructs (Raykov & Marcoulides, 2011). There is currently no psychometric measure available to consider convergent and divergent validity because the BTSS superseded all other suggestibility measures for children (e.g., Test of Statement Suggestibility, Burger, 1971; Würzburg Suggestibility Test, Bottenberg & Wehner, 1972). As such, we examined the link between BTSS scores and age and intelligence, because the wider literature proposes that these cognitive and social factors may be associated with susceptibility to suggestion (Bruck & Melnyk, 2004).

### 3.4 | Age factors

The commonplace understanding is that the younger a child is, the more likely they will be suggestible to leading questions (Endres, 1997; Saraiva & Albuquerque, 2015). Research studies in the wider literature have found a negative correlation between age and suggestibility; that is the older a child is, the less likely they will be influenced by suggestible questioning (Finnilä et al., 2003; Singh & Gudjonsson, 1992a; Singh & Gudjonsson, 1992b; Warren et al., 1991). London et al. (2013) posit that the development of Theory of Mind (ToM, an appreciation that others can have an alternative representation of reality) between the ages of 3 and 6 years old can decrease suggestibility due to the understanding of the child that they do not have to accept the alternative viewpoint (false belief) of the interviewer (Templeton & Wilcox, 2000; Welch-Ross, 1999; Welch-Ross et al., 1997).

In accordance with such research and theory, research has generally found a negative relationship between scores on the BTSS and age from preschool up until the age of 11 (Bruck & Melnyk, 2004). In the BTSS study by Endres (1997), the age effect was particularly predominant in the yes/no questions ( $r = -.40$ ;  $p < .001$ ), demonstrating

that younger children are particularly susceptible to suggestibility effects when faced with a forced-choice questioning style. Similarly, Benatti et al. (2018) sampled 1098 children aged between 4 and 11 years old and found a significant negative correlation between suggestibility and age ( $r = -.40$ ;  $p < .01$ ) on the Italian version of the BTSS. Together, this suggests that scores on the BTSS behave as expected by broader findings in the literature regarding age (i.e., BTSS scores decline with age). It also indicates that age should be considered when assessing a child's potential level of suggestibility.

However, the use of age to measure construct validity of the BTSS is complicated because other researchers have found little to no correlation between ToM and suggestibility (Melinder et al., 2006; Scullin & Ceci, 2001). It appears that age alone is unable to provide a clear, linear relationship with suggestibility but is rather reflective of a series of complex developmental variations which are associated with the suggestibility of the child (Crossman et al., 2004). Therefore, other evidence for construct validity should also be considered.

Interestingly, using the GSS, it has been demonstrated that adolescents are no more likely to yield to leading questions than adults (Gudjonsson & Singh, 1984); however, they are more likely to acquiesce to negative feedback on the Shift scale (Singh & Gudjonsson, 1992a; Warren et al., 1991). Conversely, children under the age of 12 are more likely to be suggestible to Yield and Shift scales on both the BTSS and the GSS than adolescent and adult populations (Candel et al., 2000; Gudjonsson, 1984). This indicates that children under age 12 tend to be more suggestible than those over this age; however, adolescents are likely to acquiesce when they are faced with uncertainty about the correct answer, such as when they are faced with repeated questions (Finnilä et al., 2003). Gudjonsson (1984) proposes that from the age of 16, there is very little difference in suggestibility levels compared to groups of adult participants indicating that practitioners should consider applying the GSS rather than the BTSS when assessing for interrogative suggestibility in adolescents from 16 years old.

### 3.5 | Intelligence

Researchers have long explored the possibility that those with lower intelligence (both verbal and non-verbal) are more likely to be susceptible to suggestibility (Muris et al., 2004; Richardson & Kelly, 1995). There is evidence that verbal intelligence, the ability to understand and reason using words, might be negatively associated with suggestibility in children. That is, the higher the verbal intelligence the lower the suggestibility (Chae & Ceci, 2005). There is less available evidence for the relationship between non-verbal intelligence (the ability to analyse information and solve problems using visual or hands-on reasoning) and suggestibility, however Costa and Pinho (2010) provide evidence that 8 and 9-year-olds demonstrated a negative relationship between non-verbal intelligence and suggestibility.

The BTSS also appears to have construct validity as scores on the BTSS consistently correlate negatively with cognitive variables that would be expected to be associated with suggestibility, including

intelligence (Candel et al., 2005; Klemfuss & Olaguez, 2020; Roma et al., 2011; Singh & Gudjonsson, 1992b). Interestingly, Yield demonstrates a stronger negative correlation with cognitive variables than Shift for younger children, suggesting that level of cognitive development is associated with a child's likelihood to assent to social pressure within a forensic interview or cross-examination situation (Roma et al., 2011). Volpini et al. (2016) concur, stating that older children compared to younger children are more likely to acquiesce to negative feedback from an interview and shift their answers as a result. This indicates that administration of the BTSS should be conducted in conjunction with other measures of cognitive ability, to consider other possible influences on suggestibility within the assessment.

It is important to note, however, that some researchers and statisticians have critiqued the notion of a simple linear relationship between intelligence and suggestibility. Gignac and Powell (2006) for example demonstrated a linear negative relationship on the yield scale up to an IQ of 105 (measured by matrix reasoning and vocabulary on the WASI) in a group of 5 to 13-year-olds. After 105 IQ points, this relationship ceased to exist indicating that the association between intelligence (both verbal and non-verbal) and suggestibility is not linear in nature. This should be considered when applying interrogative suggestibility measures alongside cognitive measures in forensic settings.

Moreover, there is a lack of research regarding the use of the BTSS among populations of children with intellectual disabilities. In a review of the broader suggestibility literature, research has demonstrated that children with intellectual disabilities are consistently more suggestible than comparison populations of typically developing children, but children with intellectual disabilities are no more suggestible than children with a matched mental age (Klemfuss & Olaguez, 2020; Young et al., 2003). Therefore, practitioners should be sensitive to mental age—not chronological age—in children with learning impairments when using the BTSS by using standardised measures of cognitive development (such as WISC and CPM) and comparing the outcomes to normative samples of child development.

In sum, the BTSS as a measure generally behaves in the way we would expect in relation to the development of a 'typically developing' child; the less cognitively developed the child, the more likely they are to acquiesce to outside influences, taking into account that this relationship may not always be linear (Gignac & Powell, 2006).

### 3.6 | Recommendations for practitioners and future research

Suggestibility is a multi-faceted concept and interrogative suggestibility, as a construct influenced by individual differences, is just one element of a wider picture. It is not currently clear how other psychosocial factors are associated with suggestibility or BTSS scores. This has two implications: (1) it makes interpretation of BTSS scores for an individual child difficult, and (2) we cannot be sure that the current literature base is representative of how children may have or will perform in forensic interviews or court contexts. For example, Scores on the Yield and Shift scales have previously been shown to positively

correlate with psychosocial factors such as poor assertiveness, anxiety (both trait and state), and avoidance in both adult and child populations (Gudjonsson, 1997). However, a recent review (Klemfuss & Olaguez, 2020) concluded that temperament, social avoidance and distress do not predict suggestibility. It is important to recognise that the current literature has a paucity of data from diverse populations which may help disentangle some of the individual psychosocial mediators of suggestibility; for example, children with learning difficulties, differences between genders or other individual factors, such as experience of trauma, which may be salient to child witness populations (Kanan & Plog, 2015). Klemfuss and Olaguez (2020), for example, concluded that children who had been subject to traumatic experiences were, on balance, more likely to yield to suggestible influences. Children tested using the BTSS are likely to have experienced abuse and trauma; yet there is not enough research available to understand the impact of trauma on suggestibility. The deliberate misleading of subjects inherent in the BTSS instrument is ethically questionable, particularly when applied to vulnerable witnesses, such as children or those who have experienced a traumatic episode. It is currently unclear if BTSS use could lead to further traumatising, as those in positions of power (i.e., the interviewer) provide misleading information to the vulnerable witness, which may be confusing and make the witness feel pressured to answer in certain ways. Further research exploring differences in suggestibility associated with psychosocial factors is necessary to assist practitioners to interpret and understand what an individual suggestibility score on the BTSS means for any individual child. In practice, the importance of developmental and situational factors discussed in this review, as applied to the application of the BTSS, emphasises the need for holistic assessment of young people acting as witnesses and a wider formulation of need from qualified practitioners rather than the singular application of the BTSS. We recommend that practitioners applying the BTSS should be cautious and include other forms of assessment for both situational and cognitive factors before drawing overall conclusions about suggestibility. From a practical standpoint, further research into the relationship between adverse childhood experiences and subsequent suggestibility would also serve to strengthen the literature. Practitioners should therefore be cautious in their reporting of the BTSS to lay professionals regarding this population.

Moreover, children, by virtue of their cognitive development, are more likely to be influenced by power differentials held by professionals within the forensic system than adult populations (Lamb et al., 2011) and therefore should be treated accordingly during forensic interviews. Significant headway has been made regarding the operationalisation of forensic interviews through safeguards against secondary victimisation as a result of the interview process. However, it is likely that serious situations of high stress with important outcomes for the individual and others, such as a forensic interview or cross-examination, are likely to increase susceptibility to suggestion when there is a clear power differential in the relationship (Vrij & Bush, 2000). However, Endres et al. (1999) argue that, although pre-school children are very susceptible to suggestion, there is evidence that this can be limited through robust interviewing techniques.

Indeed, other research shows that children from a young age are able to give accurate accounts if their memories are tested appropriately (e.g., La Rooy et al., 2013; La Rooy et al., 2015; Otgaar et al., 2020). Therefore, robust interviewing techniques, alongside comprehensive assessment protocols can mitigate the impact of interrogative suggestibility. Therefore, the focus should be on ensuring robust interviewing techniques, instead of using the BTSS post hoc.

Finally, it is likely that the main audience for the results of the BTSS will be professionals without psychological training. The misinterpretation caused by the lack of a clear definition of suggestibility, the role of the BTSS (measuring one discrete facet of suggestibility) and the inability of the scale to give a precise measurement of suggestibility means that stringent application is required. Future research should focus on bolstering the evidence base for the construct of interrogative suggestibility in child populations but, in particular, evidence about the internal reliability and predictive validity of the BTSS.

In sum, this critique has discussed the psychometric properties of the BTSS and considered its applicability in forensic settings. The BTSS can be considered a somewhat useful tool for measuring interrogative suggestibility in child witnesses. Yet the bar for what is a 'suitable' measure in forensic settings should be high, because of the implications of misleading assessment. Currently, the literature base is weak in places (e.g., test-retest validity, concurrent validity for shift scores and construct related validity). Therefore, if the BTSS measure is used, it should be applied with caution, interpreted within the wider context, and not be utilised in isolation.

## FUNDING INFORMATION

This work was supported by an Economic and Social Research Council grant [grant number ES/T01475X/1] and a British Academy grant.

## CONFLICT OF INTEREST STATEMENT

We have no known conflict of interest to disclose.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Kathryn Rowsell  <https://orcid.org/0000-0003-4501-3158>

Melissa F. Colloff  <https://orcid.org/0000-0001-6401-4872>

## REFERENCES

- American Psychological Association (2023). <https://dictionary.apa.org/suggestibility>
- Benatti, F., Magro, T., Zuin, M., & Monzani, M. (2018). *Standardizzazione italiana del Bonn test of statement suggestibility BTSS [PowerPoint slides]*. Università IUSVE. <http://www.associazionelaic.it/wp-content/uploads/2018/10/BTSS.pdf>
- Bottenberg, E. H., & Wehner, E. G. (1972). Suggestibilität: II. Einige persönlichkeits und leistungsdiagnostische Korrelate des Würzburger SuggestibilitätsTests (WST) [suggestibility: II. Some personality and achievement correlates of the Würzburg suggestibility test (WST)]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 21, 282–289.

- Bruck, M., & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology, 18*, 947–996. <https://doi.org/10.1002/acp.1070>
- Burger, H. (1971). *Die suggestive Beeinflussbarkeit von Aussagen über Beobachtungen. Entwicklung und erste Überprüfung eines Tests zur Aussagesuggestibilität (TAS) bei der Glaubwürdigkeitsbeurteilung kindlicher und jugendlicher Zeugen [The suggestive malleability of statements about observations. Development and first validation of a Test of Statement Suggestibility (TSS) in the credibility assessment of child and juvenile witnesses]*. Philosophical Dissertation, Universität Freiburg.
- Candel, I., Merckelbach, H., Loyen, S., & Reyskens, H. (2005). "I hit the shift-key and then the computer crashed": Children and false admissions. *Personality and Individual Differences, 38*(6), 1381–1387. <https://doi.org/10.1016/j.paid.2004.09.005>
- Candel, I., Merckelbach, H., & Muris, P. (2000). Measuring interrogative suggestibility in children: Reliability and validity of the Bonn test of statement suggestibility. *Psychology, Crime & Law, 6*(1), 61–70. <https://doi.org/10.1080/10683160008410832>
- Chae, Y., & Ceci, S. J. (2005). Individual differences in Children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology, 19*(4), 383–407. <https://doi.org/10.1002/acp.1094>
- Costa, A., & Pinho, M. S. (2010). Interrogative suggestibility in children ages 8 and 9. *Análise Psicológica, 28*(1), 193–208. <https://doi.org/10.14417/ap.266>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. <https://doi.org/10.1007/BF02310555>
- Crossman, A. M., Scullin, M. H., & Melnyk, L. (2004). Individual and developmental differences in suggestibility. *Applied Cognitive Psychology, 18*(8), 941–945. <https://doi.org/10.1002/acp.1079>
- Endres, J. (1997). The suggestibility of the child witness: The role of individual differences and their assessment. *The Journal of Credibility Assessment and Witness Psychology, 1*(2), 44–67. <http://truth.charleshontsphd.com/JCAAWP/9701/9701.html>
- Endres, J., Poggenpohl, C., & Erben, C. (1999). Repetitions, warnings and video: Cognitive and motivational components in preschool children's suggestibility. *Legal and Criminological Psychology, 4*(1), 129–146. <https://doi.org/10.1348/135532599167725>
- Endres, J., Scholz, O. B., & Summa, D. (1998). Aussagesuggestibilität bei Kindern – Vorstellung eines neuen diagnostischen verfahrens und erste Ergebnisse. In M. Stadler, T. T. Fabian, & L. Greul (Eds.), *Psychologie der zeugenaussage*. München.
- Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K., & Niemi, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology, 85*(1), 32–49. [https://doi.org/10.1016/s0022-0965\(03\)00025-0](https://doi.org/10.1016/s0022-0965(03)00025-0)
- Gignac, G. E., & Powell, M. B. (2006). A direct examination of the nonlinear (quadratic) association between intelligence and suggestibility in children. *Applied Cognitive Psychology, 20*(5), 617–623. <https://doi.org/10.1002/acp.1213>
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*(3), 303–314. [https://doi.org/10.1016/0191-8869\(84\)90069-2](https://doi.org/10.1016/0191-8869(84)90069-2)
- Gudjonsson, G. H. (1997). *The Gudjonsson suggestibility scale*. Psychology Press.
- Gudjonsson, G. H. (2013). Interrogative suggestibility and compliance. In A. M. Ridley, F. Gabbert, & D. J. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 45–61). Wiley Blackwell.
- Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1*(2), 83–104.
- Gudjonsson, G. H., & Singh, K. K. (1984). Interrogative suggestibility and delinquent boys: An empirical validation study. *Personality and Individual Differences, 5*(4), 425–430. [https://doi.org/10.1016/0191-8869\(84\)90007-2](https://doi.org/10.1016/0191-8869(84)90007-2)
- Kanan, L., & Plog, A. (2015). Trauma reactions in children: Information for parents and caregivers. *National Association of School Psychologists Communiqué, 44*(2), 27–28.
- Klemfuss, J. Z., & Olaguez, A. P. (2020). Individual differences in children's suggestibility: An updated review. *Journal of Child Sexual Abuse, 29*(2), 158–182. <https://doi.org/10.1080/10538712.2018.1508108>
- Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. Routledge.
- La Rooy, D., Brown, D., & Lamb, M. (2013). Suggestibility and witness interviewing using the cognitive interview and NICHD protocol. In A. Ridley, F. Gabbert, & D. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 197–217). Wiley-Blackwell.
- La Rooy, D., Brubacher, S. P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., Myklebust, T., Naka, M., Peixoto, C., Roberts, K., Stewart, H., & Lamb, M. (2015). The NICHD protocol: A review of an internationally-used evidence-based tool for training child forensic interviewers. *Journal of Criminological Research, Policy and Practice, 1*(2), 76–89. <https://doi.org/10.1108/JCRPP-01-2015-0001>
- Lamb, M. E., Malloy, L. C., & La Rooy, D. J. (2011). Setting realistic expectations: Developmental characteristics, capacities and limitations. In E. M. Lamb, D. J. La Rooy, C. L. Malloy, & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (2nd ed., pp. 1–15). Wiley-Blackwell.
- London, K., Henry, L. A., Conratt, T., & Corser, R. (2013). Suggestibility and individual differences in typically developing and intellectually disabled children. In A. M. Ridley, F. Gabbert, & D. J. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 129–148). Wiley Blackwell.
- Melinder, A., Endestad, T., & Magnussen, S. (2006). Relations between episodic memory, suggestibility, theory of mind, and cognitive inhibition in the preschool child. *Scandinavian Journal of Psychology, 47*, 485–495. <https://doi.org/10.1111/j.1467-9450.2006.00542.x>
- Merckelbach, H., Muris, P., Wessel, I., & van Koppen, P. J. (1998). The Gudjonsson suggestibility scale (GSS): Further data on its reliability, validity, and metacognition correlates. *Social Behavior and Personality, 26*(2), 203–209. <https://doi.org/10.2224/sbp.1998.26.2.203>
- Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology, 8*(1), 21–38. <https://doi.org/10.1348/135532503762871219>
- Muris, P., Meesters, C., & Merckelbach, H. (2004). Correlates of the Gudjonsson suggestibility scale in delinquent adolescents. *Psychological Reports, 94*(1), 264–266. <https://doi.org/10.2466/PRO.94.1.264-266>
- O'Neill, S., & Zajac, R. (2013). The role of repeated interviewing in children's responses to cross-examination-style questioning. *British Journal of Psychology, 104*(1), 14–38. <https://doi.org/10.1111/j.2044-8295.2011.02096.x>
- Otgaar, H., De Ruiter, C., Sumampouw, N., Erens, B., & Muris, P. (2020). Protecting against misinformation: Examining the effect of empirically-based investigative interviewing on misinformation reporting. *Journal of Police and Criminal Psychology, 36*, 758–768. <https://doi.org/10.1007/s11896-020-09401-2>
- Raven, J. C., Raven, J., & Court, J. H. (1998). *Coloured progressive matrices*. Oxford Psychologists Press.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge/Taylor & Francis Group.
- Richardson, G., & Kelly, T. P. (1995). The relationship between intelligence, memory and interrogative suggestibility in young offenders. *Psychology, Crime & Law, 1*(4), 283–290. <https://doi.org/10.1080/10683169508411965>
- Ridley, A. M. (2013). Suggestibility: A history and introduction. In A. Ridley, F. Gabbert, & D. LaRooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 197–217). Wiley-Blackwell.



- Roma, P., Sabatello, U., Verrastro, G., & Ferracuti, S. (2011). Comparison between Gudjonsson suggestibility scale 2 (GSS2) and Bonn test of statement suggestibility (BTSS) in measuring children's interrogative suggestibility. *Personality and Individual Differences*, 51(4), 488–491. <https://doi.org/10.1016/j.paid.2011.05.003>
- Saraiva, M., & Albuquerque, P. B. (2015). Influence of age, social desirability and memory in children's suggestibility. *Psicologia: Reflexao e Critica*, 28(2), 356–364. <https://doi.org/10.1590/1678-7153.201528216>
- Schooler, J. W., & Loftus, E. F. (1986). Individual differences and experimentation: Complementary approaches to interrogative suggestibility. *Social Behaviour*, 1, 105–112.
- Scullin, M. H., & Ceci, S. J. (2001). A suggestibility scale for children. *Personality and Individual Differences*, 30(2), 843–856. [https://doi.org/10.1016/S0191-8869\(00\)00077-5](https://doi.org/10.1016/S0191-8869(00)00077-5)
- Singh, K., & Gudjonsson, G. (1992b). The vulnerability of adolescent boys to interrogative pressure: An experimental study. *The Journal of Forensic Psychiatry*, 3(1), 167–170. <https://doi.org/10.1080/09585189208407634>
- Singh, K. K., & Gudjonsson, G. H. (1992a). Interrogative suggestibility among adolescent boys and its relationship with intelligence, memory, and cognitive set. *Journal of Adolescence*, 15(2), 155–161. [https://doi.org/10.1016/0140-1971\(92\)90044-6](https://doi.org/10.1016/0140-1971(92)90044-6)
- Templeton, L. M., & Wilcox, S. A. (2000). A tale of two representations: The misinformation effect and children's developing theory of mind. *Child Development*, 71(2), 402–416. <https://doi.org/10.1111/1467-8624.00153>
- Volpini, L., Melis, M., Petralia, S., & Rosenberg, M. D. (2016). Measuring children's suggestibility in forensic interviews. *Journal of Forensic Sciences*, 61(1), 104–108. <https://doi.org/10.1111/1556-4029.12987>
- Vrij, A., & Bush, N. (2000). Differences in suggestibility between 5-6 and 10-11 year olds: The relationship with self-confidence. *Psychology, Crime & Law*, 6(2), 127–138. <https://doi.org/10.1080/10683160008410837>
- Warren, A., Hulse-Trotter, K., & Tubbs, E. (1991). Inducing resistance to suggestibility in children. *Law and Human Behavior*, 15(3), 273–285. <https://doi.org/10.1007/BF01061713>
- Wechsler, D. (1974). *Examiner's manual: Wechsler intelligence scale for children—revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler intelligence scale for children—fourth edition*. The Psychological Corporation.
- Welch-Ross, M. K. (1999). Preschoolers' understanding of mind: Implications for suggestibility. *Cognitive Development*, 14, 101–113. [https://doi.org/10.1016/S0885-2014\(99\)80020-3](https://doi.org/10.1016/S0885-2014(99)80020-3)
- Welch-Ross, M. K., Diecidue, K., & Miller, S. A. (1997). Young children's understanding of conflicting mental representations predicts suggestibility. *Developmental Psychology*, 33, 43–53. <https://doi.org/10.1037/0012-1649.33.1.43>
- Young, K., Powell, M. B., & Dudgeon, P. (2003). Individual differences in children's suggestibility: A comparison between intellectually disabled and mainstream samples. *Personality and Individual Differences*, 35(1), 31–49. <https://doi.org/10.1016/S0191-8869%2802%2900138-1>

**How to cite this article:** Rowsell, K., & Colloff, M. F. (2023). Critique of a measure of interrogative suggestibility for children: The Bonn test of statement suggestibility. *Applied Cognitive Psychology*, 1–8. <https://doi.org/10.1002/acp.4145>